

July 2nd, 2015

Editor for *Biogeosciences*

Dear Editor Dr. Williams,

Thank you for the support of this study. We are re-submitting the manuscript entitled “Probing the past 30-year phenology trend of U.S. deciduous forests” to *Biogeosciences*.

All three reviewers suggest more discussion of model uncertainties and analyses of the impacts of species aggregation on the predicted phenological trend. As a result, we added a new paragraph in section 4.2 to discuss uncertainties from the model structure, climate variability, and data availability (or incompleteness). We performed two additional sensitivity tests (simulations 4 and 5), each of which uses parameters calibrated based on species with either the lowest or the highest temperature sensitivity, to quantify modeling uncertainties from the intraspecific variations (see section 4.3 and new Figure S18). We included a recent remote sensing study by Buitenwerf et al. (2015) in the comparison of modeling trends with the ensemble of observations (revised Figure 7). We explained more clearly how we derive observed phenological dates from different products (see supplement and many changes embedded in the main text) and how to calibrate model (see new Tables S4 and S5). We appreciate comments from all three reviewers that have improved this study.

Please find our point-to-point responses below to your decision letter as well as the comments from three reviewers. The line numbers listed in this response are correspondent to the revised manuscript without tracking changes, which has been submitted separately. A mark-up version of manuscript is attached with this response at the end. The original comments are shown in black with our responses marked as blue.

Thanks for your consideration of our submission.

Sincerely,

Xu Yue
School of Forestry and Environmental Studies
Yale University
New Haven, CT, 06511, USA
E-mail: xuyueseas@gmail.com

Responses to Editor Dr. Williams

Dear Authors:

I have now read the reviews of your submission and examined the evaluations of the authors. Overall the reviews are supportive, and your manuscript should be acceptable for publication after some revisions, which I will review.

In adjusting your manuscript, please respond to each point of the reviewers, and clearly annotate the changes.

Particularly address the following critical points:

Clearly address the issue of species-level responses, evaluating the generality of your model at species level, and the implications of community diversity on phenology. If the PFT concept is being challenged, what does this mean for your model, and its generality?

→ In response to reviewers' comments, we have estimated the impacts of species-aggregation on the estimated forest phenology in sections 4.2 and 4.3 in the following steps:

(1) We discussed the possible uncertainties of PFT-level simulation in response to the comments from reviewers 1 and 3 as follows (section 4.2, Lines 616-625):

“Third, we choose to calibrate the phenological parameterization at the level of plant functional type (PFT) because species-specific measurements are usually incomplete in time and uneven in space. Such incompleteness may influence the accuracy of derived decadal phenological records used for both model calibration and validation. At the same time, PFT-level parameterization may be too broad for the vegetation modeling because it fails to capture intraspecific variations (Van Bodegom et al., 2012; Reichstein et al., 2014). Observations at the community level suggest that the budburst of some species is sensitive to fall/winter and spring warming but with opposite signs (Cook et al., 2012). In the next subsection, we examine the records of 13 deciduous tree species at Harvard Forest”.

(2) We explained why we use PFT-level parameterization in this study (section 4.3, Lines 633-639):

“We do not perform the species-specific simulation for the following three reasons. First, the species-level measurements are usually not available on the continental scale, which influences both model calibration and validation. Second, species-level modeling increases the complexity and computational costs while decreasing predictive reliability (Prentice et al., 2015). Third, investigations at both site level and continental scale show similar temperature sensitivity of tree phenology between the species-specific and species-aggregation approaches.”

(3) We show that temperature responses of plant phenology are similar for individual species at Harvard forest (Figure 8 and corresponding discussion in section 4.3, Lines 641-669, not shown here because they are not revisions)

(4) We analyzed continental-scale phenology using derived parameters calibrated with different species, in response to comments from reviewer 2 (section 4.3, Lines 671-683):

“We perform two sensitivity runs to evaluate the modeling uncertainties due to intraspecific variations at the continental scale (Fig. S18). Simulations with either the lowest (simulation 4) or the highest (simulation 5) temperature sensitivity yield very similar phenological trends as that in the control simulation (simulation 2). In the East, simulation 4 predicts a spring advance by 0.33 day yr^{-1} while simulation 5 predicts an advance by 0.35 day yr^{-1} , both of which are close to the 0.34 day yr^{-1} from the control run. In the West and Northeast, both sensitivity runs predict autumn delay by $0.13\text{-}0.15 \text{ day yr}^{-1}$, lower than value of $0.14\text{-}0.16 \text{ day yr}^{-1}$ from control run, suggesting that site-level responses may not be necessarily consistent with responses at the continental scale. Both the similar temperature sensitivity at site level (Fig. 8) and the predicted phenological trends at continental scale (Fig. S18) support the concept of phenological modeling at the forest and PFT level, and corroborates the further investigation of phenology-climate interactions at the continental and global scale.”

The model validation is not totally clear – I am expecting thousands of evaluations against NPN data, but see only a few dozen points in fig 3. Please clarify. The continental-scale validation is critical to explain, and needs to address bias, random error and correlation.

→ We have clarified in the figure caption of Fig.3 as follows:

“The number of sites shown in the plot is much fewer than the total of 588 because only a small portion of the sites has continuous records for both years and a stringent screening process is applied to derive phenological dates (see Supplement).”

In the supplemental material (the main text is already a little bit lengthy), we added following descriptions to explain how we derive the NPN phenological dates at each site:

“We also derive phenological dates for the USA National Phenology Network (USA-NPN). The network has limited records before 2009 but is significantly enriched thereafter. Currently, there are >600000 records for >20000 individual plants at >6500 sites. Due to the large data amount, we select observations during 2011-2012 for 52 deciduous tree species that are the most common in the U.S. (Table S1). In total, we have records at 588 sites (Table 1); however, not all of these sites provide continuous year-round observations. We apply the following stringent screening filter to derive the phenological dates with high confidence. Each USA-NPN record reports 2-10 phenological statuses, including breaking leaf buds, $\geq 75\%$ of full leaf size, $\geq 50\%$ of

leaves colored, $\geq 50\%$ of leaves fallen, all leaves fallen, and so on. We determine budburst date as the first day with ‘breaking leaf buds’ and dormancy onset as the first day with ‘all leaves fallen’. To ensure that the selected date is the real ‘first day’ for the phenological shift, we require that there is at least one record beforehand showing no onset of the phenological events. For example, a record with true status of ‘breaking leaf buds’ is not used if no records in the earlier days of this year report a false status for the same event. The derived dates for individual trees are averaged among species for every site to achieve the spatial distribution of phenology.”

We have added following analyses for the continental-scale validation as suggested.

“The simulated spatial pattern is consistent with phenology records from the USA-NPN network, with a correlation coefficient of 0.77 over 46 sites (Fig. 3b). However, the predicted budburst date at these sites is 5 days on average later than the observations. Causes of such bias are unclear but might be related to the uneven spatial distribution of network sites, the distinct nature of the protocols between calibration and validation data, and the incompatibility between model parameters derived at regional scale versus those for the continental scale.” (Lines 445-451)

“The validation against observations from 23 USA-NPN sites yields a significant correlation coefficient of 0.80 for the simulated autumn phenology (Fig. 3d). Similar to the spring budburst, predicted dormancy onset is later by 11 days than the observations.” (Lines 454-457)

In response to your next comment, we also added more analyses and discussion of the temporal variations of simulated phenology at continental scale (see the responses below).

I agree with the reviewer that “ $p < 0.2$ is a very generous threshold”. I do not follow your response. Please think carefully about this issue.

→ We have recalculated the correlations with a more stringent threshold of $p < 0.1$ in this revision. We also explained the implications of these new results as follows (Lines 463-472):

“The correlations between modeled and observed budburst dates are larger than 0.3 for 47 out of 59 sites, among which 26 are significant ($p < 0.1$), suggesting that the predicted interannual variation and long-term trend of spring phenology are generally reasonable on the continental scale. The insignificant correlations at 33 sites are in part attributed to the deviations in species between model (PFT level) and measurements (species level). Furthermore, the record length may also contribute to these biases because 28 out of the 33 sites with insignificant correlations have records shorter than 8 years. The large interannual variability in the spring phenology (see section 4.1) may affect the correlations especially for time series with short record length.”

I looked at the Buitenwerf paper and it seems to me there are comparisons that you could usefully make with their figure 2 and 3. Figure 2c might be at odds with your results, and would be worth discussing.

→ We compared our results with that in Buitenwerf et al. (2015) as suggested. However, we based our comparison on Figures S2 and S3, instead of Figures 2 and 3. The figures in the main text show standard deviations, which are not the absolute changes in phenological dates but a measure of the severity of change. The figures in the supplement show changes in the seasonal cycle of NDVI for the 1980s and 2010s, which can be used as an indication of phenological changes. In the main text, we compared our results with that in Buitenwerf et al. (2015) as follows: (section 3.3, Lines 521-530)

“In a recent study, Buitenwerf et al. (2015) found an overall extension of the growing season over boreal and temperate forests during 1981-2012 based on the Normalized Difference Vegetation Index (NDVI) from satellite data. However, the exact phenological changes that underlie such overall greening differ among regions. For U.S. forests, the longer growing season is primarily driven by later leaf-off dates, though regional advance of spring is also observed. Our results are generally consistent with their conclusions but with some deviations. For example, they observed later autumn in almost all the eastern U.S., where we predict delays only in the North and Northeast (Fig. 5b). Such discrepancies reflect prediction biases, and may also be a consequence of satellite retrieval uncertainties (Table 1).”

We also added the results of Buitenwerf et al. (2015) in Table 1 and Figure 7 to extend our comparisons with ensemble estimates from multiple literature using remote sensing data. The updated Figure 7 shows distributions of phenological changes even closer to our prediction (Figure 5) than the original results, further supporting our estimate of the phenological trends in U.S. using the well-calibrated model.

Please also examine the structure of the text to improve the focus of the Results section and provide a longer and separate Discussion section. This helps the readers

→ We have moved all the original section 3.4, which discusses the uncertainties of the simulation, to an independent section 4 as Discussion section. Relative to the original version, we have added the following paragraphs to further discuss model uncertainties:

(1) Uncertainties due to model structure (i.e. whether the model should include chilling and photoperiod limits) (Section 4.2, Lines 602-625)

“Our investigation of the roles of chilling and photoperiod is sensitive to the model structure, climate variability, and data availability. First, the similar performance between spring warming and chilling models might also result from the inaccurate representation of chilling / photoperiod mechanisms. For example, the chilling units used in our parameterization is calculated based on daily average temperature, while Piao et al.

(2015) suggested that leaf unfolding dates during 1982-2011 are triggered by daytime more than by nighttime temperature. The up-to-date autumn phenology model fails to capture interannual variability of dormancy onset (Fig. 2), suggesting that unknown processes are involved in the autumn leaf fall (Keenan and Richardson, 2015). It is unclear whether these processes are related to the variations of photoperiod. Second, the decadal changes in temperature may mask the role of chilling. The trend of winter warming is not significant for most areas in the U.S. (Fig. S14a), suggesting that chilling requirements have been fulfilled in the past 3 decades. However, it is unclear whether the winter warming will intensify in the future, which may slow the advancement of spring budburst. Third, we choose to calibrate the phenological parameterization at the level of plant functional type (PFT) because species-specific measurements are usually incomplete in time and uneven in space. Such incompleteness may influence the accuracy of derived decadal phenological records used for both model calibration and validation. At the same time, PFT-level parameterization may be too broad for the vegetation modeling because it fails to capture intraspecific variations (Van Bodegom et al., 2012; Reichstein et al., 2014). Observations at the community level suggest that the budburst of some species is sensitive to fall/winter and spring warming but with opposite signs (Cook et al., 2012). In the next subsection, we examine the records of 13 deciduous tree species at Harvard Forest.”

(2) Uncertainties due to species aggregation (Section 4.3, Lines 671-683)

“We perform two sensitivity runs to evaluate the modeling uncertainties due to intraspecific variations on the continental scale (Fig. S18). Simulations with either the lowest (simulation 4) or the highest (simulation 5) temperature sensitivity yield very similar phenological trends as that in the control simulation (simulation 2). In the East, simulation 4 predicts a spring advance by 0.33 day yr^{-1} while simulation 5 predicts an advance by 0.35 day yr^{-1} , both of which are close to the 0.34 day yr^{-1} from the control run. In the West and Northeast, both sensitivity runs predict autumn delay by $0.13\text{-}0.15 \text{ day yr}^{-1}$, lower than value of $0.14\text{-}0.16 \text{ day yr}^{-1}$ from control run, suggesting that site-level responses may not be necessarily consistent with responses at the continental scale. Both the similar temperature sensitivity at site level (Fig. 8) and the predicted phenological trends at continental scale (Fig. S18) support the concept of phenological modeling at the forest and PFT level, and corroborates the further investigation of phenology-climate interactions at the continental and global scale.”

best wishes

Mathew Williams

Responses to Reviewers

Anonymous Referee #1

We thank the referee for the positive, helpful and comprehensive review that has improved the manuscript. Detailed point-by-point responses to the reviewer comments are provided below. The reviewers' comments are shown in black with our responses are marked as blue. The line numbers below refer to the revised manuscript to be submitted separately.

General comments

In this contribution, the authors parameterize an ensemble of models of spring and autumn phenophases and use the models (1) to derive continental maps of phenological trends in US deciduous forests over the period of, at most, 1982 to 2012 and (2) to evaluate the role of model components (chilling function, role of photoperiod, GDD/CDD summation) on the predicted trends.

The paper is overall well written though a bit lengthy. I do appreciate the details the authors give on methodological aspects, and the multiple checks (against ground observation and remote sensing data) they present to corroborate their results, but probably some of this material would better be placed in the supplementary material section (see my specific comments). As a formal comment, notice that the “Results” section is more a mixture of results and discussion, while the “Conclusion and discussion” section is rather a “conclusion”. I suggest the authors to rename the sections accordingly.

The science is sound, Results are well presented and, considering the spatial extent covered in the paper, new for the US zone (similar results centred on the European zone have already been published, see notably Menzel et al., 2006; Menzel, 2013).

→ We thank the reviewer for positive evaluations. We have moved several parts to the supplementary material as suggested. We changed “Conclusions and discussion” to “Conclusions” and added a separate “Discussion” section.

Some aspects of the paper deserve further discussion / modifications:

(1) Phenological models are usually parameterized at the species level (parameters are believed to be species-specific, referring to species proper characters). The authors choose to parameterize the models on phenological metrics derived from LAI time series acquired in mixed deciduous forests. The model validation is undertaken against an ensemble of data, including a spectrum of forest and non-forest (e.g. Lilac) woody species, which (partly) differ from the species for which the models were calibrated.

Questions related to that aspect:

- The model validation is possibly affected by the choice of averaging at a given site all available phenological data (L2-3 p 6044). The local species composition will obviously affect the observed mean. Furthermore, interrupted pheno time series are frequent, so that I suppose the number of species observed at a given location can change over time, and bias the calculation of pheno average dates.

- The main results of the paper (trends derived from the simulations and the relative little influence of the chilling function / photoperiod on determining the date of budburst / senescence) are dependent on the authors choice (fitting ensemble models over all available data, with no attention to the particular response of a species). Yet, there is evidence in the literature that different, co-existing, species exhibit different phenological trends (in response to temperature changes: see e.g. Vitasse et al., 2009; Fu et al., 2014a). I expect the authors to at least discuss that aspect, and at best to test the genericity of their results with respect to particular species, known to be early/late flushers/senescent.

- The question of model parameterization at the PFT level (“temperate deciduous broadleaf species”, DBF) should be linked with recent attempts to re-define PFTs. The validity of the PFT concept for vegetation modelling is under discussion (see e.g. Reichstein et al., 2013; van Bogedom et al., 2012). If all DBF species show similar trends (see my point above), then using a PFT-parameterized model is OK when projecting the future / past of forest phenology

→ We agree that data incompleteness and species-aggregation may affect our conclusions. Other reviewers have also raised similar concerns. We considered all the related comments and added the following discussion in the revised paper: (Section 4.2, 602-625)

“Our investigation of the roles of chilling and photoperiod is sensitive to the model structure, climate variability, and data availability. First, the similar performance between spring warming and chilling models might also result from the inaccurate representation of chilling / photoperiod mechanisms. For example, the chilling units used in our parameterization are calculated based on daily average temperatures, while Piao et al. (2015) suggested that leaf unfolding dates during 1982-2011 are triggered by daytime more than by nighttime temperature. The up-to-date autumn phenology model fails to capture interannual variability of dormancy onset (Fig. 2), suggesting that unknown processes are involved in the autumn leaf fall (Keenan and Richardson, 2015). It is unclear whether these processes are related to the variations of photoperiod. Second, the decadal changes in temperature may mask the role of chilling. The trend of winter warming is not significant for most areas in the U.S. (Fig. S14a), suggesting that chilling requirements have been fulfilled in the past 3 decades. However, it is unclear whether the winter warming will intensify in the future, which may slow the advancement of spring budburst. Third, we choose to calibrate the phenological parameterization at the level of plant functional type (PFT) because species-specific measurements are usually incomplete in time and uneven in space. Such incompleteness may influence the accuracy of derived decadal phenological records used for both model calibration and validation.

At the same time, PFT-level parameterization may be too broad for the vegetation modeling because it fails to capture intraspecific variations (Van Bodegom et al., 2012; Reichstein et al., 2014). Observations at the community level suggest that the budburst of some species is sensitive to fall/winter and spring warming but with opposite signs (Cook et al., 2012). In the next subsection, we examine the records of 13 deciduous tree species at Harvard Forest.”

(2) Protocol of model validation: ground observation data from phenological networks are by far the most abundant source of data for validation (see Table 4). The reader would like to see a figure comparing dates derived from LAI metrics to autumn dates derived from phenological observations (Fig. 2 does not report such a comparison for autumn).

→ We have shown more date comparisons between phenological records and LAI-derived ones in the updated Fig. S2. The LAI-derived dates at Harvard Forest are not complete and we mentioned in the SI that: “However, for US-Ha1, date records are comprehensive while LAI measurements are incomplete and as a result we do not use the LAI-derived dates in the model validations at this site (e.g. Fig. 1)”. In the main text section 3.4.3, we also explained that: “The 21-year average of the ensemble dormancy onset date is DOY 306, again close to the estimate of DOY 310 based on LAI (Table 3).” These statements and results show that LAI-derived dates are a reasonable proxy for missing phenological records for the model validation.

(3) The use of the “dormancy” term is not adequate to qualify leaf fall (as derived from LAI measurements or Phenocams). Please remind that dormancy is a physiological state of the bud, starting right after budset (i.e. concomitantly to the timing of height growth cessation, i.e. in the middle of summer, when leaves are still present and green) and ending at budburst during next spring (e.g. Delpierre et al., 2015). Hence, please replace occurrences of “dormancy” by “leaf fall” throughout the manuscript.

→ In the first paragraph of section 2.2, we added the following statement to explain the ‘dormancy onset’ date: “Most of phenological records are discrete and evaluation of the annual cycle of tree phenology is difficult. Following definitions in earlier literatures (e.g., Zhu et al., 2012), we validate spring budburst dates (or the onset of growing season, the dates D1 in Fig. S1) and dormancy onset dates (or the end of leaf fall period, the dates D2 plus falling length L2 in Fig. S1) predicted by phenology models with the site-level records. The dormancy onset date defined here is based on the canopy level instead of the bud dormancy examined in a recent review paper by Delpierre et al. (2015).” (Lines 172-179)

Specific comments

L25, p 6039: the “temperature sensitivity to altitudinal trends” is not clearly defined. Altitudinal trends are first and foremost T-related. Please rephrase.

→ We have removed this sentence because we do not examine the impacts of altitude on

phenological responses in this study.

L6-14 p 6041: models calibrated at the species scale cannot pretend to estimate phenological trends for "US deciduous forests", merely for certain species

→ The models were calibrated with derived phenological dates from LAI measurements, which represent the average phenophase of multiple species. We also validated the selected models with average dates across species, though uncertainties exist due to data incompleteness. "Both the similar temperature sensitivity at site level (Fig. 8) and the predicted phenological trends at continental scale (Fig. S18) support the concept of phenological modeling at the forest and PFT level, and corroborates the further investigation of phenology-climate interactions at the continental and global scale." (Lines 680-683)

L22 p 6041: were data from 1000 ground observation sites used? Not comparable with the 4 ground observation sites used for calibration.

→ We clarify that "In total, phenological observations from 1151 sites were used for model validation." (Line 134)

L23 p 6042 – and on: unclear to me where are the four calibration sites located in the US. Refer to Figure 1 here.

→ In the revised paper, we refer to Fig. 1 even earlier at the first sentence of section 2.1 as follows: "Decadal measurements ... to calibrate parameters of the phenology model (Table 2 and Fig. 1)" (Lines 141-143). We also refer to Fig. 1 in one sentence before the original L23 P6042. (Now Line 166).

P 6042 §2: only four out of the 5 calib sites are mentioned (Hubbard Brook missing). Is it because the LAI data were not available at the US-HB site? In that case, against which data were the model calibrated for US-HB?

→ No, the LAI data were not available at the US-HB site. We do not use data from US-HB site to calibrate model parameters. Instead, we use the phenological dates at this site to validate the calibrated models and to check whether these models could perform well at locations beyond the calibration sites.

P 6045: equations 6 and 7: which arguments have driven the choice of the parameters to optimize / to fix?

→ The fixed parameters in Equation 6 and 7 were adopted from Chuine et al. (1999). We calibrated other parameters based on RMSE between model prediction and observations. We presented all the calibrated parameters in Table S4.

Equations 2 to 12: report parameters units in the text, or in a dedicated Table.

→ Both the values and units of the optimized parameters have been shown in the new Table S4.

Equations 8 and 11 share the same variable name (f_T) for two independent variables. Use different names.

→ The name of spring phenology has been changed to f_S .

L17-20 p 6049: conditions of the SA are unclear to me. Please reconsider and rephrase.

→ We further explained the sensitivity analysis as follows: “In this run, we do not include chilling constraint for the spring phenology by using a fixed and calibrated forcing threshold F^* of 50 degree days. As a result, forcing value S_f begins accumulation from winter solstice and budburst occurs if only $S_f > F^*$. The whole process is not dependent on the value of chilling units S_c . Meanwhile, we lift the photoperiod cap for leaf senescence by setting $f_p = 1$, so that the autumn phenology is only determined by temperature ($f_A = f_T$).” (Lines 367-372)

L10 p 6050: replace “decreasing” by “increasing” (higher AIC means the model has a worse accuracy-parameterization trade-off).

→ Yes, we have corrected ‘decreasing’ to ‘increasing’ as indicated.

L15-16 p 6050: which physiological processes do you point here? What are the “synthesis, viscosity, diffusion” processes you mention? Which role do they / are they supposed to play in leaf senescence? If no precision is given, this sentence should be deleted.

→ Physiological processes are important for plant phenology. For example, “the abundance or concentration of enzymes is determined by the rates of synthesis and breakdown. It is known that temperature and photoperiod play a prominent role in controlling these processes. Temperature, for instance, can act through purely physical mechanisms, as when it influences viscosity and diffusion” (Schaber and Badeck, 2003). However, we understand that physiological impacts are very uncertain and the specific pathway is unclear. We have deleted both this sentence and the one in the discussion section for clarity.

L 24 p 6051 and on: I’m not a native English speaker, but I’ve always read calendar date to be reported as DoY XXX rather than XXX DoY

→ We have changed all the ‘XXX DOY’ to ‘DOY XXX’.

L15-17 P 6053: compare with altitudinal trends in Vitasse et al. 2009

→ We have added the following statement: “The temperature sensitivity of spring budburst date is also within the range from -2.05 to -7.48 days $^{\circ}\text{C}^{-1}$ for different species

based on the field experiments performed by Vitasse et al. (2009).” (Lines 505-507)

L9 p 6056: the question of the species-specificity of phenological model parameters is well established (see Vitasse et al., 2011 for instance). I would write “has been shown” instead of “is thought”.

→ We have changed “are thought” to “have been shown” as suggested.

L10 p 6056: a key missing reference reporting results for one of the datasets used in the paper is Archetti et al. (2013)

→ Yes. We have cited Archetti et al. (2013) in the revised manuscript.

L26 P 6058: The sentence beginning with “Missing. . .” should be rewritten, with appropriate places for parentheses and citations. Again, the notion of “synthesis” cited along the Schaber & Badeck (2003) paper, should be precised or deleted.

→ We have corrected the format errors and removed citation of Schaber and Badeck (2003) for clarity.

L 1 p 6059: the potential link between spring and autumn phenophases was demonstrated in the Fu et al. (2014b) paper. This citation should appear along with the Keenan & Richardson (2015) paper.

→ We have cited Fu et al. (2014) PNAS as suggested.

Section 3.1.2 and Figure 3: predicted dates are systematically biased (occur later than the ground observed dates). Probably caused by the distinct nature of the protocols used for obtaining calibration data vs. validation data. This should be investigated and reported in the paper.

→ We examined this issue in the revised manuscript: “However, the predicted budburst date at these sites is 5 days on average later than the observations. Causes of such bias are unclear but might be related to the uneven spatial distribution of network sites, the distinct nature of the protocols between calibration and validation data, and the incompatibility between model parameters derived at regional scale versus those for the continental scale.” (Lines 447-451) For autumn phenology, we added: “Similar to the spring budburst, predicted dormancy onset is later by 11 days than the observations.” (Line 456)

Figure 1: 5 sites are mentioned on the map, only 4 sites appear on the graphs. Clarify. Replace “literature-based phenology model” by “models S9 and A4” in the caption

→ We added the following figure caption to clarify: “Derived phenological dates at US-MOz are used for model calibration but not validation and are not shown here.” We

corrected the names of phenology model as suggested.

Figure 4: no autumn data appear on the figure, contrary to what is reported in the caption.

→ We have removed the words “and autumn model A4” to make sure that this figure is for validation of spring dates.

Figures 6 should be moved to the Suppl. Mat. Section.

→ The original Figure 6 has been moved to SI (now Fig. S14) as suggested.

Table 1: numeric values of the trends should be reported, and discussed in section 3.3

→ Reviewer #3 also proposed the similar comment. This is a good suggestion but might be difficult to implement for two reasons. First, regional phenological changes varied significantly across the U.S., while most of previous studies, if provided numeric values, show only numbers on the continental scale or for the latitude belts. For example, Jeong et al. (2011) reported almost zero changes in start of season over U.S. for 1982-1999. However, at the regional scale, this study showed both positive and negative trends in U.S. Second, the definition of phenological events varies among different studies (White et al., 2009), making it difficult to quantitatively perform the inter-comparison. As a result, we qualitatively compared phenological changes based on spatial maps from those studies, so as to estimate the ensemble spatial distribution of phenological changes in the past decades.

Table 2-3: Four sites appear, when sites are cited in the text (p 6042) for calibration. Clarify.

→ We clarified as follows: “Data from three out of the four calibration sites (US-Ha1, US-UMB, US-MMS, Table2 and Fig. 1) are also used for validation; however, we use them in different ways.”

Table S3: categorize spring and autumn parameters.

→ We categorized parameters for spring and autumn models as suggested.

Figure S8: “CDD-photoperiod model” has changed name to “literature-based” (compared to e.g. Fig. 9)

→ The figure title has been changed to “CDD-photoperiod model” as indicated.

SI, section 1 (“Derivation of phonological observations”): no mention is made to the autumn pheno dates. Are the LAI threshold used identical to those for spring?

→ Yes. We clarify in the SI as follows: “The same spring LAIt is applied to derive dormancy onset dates, which also show low RMSE against other sources of autumn phenology (Fig. S2b).”

Anonymous Referee #2

We thank the referee for the positive, helpful and comprehensive review that has improved the manuscript. Detailed point-by-point responses to the reviewer comments are provided below. The reviewers' comments are shown in black with our responses are marked as blue. The line numbers below refer to the revised manuscript to be submitted separately.

General comments

I have read the discussion paper "Probing the past 30-year phenology trend of U.S. deciduous forests" by Yue et al. with great interest.

In this paper, the authors test a suite of phenology models against site-level data in the US. Models represent spring budburst and autumn dormancy with various degrees of complexity as shown by differences in the required input data and the number of parameters. Models are first calibrated against phenology observations at four deciduous forests. Then, the selected best spring and autumn models are used to produce time series of phenology for the last 30 years over the whole conterminous US. The authors conclude to a temperature-driven advance in the spring budburst in the East and a delay in dormancy in the Northeast and West with large regional variations.

Phenology in ecosystem models is a topical question and this study falls well within the scope of the journal. The paper is well-written and easy to follow but I have several comments about the modelling procedure that needs to be clarified before publication (see below).

My main comment is that most of the spring phenology models have been derived from the same site, and calibrated against only 4 sites. How reliable are these models once applied to represent forests with varying species composition under very different climatic conditions? An assessment of the uncertainty in at least the models used to produce the regional maps is needed to check whether trends are a robust, or a result of over-fitted models.

→ We fully agree that the assessment of modeling uncertainty is important for the understanding of the robustness of our model predictions. In this study, we have quantified impacts of climate variability, model structure, and species aggregation on the predicted phenological trends. We also discussed the uncertainties from other possible sources, such as the incompleteness of observations, unrealistic representation of chilling/photoperiod limits, incompatibility of model parameters at site-level and continental scales, and so on.

We performed two additional sensitivity tests and added a new figure S18 following the reviewer's comment. "We analyze species-specific temperature sensitivity of tree phenology at Harvard Forest (section 4.3). Based on these results, we perform two additional sensitivity tests to evaluate modeling uncertainties from the intraspecific

variations. In the first run (simulation 4), phenological parameters are derived based on records of species with the lowest temperature sensitivity for both spring (Sweet Birch, *Betula Lenta*) and autumn (Paper Birch, *Betula Papyrifera*). In the other run (simulation 5), parameters are derived using records of species with the highest temperature sensitivity for spring (Striped Maple, *Acer Pensylvanicum*) and autumn (Black Oak, *Quercus Velutina*). We applied the derived parameters for the whole domain of U.S. by ignoring the realistic fractional coverage of specific species, so as to estimate the maximum uncertainty of prediction due to the intraspecific variations.” (Lines 374-384)

“We perform two sensitivity runs to evaluate the modeling uncertainties due to intraspecific variations on the continental scale (Fig. S18). Simulations with either the lowest (simulation 4) or the highest (simulation 5) temperature sensitivity yield very similar phenological trends as that in the control simulation (simulation 2). In the East, simulation 4 predicts a spring advance by 0.33 day yr^{-1} while simulation 5 predicts an advance by 0.35 day yr^{-1} , both of which are close to the 0.34 day yr^{-1} from the control run. In the West and Northeast, both sensitivity runs predict autumn delay by $0.13\text{-}0.15 \text{ day yr}^{-1}$, lower than value of $0.14\text{-}0.16 \text{ day yr}^{-1}$ from control run, suggesting that site-level responses may not be necessarily consistent with responses at the continental scale. Both the similar temperature sensitivity at site level (Fig. 8) and the predicted phenological trends at continental scale (Fig. S18) support the concept of phenological modeling at the forest and PFT level, and corroborates the further investigation of phenology-climate interactions at the continental and global scale.” (Lines 671-683)

Specific comments

P6042 L2: Please define "long-term"? Perhaps the term "decadal" would be more suited to describe the station data.

→ Yes. We have revised “long-term” to “decadal” as suggested.

P6043 L18: Would it be more correct to estimate budburst and dormancy based on significant changes in the LAI time series rather than a threshold?

→ It is a good suggestion to estimate phenological dates based on rapid changes in LAI. However, LAI records are not continuous, making it difficult to identify those phase changes. For example, measurements at US-UMB in 1999 started at DOY 121 with LAI of $1.5 \text{ m}^2 \text{ m}^{-2}$, which could be considered as the background value of LAI. However, in 2007, measurements began on DOY 132 with LAI of $1.93 \text{ m}^2 \text{ m}^{-2}$. For the year 2007, budburst date could be calculated only by extrapolating LAI to the day with value of $1.5 \text{ m}^2 \text{ m}^{-2}$.

P6045 L1: See my main comments. Also, why is Jolly et al.’s (2005) phenology model not used for spring?

→ We have responded to your main comments. All spring models used for the inter-

comparison considers chilling requirement while Jolly et al. (2005) does not implement such an effect.

P6046 L11: More detailed are needed to describe the calibration method. I am concerned that calibrating "by hand", as it appears, rather than using an automated tool does not allow to find the "true" optimum. Generally, using different model structures introduces more uncertainty than the equifinality of parameters of a single model. However, a few words about how well-defined the parameters are is required here.

→ In the text, we explained that: "For each model in Table 5, we apply the exhaustive enumeration method to evaluate all combinations of the discrete parameters. We select the optimized parameters that jointly predict the lowest RMSE for the long-term budburst dates at the four calibration sites." (Lines 279-282) We have shown the values of optimized parameters in Tables S3 and S4.

P6051 Section 3.1.2: Can you compare these results with satellite based estimates from Buitenwerf et al. (2015)?

→ Section 3.1.2 evaluates spatial distribution of simulated phenology while Buitenwerf et al. (2015) estimates changes in the standard deviation of phenological metrics. The two studies are investigating different aspects of phenology, making it difficult to perform the comparison directly.

P6052 L12: $p < 0.2$ is a very generous threshold

→ We have recalculated the correlations with a more stringent threshold of $p < 0.1$ in this revision. We also explained the implications of these new results as follows (Lines 463-472):

"The correlations between modeled and observed budburst dates are larger than 0.3 for 47 out of 59 sites, among which 26 are significant ($p < 0.1$), suggesting that the predicted interannual variation and long-term trend of spring phenology are generally reasonable on the continental scale. The insignificant correlations at 33 sites are in part attributed to the deviations in species between model (PFT level) and measurements (species level). Furthermore, the record length may also contribute to these biases because 28 out of the 33 sites with insignificant correlations have records shorter than 8 years. The large interannual variability in the spring phenology (see section 4.1) may affect the correlations especially for time series with short record length."

Figures Please add coordinates on Figures 3 to 7.

→ Coordinates have been added as suggested.

In Figure 3, why do the coloured surface appear to have different shapes in panels a and c, especially in the Southeast? Is it an artefact due to the colour scale that draws missing

areas and areas with lower values in white? More generally, why are the coloured areas in all maps different from the map in Figure 1?

→ Yes, the inconsistency between a and c is due to lower values in white. A full coverage of DBF has been shown in Figure 1. We have clarified in the caption of Figure 3 as follows: “The coverage of colored patches in (a) and (c) differs from that in Figure 1 because values at and beyond the low end of color scales have been shown in white.”

Anonymous Referee #3

We thank the referee for the positive, helpful and comprehensive review that has improved the manuscript. Detailed point-by-point responses to the reviewer comments are provided below. The reviewers' comments are shown in black with our responses are marked as blue. The line numbers below refer to the revised manuscript to be submitted separately.

General comments

Yue et al studied the changes of both spring and autumn phenology using different spring and autumn phenology models and focus on multiple scales, i.e. in situ and continent, and they found spatial difference in phenology trends over the period 1980-2012. Importantly, they concluded the temperature is the dominant driver of spring/autumn phenology, because phenology models including a chilling requirement or photoperiod limitation does not improve the model performance. The results are interesting, while there are some major comments in the model calibration and evaluation, as well as the explanation of the model results.

The phenology models are normally parametrized for specific species, i.e. difference specie holds different parameters, even at different sites for same specie ... the authors calibrated the model using LAI-based dates (species-mixed), and applied these models in the forest sites that probably have different species composition, as well as for the shrub i.e. Lilac. Uncertainty would be raised, therefore, the best model, which was selected for the continent scale prediction (while, actually the US scale), may be not the best, as the parameters might be not accurate. In additions, the models were calibrated using four site dates. How you calibrate these models, i.e. using mixed- dates from all four sites or using the yearly-average-dates across the 4 sites? Need to be clarify. . .

→ The uncertainty of model parameters is one of the main sources of uncertainty in phenological modeling. We include the following text to address this in the revised manuscript: “The validation shows that the predicted spatial pattern is reasonable and the long-term average matches observations within sampling uncertainty (Figs. 3-4). However, due to the data scarcity, all the selected sites are located in temperate areas ranging from 38°-46°N, suggesting that the model should be used cautiously at other latitudes and parameters may require re-calibration.” (Lines 710-714)

We clarified how to calibrate models as follows: “The measurements of LAI are not evenly distributed from year to year, and data at some years are too sparse to form the full annual cycle. As a result, we derive the decadal average phenological dates by regressing against all available LAI records at one site.” (Lines 153-156) “For each model in Table 5, we apply the exhaustive enumeration method to evaluate all combinations of the discrete parameters. We select the optimized parameters that jointly predict the lowest RMSE for the long-term budburst dates at the four calibration sites.” (Lines 279-282)

For the most important conclusion, i.e. the temperature is the dominant driver of spring

and autumn phenology, because the chilling and photoperiod models does not improve the model performance. It's is not completely right. (1) First, for spring phenology, the chilling requirement may be fulfilled, so the similar model performance can be expected, as reported by previous studies that compared different phenology models. So you could not evaluate the chilling effect from the model performance only...(2) Second, the similar model performance between one- and two- phase models may also suggest that the chilling / photoperiod mechanisms may not be accurately represented in these models, such as the chilling units that are counted as the chilling days or is a daily temperature function, while the day and night temperature may be play a different role in phenology (see Piao et al 2015 Leaf onset in the northern hemisphere triggered by daytime temperature). (3) Given the models are reliable, and the temperature might be a dominant driver for phenology in current climate, while with climate warming, the chilling and photoperiod may play an important role, might be dormant in the future. At least, these issues should be discussed in the manuscript. . .

→ We appreciate reviewer's thoughtful comments that reveal possible uncertainties in this study. In combination with comments from other reviewers, we added the following discussion in the revised manuscript:

“Our investigation of the roles of chilling and photoperiod is sensitive to the model structure, climate variability, and data availability. First, the similar performance between spring warming and chilling models might also result from the inaccurate representation of chilling / photoperiod mechanisms. For example, the chilling units used in our parameterization are calculated based on daily average temperatures, while Piao et al. (2015) suggested that leaf unfolding dates during 1982-2011 are triggered by daytime more than by nighttime temperature. The up-to-date autumn phenology model fails to capture interannual variability of dormancy onset (Fig. 2), suggesting that unknown processes are involved in the autumn leaf fall (Keenan and Richardson, 2015). It is unclear whether these processes are related to the variations of photoperiod. Second, the decadal changes in temperature may mask the role of chilling. The trend of winter warming is not significant for most areas in the U.S. (Fig. S14a), suggesting that chilling requirements have been fulfilled in the past 3 decades. However, it is unclear whether the winter warming will intensify in the future, which may slow the advancement of spring budburst. Third, we choose to calibrate the phenological parameterization at the level of plant functional type (PFT) because species-specific measurements are usually incomplete in time and uneven in space. Such incompleteness may influence the accuracy of derived decadal phenological records used for both model calibration and validation. At the same time, PFT-level parameterization may be too broad for the vegetation modeling because it fails to capture intraspecific variations (Van Bodegom et al., 2012; Reichstein et al., 2014). Observations at the community level suggest that the budburst of some species is sensitive to fall/winter and spring warming but with opposite signs (Cook et al., 2012). In the next subsection, we examine the records of 13 deciduous tree species at Harvard Forest.” (Lines 602-625)

One more comment, the authors compared the trends of phenology between the modeling and the RS results, but you could not identify the robust of the model results using the RS

data, because the two results are not comparable, i.e. the modeling outputs and the RS data may refer to different phenophase. The statement need to be improved.

→ We agree that the definition of phenological events with RS data is very uncertain. We added the following statement in the revised manuscript: “Since the definition of phenological events varies among different studies (White et al., 2009), we qualitatively compare the simulations with the remote sensing retrievals so as to evaluate the ensemble spatial distribution of phenological changes in the past decades.” (Lines 538-541)

Specific comments

P6039 L2 : Körner, not Korner, please check through the manuscript;
Körner and Basler discussed the importance of photoperiod on the spring phenology in their science paper. Better to rephrase it as ‘is sensitive to temperature variation’, and cite review papers, such as Cleland et al 2007; Polgar,C.A.,Primack,R.B.,2011.

→ We have corrected the author name of the reference to Körner. We have rephrased the sentence and added those review papers as suggested.

P6039 L10, for the phenology changes in Europe, you should also cite Menzel et al 2006 GCB

→ We have added the suggested reference in the revised manuscript.

P6039 L16, please rephrase as ‘some species may require cold temperatures’, the chilling may not only occur in winter, also in early spring. . .

→ We have removed the word “winter” as suggested.

P6039 L20, for the tree age, you can cite Vitasse et al 2013: Ontogenic changes rather than difference in temperature cause understory trees to leaf out earlier

→ Yes, we have cited the paper as suggested.

P6039 L25, what is the ‘temperature sensitivity to altitudinal trends’? Please rephrase..

→ We have removed this sentence because we do not discuss it in the paper.

P6041 L4-8, how you calibrate the models, using average date of all species from the four forests? No clear..

→ We rephrased it as: “We first calibrate each model using derived phenological dates based on the decadal ground observations of leaf area index (LAI) at four deciduous forests.” (Lines 115-116) In the discussion section 4.3, we showed that LAI-based phenological dates could be a good indicator of species-aggregated phenology.

P6041 L18, define chilling requirements

→ We have defined it in the introduction section: “Some species may also require cold temperatures before budburst (named chilling requirement)”

P6041 L22, ~1000? Why not provide the exact number of sites?

→ We changed it to the accurate site number of 1151 in the revised paper.

P6042 L7, because you determined the phenological dates from the LAI, better use ‘Start of growing season’ or ‘onset of growing season’, instead of ‘budburst’;

→ Our comparison showed that the ‘onset of growing season’ derived from LAI is very close to the ‘budburst’ from phenological records (Fig. S2). As a result, we continue to use ‘budburst’ for the LAI-derived spring phenology. In the revised text, we have the following definitions about spring budburst: “spring budburst date (or the onset of growing season, the dates D1 in Fig. S1).”

P6042 L6-10, D1, L1, D2 and L2, you should point these dates out in a Fig S?

→ We have shown D1, L1, D2, and L2 on the revised Fig. S1a.

P6042 L6-10, you determined the phenological dates for each year, and then calculated the average dates for each site, right? After that, you calibrated the model using only four date, i.e. four average dates from four sites? I’m not sure that, maybe need to rephrase to make it clear;

→ In the revised paper, we clarified as follows: “The measurements of LAI are not evenly distributed from year to year, and data at some years are too sparse to form the full annual cycle. As a result, we derive the decadal average phenological dates by regressing against all available LAI records at one site.” (Lines 153-156)

P6042 L14-18, this is results, move to the results section. . .

→ It is a good suggestion. However, as you may read, the result section presents model calibration and validation. The sentences you mentioned here are the introductions of sites and data, and as a result might be suitable for the method section.

P6043 L12-13, to evaluate the model, you used the averaged dates over all trees and species at each site. This means that the difference of average phenological dates among sites not only determined by the climatic variables, but also the composition of species. . . So, the questions is that can the parameters determined at four-sites be applied in other sites that have different species composition?

→ As we respond to the general comments, we tried the best to collect the most complete datasets (phenological records, LAI, and photos) to calibrate parameters and validate

model performance at sites and locations outside the calibration sites. “The validation shows that the predicted spatial pattern is reasonable and the long-term average matches observations within sampling uncertainty (Figs. 3-4). However, due to the data scarcity, all the selected sites are located in temperate areas ranging from 38°-46°N, suggesting that the model should be used cautiously at other latitudes and parameters may require re-calibration.” (Lines 710-714) We also discussed the uncertainties due to the differences in species composition in the section 4.3.

P6043 L21-25, define the budburst and dormancy when it first occurred. . .

→ We defined them in the revised manuscript: “Following definitions in earlier literatures (e.g., Zhu et al., 2012; Richardson et al., 2013), we validate spring budburst date (or the onset of growing season, the dates D1 in Fig. S1) and dormancy onset date (or the end of leaf fall period, the dates D2 plus falling length L2 in Fig. S1) from phenology models with the site-level records.” (Lines 173-177)

P6043 L22, how you define the rapid change from gray to light green? Using the maximum change rate? Difficult to understand the ‘the middle of the few days when tree colors change rapidly. . .’. Please rephrase, as well as the definition of Dormancy start ...

→ We explained it more clearly in the revised text: “Based on the changes of tree color in these photos, we can easily identify the phase changes in phenology. For example, changes from gray to light green in spring could occur within several days. We select the middle of these few days as the budburst date. Similarly, changes from brown to gray in late autumn may happen within one week and the middle day of the week is selected as the dormancy onset date. An example of autumn dormancy at US-UMB is shown in Fig. S3. The dates derived from photos may have comparable precision as the observations from site-level phenological records (e.g., Fig. S2c), because the latter are also reported weekly or half-weekly.” (Lines 197-205)

P6043 L23, the dormancy start normally is around the date when the bud set, i.e. much earlier than the leaf coloring. . .better use other terms, such as offset of growing season?

→ The ‘dormancy’ here represents the ending period of leaf fall. It is defined based on the canopy level instead of bud level. In the revised paper, we added following statement: “The dormancy onset date defined here is based on the canopy level instead of the bud dormancy examined in a recent review paper by Delpierre et al. (2015).” (Lines 177-179)

P6044 L22-26, similar model performance of 1- and 2- phase models suggest the chilling maybe sufficient over the study period. You can expect the chilling effect as the large scale, but I did not find the correlation between model performance and difference in phenological response among species, please rephrase ...

→ We have changed the phrase “for most species” to “at the site level”.

P6044 L28, The parameter values of the Sarvas function were determined from the experimental results on *Betula pendula* *Betula pubescens* and *Populus tremula* in Finland (Sarvas, 1972). Whether the ‘northern’ parameters can be used in the temperate trees? At least, you need mention it in the text, and discuss these issues.

→ In the text, we mentioned this problem as follows: “We apply the same fixed thresholds (e.g., 3.4 and 10.4) for equations (6)-(7) as that in Chuine et al. (1999); however, we re-calibrate other parameters (e.g. T_c and C^*) so that these functions adapt to the phenological changes in U.S. deciduous forest.” (Lines 268-271)

We discussed this issue as follows: “However, due to the data scarcity, all the selected sites are located in temperate areas ranging from 38°-46°N, suggesting that the model should be used cautiously at other latitudes and parameters may require re-calibration.” (Lines 712-714)

P6046 L20, provide reference for the statement: temperature and photoperiod affects the autumn phenology.

→ We have cited Delpierre et al. (2009) and Richardson et al. (2013) in the revised paper.

P6050 L5-13, the figure S3-S9, no RMSE, AIC and correlation coefficient values, better to show these values in a table S.

→ We have added a Table S5 to summarize those numbers.

P6050 L14-15, no model could predict the autumn phenology, i.e. correlation smaller than 0.5, how you conclude the temperature dominant the autumn phenology process?

→ We conclude that temperature dominates autumn phenology because simulations with and without photoperiod limit show similar results. However, as we discussed (refer to the responses to the general comments), this conclusion is model-dependent and is uncertain.

P6051 L18, you studied the country-scale, i.e. USA, no on the continental scale..

→ The scale of USA is large enough to be considered as continental scale. Similar definition could also seen in previous studies. For example, Schwartz and Hanes (2010) who investigated phenology for western U.S.

P6052 L11 the model-observation correlation -> the correlations between modeled and observed budburst dates. . .

→ Changed as suggested.

P6053 3.3 section, you should discuss the difference between the remote sensing based

phenology and the model results. . .

→ In section 3.3, we added a new remote sensing study from Buitenwerf et al. (2015) for comparison. The updated Figure 7 shows distributions of phenological changes even closer to our prediction (Figure 5) than the original results, further supporting our estimate of the phenological trends in U.S. using the well-calibrated model. We discussed differences between RS and simulations as follows:

“For U.S. forests, the longer growing season is primarily driven by later leaf-off dates, though regional advance of spring is also observed. Our results are generally consistent with their conclusions but with some deviations. For example, they observed later autumn in almost all the eastern U.S., where we predict delays only in the North and Northeast (Fig. 5b). Such discrepancies reflect prediction biases, and may also be a consequence of satellite retrieval uncertainties (Table 1)” (Lines 525-530)

“Our results show similar changes in spring phenology as the ensemble of the remote sensing studies, except that we predict smaller delays in the northern states (Fig. 5a). In addition, our data-informed model simulates significant spring advances in the central U.S., while remote sensing studies largely disagree over this area.” (Lines 544-548)

P6055 3.4.2 section, please see the general comments.

→ We have carefully discussed the uncertainties in our conclusions about the impact of chilling and photoperiod.

P6057, the conclusion and discussion, it's kind of a 'Conclusion'... in the results section, you have a lot discussion, I would suggest to make a separate discussion section. . .

→ The similar comment has been proposed by the Reviewer #1 and we have changed “Conclusions and discussion” to “Conclusions” and added a separate “Discussion” section.

Table 1, the results, better providing the values of changes if you can, such as days per year or decade.

→ Reviewer #1 also proposed the similar comment. This is a good suggestion but might be difficult to implement for two reasons. First, regional phenological changes varied significantly across the U.S., while most of previous studies, if provided numeric values, show only numbers on the continental scale or for the latitude belts. For example, Jeong et al. (2011) reported almost zero changes in start of season over U.S. for 1982-1999. However, at the regional scale, this study showed both positive and negative trends in U.S. Second, the definition of phenological events varies among different studies (White et al., 2009) making it difficult to quantitatively perform the inter-comparison. As a result, we qualitatively compared phenological changes based on spatial maps from those

studies, so as to estimate the ensemble spatial distribution of phenological changes in the past decades.

Table 2, define the ‘n’ in the legend;

→ Defined it in a table footnote.

Table 3, the phenology dates are the average dates, right? Provide the std . .

→ The regressed (now has changed the name to “LAI-derived” in Table 3) dates are based on one regression against all available records. As a result, there are no year-to-year values and standard deviations are not available. The modeled dates are based on interannually varied meteorology and the standard deviations have been shown in the revised Table 3.

Reference

- Buitenwerf, R., Rose, L., and Higgins, S. I.: Three decades of multi-dimensional change in global leaf phenology, *Nat Clim Change*, 5, 364-368, 2015.
- 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, doi:10.1046/J.1365-3040.1999.00395.X, 1999.
- Cook, B. I., Wolkovich, E. M., and Parmesan, C.: Divergent responses to spring and winter warming drive community level flowering trends, *P Natl Acad Sci USA*, 109, 9000-9005, doi:10.1073/Pnas.1118364109, 2012.
- Delpierre, N., Dufrene, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J., and Francois, C.: Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France, *Agr Forest Meteorol*, 149, 938-948, doi:10.1016/J.Agrformet.2008.11.014, 2009.
- Delpierre, N., Vitasse, Y., Chuine, I., Guillemot, J., Bazot, S., Rutishauser, T., and Rathgeber, C. B. K.: Temperate and boreal forest tree phenology: from organ-scale processes to terrestrial ecosystem models, *Annals of Forest Science*, doi:10.1007/s13595-015-0477-6, 2015.
- 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, *Global Change Biol*, 17, 2385-2399, doi:10.1111/J.1365-2486.2011.02397.X, 2011.
- Jolly, W. M., Nemani, R., and Running, S. W.: A generalized, bioclimatic index to predict foliar phenology in response to climate, *Global Change Biol*, 11, 619-632, doi:10.1111/J.1365-2486.2005.00930.X, 2005.
- Keenan, T. F., and Richardson, A. D.: The timing of autumn senescence is affected by the time of spring phenology: implications for predictive models, *Global Change Biol*, 21, 2634-2641, doi:10.1111/gcb.12890, 2015.
- Piao, S. L., Tan, J. G., Chen, A. P., Fu, Y. H., Ciais, P., Liu, Q., Janssens, I. A., Vicca, S., Zeng, Z. Z., Jeong, S. J., Li, Y., Myneni, R. B., Peng, S. S., Shen, M. G., and Pennuelas, J.: Leaf onset in the northern hemisphere triggered by daytime temperature, *Nat Commun*, 6, 6911, doi:10.1038/Ncomms7911, 2015.
- Prentice, I. C., Liang, X., Medlyn, B. E., and Wang, Y.-P.: Reliable, robust and realistic: the three R's of next-generation land-surface modelling, *Atmos. Chem. Phys.*, 15, 5987-6005, doi:10.5194/acp-15-5987-2015, 2015.
- Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking plant and ecosystem functional biogeography, *P Natl Acad Sci USA*, 111, 13697-13702, doi:10.1073/Pnas.1216065111, 2014.
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M.: Climate change, phenology, and phenological control of vegetation feedbacks to the climate system, *Agr Forest Meteorol*, 169, 156-173, 2013.
- Schaber, J., and Badeck, F. W.: Physiology-based phenology models for forest tree species in Germany, *Int J Biometeorol*, 47, 193-201, doi:10.1007/S00484-003-0171-5, 2003.

- Schwartz, M. D., and Hanes, J. M.: Continental-scale phenology: warming and chilling, *Int J Climatol*, 30, 1595-1598, doi:10.1002/Joc.2014, 2010.
- Van Bodegom, P. M., Douma, J. C., Witte, J. P. M., Ordonez, J. C., Bartholomeus, R. P., and Aerts, R.: Going beyond limitations of plant functional types when predicting global ecosystem-atmosphere fluxes: exploring the merits of traits-based approaches, *Global Ecol Biogeogr*, 21, 625-636, doi:10.1111/J.1466-8238.2011.00717.X, 2012.
- Vitasse, Y., Delzon, S., Dufrene, E., Pontailier, J. Y., Louvet, J. M., Kremer, A., and Michalet, R.: Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses?, *Agr Forest Meteorol*, 149, 735-744, doi:10.1016/J.Agrformet.2008.10.019, 2009.
- White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., O'Keefe, J., Zhang, G., Nemani, R. R., van Leeuwen, W. J. D., Brown, J. F., de Wit, A., Schaepman, M., Lin, X. M., Dettinger, M., Bailey, A. S., Kimball, J., Schwartz, M. D., Baldocchi, D. D., Lee, J. T., and Lauenroth, W. K.: Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006, *Global Change Biol*, 15, 2335-2359, doi:10.1111/J.1365-2486.2009.01910.X, 2009.
- Zhu, W. Q., Tian, H. Q., Xu, X. F., Pan, Y. Z., Chen, G. S., and Lin, W. P.: Extension of the growing season due to delayed autumn over mid and high latitudes in North America during 1982-2006, *Global Ecol Biogeogr*, 21, 260-271, doi:10.1111/J.1466-8238.2011.00675.X, 2012.

Probing the past 30-year phenology trend of U.S. deciduous forests

X. Yue ¹, N. Unger ¹, T. F. Keenan ², X. Zhang ³, C. S. Vogel ⁴

¹ School of Forestry and Environmental Studies, Yale University, 195 Prospect Street,
New Haven, CT 06511, USA

² Department of Biological Sciences, Macquarie University, Sydney, New South Wales
2109, Australia

³ Geospatial Sciences Center of Excellence, South Dakota State University, Brookings,
SD 57007, USA

⁴ The University of Michigan Biological Station, Pellston, MI 49769, USA

Correspondence to: X. Yue (xuyueseas@gmail.com)

Abstract

Phenology is experiencing dramatic changes over deciduous forests in the U.S. Estimates of trends in phenology on the continental scale are uncertain, however, with studies failing to agree on both the magnitude and spatial distribution of trends in spring and autumn. This is due to the sparsity of *in situ* records, uncertainties associated with remote sensing data, and the regional focus of many studies. It has been suggested that reported trends are a result of recent temperature changes, though multiple processes are thought to be involved and the nature of the temperature forcing remains unknown. To date, no study has directly attributed long-term phenological trends to individual forcings across the U.S. through integrating observations with models. Here, we construct an extensive database of ground measurements of phenological events across the U.S., and use it to calibrate and evaluate a suite of phenology models. The models use variations of the accumulative temperature summation, with additional chilling requirements for spring phenology and photoperiod limitation for autumn. Including a chilling requirement or photoperiod limitation does not improve model performance, suggesting that temperature change, especially in spring and autumn, is [likely](#) the dominant driver of the observed trend during the past 3 decades. Our results show that phenological trends are not uniform over the contiguous U.S., with a significant advance of 0.34 day yr^{-1} for the spring budburst in the East, a delay of 0.15 day yr^{-1} for the autumn dormancy onset in the Northeast and West, but no evidence of change elsewhere. Relative to the 1980s, the growing season in the 2000s is extended by about 1 week (3-4%) in the East, New England, and the upper Rocky Mountains forests. [Additional sensitivity tests show that intraspecific variations may not influence the predicted phenological trends.](#) These results help reconcile conflicting reports of phenological trends in the literature, and directly attribute observed trends to long-term changes in temperature.

Keywords: phenology, trend, interannual variability, deciduous forest, chilling requirement, growing season

1 Introduction

Plant phenology, such as the timing of spring budburst and autumn leaf fall, is sensitive to [temperature variation](#) ([Körner and Basler, 2010](#); [Polgar and Primack, 2011](#); Richardson et al., 2013) and is thus exhibiting a long-term trend with the changing climate ([Badeck et al., 2004](#); [Cleland et al., 2007](#); Gordo and Sanz, 2009; Jeong et al., 2011). Long-term changes in phenology may be affecting ecosystem carbon assimilation (Keenan et al., 2014), surface water and energy balance (Schwartz and Crawford, 2001), and forest composition and evolution (Forrest and Miller-Rushing, 2010). Emerging observations have shown advanced spring and delayed autumn over the Northern Hemisphere, especially in Europe, during the past several decades (Menzel and Fabian, 1999; Fitter and Fitter, 2002; [Menzel et al., 2006](#); Gordo and Sanz, 2009). However, the extent of regional phenological trends in U.S. remains uncertain as different studies present inconsistent and even opposite results (Table 1).

The uncertainty of the phenological changes in U.S. forests could be attributed to genetic, geographic, and temporal factors. First, experiments have suggested that different species may have different phenological sensitivity to temperature (Vitasse et al., 2009). Some species may also require cold [temperatures before budburst](#) ([named chilling requirement](#)), leading to divergent responses of U.S. plants to spring and winter warming at the community level (Cook et al., 2012) and the continental scale (Zhang et al., 2007). In addition, it is not clear whether other biotic and/or abiotic factors (e.g. humidity, photoperiod, tree age, and tree species) may play a role (Morin et al., 2009; Basler and Korner, 2012; [Vitasse, 2013](#); Caldararu et al., 2014; Laube et al., 2014). Second, most deciduous forests in the U.S. are found at mid-latitudes, where temperature increases have not been uniform, and are not as strong as those at high latitudes (Hartmann et al., 2013). [Third](#), differences in the time frames used in different studies may lead to apparently inconsistent trends (Badeck et al., 2004).

There are generally three approaches for estimating phenology at regional and continental scales: ground networks, remote sensing, and numerical modeling. Ground-based

Xu Yue 7/2/15 11:18 AM

Deleted: climatic conditions

Xu Yue 7/2/15 11:18 AM

Deleted: Korner

Xu Yue 7/2/15 11:18 AM

Deleted: winter

Xu Yue 7/2/15 11:18 AM

Deleted: ,

Xu Yue 7/2/15 11:18 AM

Deleted: The high elevation in the western U.S. also induces additional complexity for the phenological responses, as temperature sensitivity to altitudinal trends also differs among species (Vitasse et al., 2009).

90 measurements can provide the most accurate phenological dates, such as budburst,
91 flowering, and leaf fall. Some records last for decades and even centuries (Sparks and
92 Menzel, 2002), making it possible to study long-term phenological change. However,
93 such measurements usually have very limited spatial coverage. Ground-based networks,
94 such as North American Lilac Network (Schwartz and Reiter, 2000), improve the spatial
95 coverage but focuses only on 1-2 species, which may not represent the average
96 phenological status of local plants. More extensive networks, such as the North American
97 Phenology Network (www.usanpn.org) or the European Phenology Network
98 (www.pep725.eu), contain many more species but typically do not have long data records
99 (with some exceptions). Remote sensing provides a way to examine phenological
100 changes over large scales but is inherently limited by short time scales or infrequent
101 retrieval times and must be validated using ground measurements. Most of the recent
102 estimates of phenological changes on the continental scale are performed using satellite
103 retrievals (e.g., Zhang et al., 2007; Jeong et al., 2011; Piao et al., 2011). The lack of a
104 universally accepted definition of phenological status for this method may lead to
105 discrepancies up to 60 days for the timing of events among different algorithms and
106 products (White et al., 2009). Moreover, date retrieval is often hampered, e.g., by cloud
107 cover, which can lead to poor correlations with ground observations (Badeck et al., 2004;
108 Schwartz and Hanes, 2010).

109

110 Phenological models are useful tools for diagnosing causes of phenological changes and
111 also for understanding the feedback of those changes to the Earth system (Richardson et
112 al., 2013; Zhao et al., 2013). Evaluations of well-calibrated phenological models have
113 shown high correlations between predictions and observations (e.g., White et al., 1997;
114 Richardson et al., 2006; Delpierre et al., 2009; Vitasse et al., 2011). However, most of
115 these state-of-art schemes are not evaluated at continental or even larger scales, thus
116 limiting their applicability in dynamic vegetation models and climate models. Recent
117 model-data comparisons have shown that the bias in the prediction of vegetation
118 phenology is a large source of uncertainty in models of ecosystem carbon uptake
119 (Richardson et al., 2012). This necessitates the development and evaluation of continental
120 scale phenology models with continental scale observations.

Xu Yue 7/2/15 11:18 AM

Deleted: of course

122

123 In this study, we use an extensive dataset of phenological observations to calibrate and
124 evaluate 13 models (9 for spring and 4 for autumn) of deciduous tree phenology across
125 the U.S. We first calibrate each model using [derived phenological dates based on the](#)
126 [long-term ground observations of leaf area index \(LAI\)](#) at four deciduous forests. We
127 then examine modeled interannual variability and trends, along with regional
128 phenological differences, using an extensive network of phenological observations. The
129 phenology model best supported by the observations is then applied to: (1) estimate the
130 trend of both spring and autumn phenology of U.S. deciduous forests over the last three
131 decades; (2) compare our results with other approaches (ground network, remote sensing,
132 and model based) to identify robust changes and assess discrepancies; and (3) examine
133 the underlying drivers of both the observed trends and interannual variability.

134

135

136 2 Materials and methods

137

138 We assembled and compared a suite of published models of spring and autumn
139 phenology. Most of these models are built using cumulative thermal summations with
140 constraining processes, such as chilling requirements and photoperiod limits. Model
141 parameters were calibrated using long-term observations at four deciduous forest sites,
142 with some model constants estimated based on literature values. An independent dataset
143 of ground measurements was compiled and used to validate the performance of these
144 models. In total, phenological observations from [1151](#) sites were used [for model](#)
145 [validation](#). In this section we first present the observations used for calibration and
146 validation, followed by a description of the various model formulations tested and
147 simulations performed.

148

149 2.1 Ground measurements for calibration

150

151 [Decadal](#) measurements of [LAI](#) from four U.S. deciduous broadleaf sites are collected
152 from the Ameriflux network (<http://ameriflux.ornl.gov/>) to calibrate parameters of the

Xu Yue 7/2/15 11:18 AM

Deleted: phenology

Xu Yue 7/2/15 11:18 AM

Deleted: ~1000

Xu Yue 7/2/15 11:18 AM

Deleted: Long-term

Xu Yue 7/2/15 11:18 AM

Deleted: leaf area index (

Xu Yue 7/2/15 11:18 AM

Deleted:)

phenology model (Table 2 and Fig. 1). We derive annual cycles of phenology by normalizing individual LAI values to the maximum and minimum LAI in each year for each site (Fig. S1, top panel). Since the measurements are discrete, we estimate the budburst dates (D1, marked on Fig. S1a), growing length (L1), offset start dates (D2), and falling length (L2) based on segmented regressions, which yield the minimum root mean square error (RMSE) against observations (Fig. S1 middle and bottom panel):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (1)$$

where O_i is the observation, P_i is the regression or prediction, and n is the number of samples. The measurements of LAI are not evenly distributed from year to year, and data at some years are too sparse to form the full annual cycle. As a result, we derive the decadal average phenological dates by regressing against all available LAI records at one site. The derived phenological dates are presented in Table 3. The average budburst date at US-MMS and US-MOz is earlier by three weeks than that at US-Ha1 and US-UMB, probably because the former sites are ~5 °C warmer than the latter. However, the start of leaf senescence is similar at all four sites, suggesting that photoperiod may also play an important role in regulating the autumn phenology, especially at the two warmer sites.

2.2 Ground measurements for validation

We use >75000 records for deciduous trees to evaluate the temporal variation and spatial distribution of simulated phenology (Table 4 and Table S1-S2). Data from three out of the four calibration sites (US-Ha1, US-UMB, US-MMS, Table 2 and Fig. 1) are also used for validation; however, we use them in different ways. For calibration, we use the decadal average phenology derived from the multiple-year LAI measurements, so that every calibrated model can capture the spatial pattern of phenology events on the continental scale. For validation, we use year-to-year phenological dates estimated from date records, photos, and LAI at each year, so as to identify the model that best captures the temporal variations. Most of the phenological records are discrete and evaluation of

Xu Yue 7/2/15 11:18 AM

Deleted: long-term average

Xu Yue 7/2/15 11:18 AM

Deleted: Although data at some

Xu Yue 7/2/15 11:18 AM

Deleted: , such as

Xu Yue 7/2/15 11:18 AM

Deleted: and

Xu Yue 7/2/15 11:18 AM

Deleted: (Table 2),

Xu Yue 7/2/15 11:18 AM

Deleted: calibration

Xu Yue 7/2/15 11:18 AM

Deleted: long-term

the annual cycle of tree phenology is difficult. Following definitions in earlier literatures (e.g., Zhu et al., 2012; Richardson et al., 2013), we validate spring budburst date (or the onset of growing season, the dates D1 in Fig. S1) and dormancy onset date (or the end of leaf fall period, the dates D2 plus falling length L2 in Fig. S1) from phenology models with the site-level records. The dormancy onset date defined here is based on the canopy level instead of the bud dormancy examined in a recent review paper by Delpierre et al. (2015).

The two New England sites, Harvard Forest (<http://harvardforest.fas.harvard.edu/>) and Hubbard Brook (US-HB1) Forest (<http://www.hubbardbrook.org/>), have decadal measurements back to 1990. The full records at Harvard Forest include 34 species, 16 of which are deciduous trees. The forest within the tower footprint is dominated by red oak (*Quercus Rubra*, 60% basal area), red maple (*Acer Rubrum*, 23% basal area), and secondary deciduous species. Hubbard Brook has three species, namely sugar maple (*Acer Saccharum*), American beech (*Fagus Grandifolia*), and yellow birch (*Betula Alleghaniensis*). We average over all trees and species at each site to generate average phenological dates for each year. Phenological observations are incomplete at two of the Ameriflux sites, US-UMB and US-MMS. We derive the missing phenological dates based on LAI data from Ameriflux and images from the PhenoCam project (<http://phenocam.sr.unh.edu/webcam/>). If the year-round LAI data are available at one site, we estimate budburst and dormancy start dates as the days when the interpolated or extrapolated LAI is equal to a selected threshold (see Supplement). Otherwise, we qualitatively estimate phenological dates based on photos from PhenoCam, which is a near-surface remote sensing network that observes phenology changes with high-resolution digital cameras (Sonnentag et al., 2012). Based on the changes of tree color in these photos, we can easily identify the phase changes in phenology. For example, changes from gray to light green in spring could occur within several days. We select the middle of these few days as the budburst date. Similarly, changes from brown to gray in late autumn may happen within one week and the middle day of the week is selected as the dormancy onset date. An example of autumn dormancy at US-UMB is shown in Fig. S3. The dates derived from photos may have comparable precision as the observations

Xu Yue 7/2/15 11:18 AM

Deleted: long-term

Xu Yue 7/2/15 11:18 AM

Deleted: descriptions in the Supporting Information).

Xu Yue 7/2/15 11:18 AM

Deleted: We define

Xu Yue 7/2/15 11:18 AM

Deleted: budburst date as the middle

Xu Yue 7/2/15 11:18 AM

Deleted: the few days when

Xu Yue 7/2/15 11:18 AM

Deleted: colors change rapidly

Xu Yue 7/2/15 11:18 AM

Deleted: . On the contrary, a dormancy start date is defined as the

Xu Yue 7/2/15 11:18 AM

Deleted: days with rapid color

Xu Yue 7/2/15 11:18 AM

Deleted: .

from site-level phenological records (e.g., Fig. S2c), because the latter are also reported weekly or half-weekly.

Data from ground networks was used to evaluate the model performance on the continental scale. The USA National Phenology Network (USA-NPN) is a nationwide project collecting standardized ground phenology observations by researchers, students, and volunteers. The network has limited records before 2009 but is significantly enriched thereafter. We select observations during 2011-2012 for 52 deciduous tree species that are most common in the U.S. (Table S1). The derived phenological dates for individual trees are averaged if they are observed at the same location (see Supplement). We also used observations from The North American Lilac Network (NALN), which provides records of the first leaf and first bloom dates of two lilac species, Common Lilac (*Syringa Vulgaris*) and Red Rothomagensis lilac (*Syringa Chinensis*), for the period of 1956-2003 (Schwartz and Reiter, 2000). As we shown in section 4.3, the phenology of individual species may vary by up to 3 weeks, however, the responses of phenology to temperature changes are relatively similar across species. We calculate correlations of budburst dates between observations and simulations at the available sites of NALN to validate the simulated temporal variations of phenology. We also adopt the limited long-term records from USA-NPN (Table S2) to evaluate the model over regions not covered by NALN.

2.3 Spring phenology models

Dozens of spring phenology models have been evaluated and inter-compared in the past two decades (Chuine et al., 1999; Linkosalo et al., 2008; Vitasse et al., 2011; Fu et al., 2012a; Fu et al., 2012b; Migliavacca et al., 2012; Melaas et al., 2013). These models may have different formats and parameters, but are generally dependent on temperature and photoperiod and could be divided into two categories, spring warming (or 1-phase) and chilling (or 2-phase), based on their assumptions of how warm and cold temperatures control the phenology development (Migliavacca et al., 2012). Although regional studies have demonstrated that the 1-phase models are as efficient as 2-phase models at the site level (e.g., Vitasse et al., 2011; Fu et al., 2012a; Migliavacca et al., 2012), we consider

Xu Yue 7/2/15 11:18 AM

Deleted: or site

Xu Yue 7/2/15 11:18 AM

Deleted: 3.

Xu Yue 7/2/15 11:18 AM

Deleted: for most species

that chilling requirement may be necessary for the phenology at the continental and global scales where divergent phenological responses are observed (Zhang et al., 2007; Cook et al., 2012).

The chilling models have different formulations based on the sequences (sequential, parallel, or alternating) and forms (thermal summation or the Sarvas function) of chilling and forcing (Chuine et al., 1999). According to these differences, Migliavacca et al. (2012) summarized and compared eight models, listed as S1-S8 in Table 5, to fit phenology data at Harvard Forest. The sequential models require that a chilling threshold (C^*) must be achieved before the forcing (S_f) is effective. The parallel and alternating models calculate chilling units (S_c) and S_f at the same time, however, the increases in S_c can reduce the budburst threshold (F^*) for S_f following an exponential relationship $F^* = a \exp(b \times S_c)$. The functions of S_c and S_f are calculated as the cumulative thermal unit as follows:

$$S_c(t) = \sum_{t_1}^t R_c(x_t) \quad (2)$$

$$S_f(t) = \sum_{t_2}^t R_f(x_t) \quad (3)$$

where x_t is the daily temperature. The thermal unit may have two different formats. In the thermal summation approach (CF1, Equations 4 and 5), S_c is the number of chilling days ($< T_c$) from a starting day t_1 and S_f is the cumulative temperature higher than T_f (commonly named growing degree day, GDD) from day t_2 . In the other approach (CF2, Equations 6 and 7), both R_c and R_f are functions of daily temperature (Chuine et al., 1999).

$$CF1: R_c(x_t) = \begin{cases} 0, & x_t \geq T_c \\ 1, & x_t < T_c \end{cases} \quad (4)$$

Xu Yue 7/2/15 11:18 AM

Deleted: much

$$CF1: R_f(x_t) = \begin{cases} x_t - T_f, & x_t \geq T_f \\ 0, & x_t < T_f \end{cases} \quad (5)$$

299

$$CF2: R_c(x_t) = \begin{cases} 0, & x_t \leq -3.4 \text{ or } x_t \geq 10.4 \\ \frac{x_t + 3.4}{T_c + 3.4}, & -3.4 < x_t < T_c \\ \frac{x_t - 10.4}{T_c - 10.4}, & T_c < x_t < 10.4 \end{cases} \quad (6)$$

301

$$CF2: R_f(x_t) = \begin{cases} 0, & x_t \leq 0 \\ \frac{28.4}{1 + e^{-0.185(x_t - 18.4)}}, & x_t > 0 \end{cases} \quad (7)$$

303

304 For both parallel and alternating models, t_1 is equal to t_2 , and for the latter, T_c is equal to
 305 T_f . For sequential models, t_2 is the first day when $S_c > C^*$. [We apply the same fixed](#)
 306 [thresholds \(e.g., 3.4 and 10.4\) for equations \(6\)-\(7\) as that in Chuine et al. \(1999\);](#)
 307 [however, we re-calibrate other parameters \(e.g. \$T_c\$ and \$C^*\$ \) so that these functions adapt to](#)
 308 [the phenological changes in U.S. deciduous forest.](#)

309

310 In a modified alternating scheme (S9), we decrease model complexity by fixing some
 311 parameters based on literature values. First, we fix t_1 as the winter solstice (December
 312 22nd in Northern Hemisphere, NH), after which photoperiod increases gradually. Second,
 313 we set T_c to 5°C, a value widely used for woody species (Murray et al., 1989; Kaduk and
 314 Heimann, 1996; Sitch et al., 2003). Third, we redefine the format of [the](#) forcing threshold
 315 as $F^* = a + b \exp(r \times S_c)$ following Murray et al. (1989) and set $r = -0.01$, a value used for
 316 temperate trees (e.g. beech and black locust). For each model in Table 5, we [apply the](#)
 317 [exhaustive enumeration method to evaluate](#) all [combinations of the discrete](#) parameters.
 318 [We](#) select the optimized parameters that jointly predict the lowest RMSE for the long-
 319 term budburst dates at the four calibration sites.

320

Xu Yue 7/2/15 11:18 AM

Deleted: adjust

Xu Yue 7/2/15 11:18 AM

Deleted: unfixed

Xu Yue 7/2/15 11:18 AM

Deleted: step by step and

324 We assume the green up process is linearly dependent on forcing S_f as follows,

325

326
$$f_s = \begin{cases} 0, & S_f < F^* \\ \frac{S_f - F^*}{L_g}, & F^* \leq S_f \leq F^* + L_g \\ 1, & S_f > F^* + L_g \end{cases} \quad (8)$$

327

328 where f_s is [spring](#) phenology ranging from 0 to 1. The parameter L_g is a growing length
329 constraint calibrated based on the cycle of forest phenology (Fig. S1).

330

331 2.4 Autumn phenology models

332

333 Autumn phenology is more uncertain than budburst because it is affected by both
334 temperature and photoperiod, (Delpierre et al., 2009; Richardson et al., 2013). Three
335 models have been developed to predict leaf fall with constraint from temperature and
336 photoperiod, namely the continental phenology model by White et al. (1997), the
337 growing season index (GSI) by Jolly et al. (2005) and the cold-degree day photoperiod-
338 dependent model by Delpierre et al. (2009). The White et al. (1997) scheme is not
339 compared in this study as it depends on soil temperature, which is not available at some
340 sites. Jolly et al. (2005) calculated global phenology as the product of three segmented
341 functions, which depend on the upper and lower limits in temperature (T_x and T_i), vapor
342 pressure deficit (VPD) (V_x and V_i), and photoperiod (P_x and P_i), respectively. The value
343 of VPD function is set to constant of 1 for temperate forests with no water stress.
344 Delpierre et al. (2009) calculated the cumulative products of the functions of temperature
345 and photoperiod. Those functions may have power indexes ranging from 0 to 2,
346 suggesting that autumn phenology could be unrelated, linearly related, or exponentially
347 related with the constraints from temperature and photoperiod. We calibrate all model
348 parameters based on the observations at U.S. deciduous forests (A2-A3 in Table 5). We
349 also use the original parameters from Jolly et al. (2005), which have been validated based
350 on remote sensing data on the global scale (A1 in Table 5).

Xu Yue 7/2/15 11:18 AM

Deleted: $f_T = \begin{cases} 0, & S_f < l \\ \frac{S_f - F^*}{L_g}, & F^* \leq S_f \leq \\ 1, & S_f > F^* \end{cases}$

Unknown

Field Code Changed

Xu Yue 7/2/15 11:18 AM

Deleted: f_T

Xu Yue 7/2/15 11:18 AM

Deleted: a temperature-dependent

Xu Yue 7/2/15 11:18 AM

Deleted: .

355

356 We also construct a simple scheme based on cumulative cold degree-days. The scheme,
 357 named ‘CDD-photoperiod’ (A4 in Table 5), calculates cold degree days (CDD) C_a
 358 following Richardson et al. (2006):

359

$$360 \quad C_a(t) = \sum_{t_3}^t R_a(x_t) \quad (9)$$

361

$$362 \quad R_a(x_t) = \begin{cases} T_b - x_t, & x_t < T_b \\ 0, & x_t \geq T_b \end{cases} \quad (10)$$

363

364 where t_3 is the starting day set to summer solstice (June 22nd in NH), and T_b is a base
 365 temperature of 20 °C as that in Dufrene et al. (2005) and Richardson et al. (2006). The
 366 leaf fall is triggered if C_a is higher than a threshold F_s and the length of falling period is
 367 determined by L_f as follows,

368

$$369 \quad f_T = \begin{cases} 0, & C_a \leq F_s \\ \frac{C_a - F_s}{L_f}, & F_s < C_a < F_s + L_f \\ 1, & C_a \geq F_s + L_f \end{cases} \quad (11)$$

370

371 here f_T is the temperature-dependent phenology ranging from 0 to 1. We also define a
 372 photoperiod-limited phenology following Jolly et al. (2005),

373

$$374 \quad f_P = \begin{cases} 0, & P \leq P_i \\ \frac{P - P_i}{P_x - P_i}, & P_i < P < P_x \\ 1, & P \geq P_x \end{cases} \quad (12)$$

375

where P is the daylength in minutes. P_i and P_x are the lower and upper limits of daylength during the period of leaf fall. Following Jolly et al. (2005), the final autumn phenology f_A of deciduous forest is determined as the product of f_T (Equation 11) and f_P (Equation 12).

2.5 Simulations

We perform both site-level and continental-scale simulations. For standalone simulations (simulation 1), phenology models are driven with daily surface air temperature sampled at each site (<http://ameriflux.ornl.gov/>). We gap-filled *in situ* temperature with daily reanalysis data from the Modern-Era Retrospective Analysis for Research and Applications (MERRA, Reichle et al., 2011; Rienecker et al., 2011), which is interpolated to each site based on the site location. The time span of each site-level simulation varies depending on the availability of the phenology observations. We perform a model inter-comparison to determine which model is most supported by observations. The statistical metrics we used for evaluations include correlations, RMSE (Equation 1) and the Akaike Information Criterion (AIC), a measure of the trade-off between model predictability and model complexity (Akaike, 1973; Burnham and Anderson, 2002),

$$AIC = n \cdot \log \sigma^2 + 2p + \frac{2p(p+1)}{n-p-1} \quad (13)$$

where n is the number of samples, p is the number of fit parameters for the model, and σ^2 is the square of RMSE between prediction and observations. A good prediction usually has high correlation coefficients but low RMSE and AIC values with observations.

For the regional simulation (simulation 2), we utilize daily surface air temperature from MERRA to drive the selected model on a resolution of 1° by latitude and 1.33° by longitude for 1982-2012. The uncertainty of predicted phenology is very sensitive to that of drivers (Migliavacca et al., 2012), as a result, we compare the MERRA forcing with ground observations from the United States Historical Climatology Network (USHCN, Easterling et al., 1996), which provides a high quality data set of daily and monthly

406 temperature from 1218 observing stations across the contiguous United States. We
407 analyze the phenological trend for different time periods so as to understand how the
408 selected time frame and interannual variability may influence our conclusions.

409

410 We perform a sensitivity analysis (simulation 3) to evaluate the uncertainty due to
411 phenological schemes. In this run, we do not include chilling constraint for the spring
412 phenology by using a fixed and calibrated forcing threshold F_v^* of 50 degree days. As a
413 result, forcing value S_f begins accumulation from winter solstice and budburst occurs if
414 only $S_f > F^*$. The whole process is not dependent on the value of chilling units S_c .
415 Meanwhile, we lift the photoperiod cap for leaf senescence by setting $f_p = 1$, so that the
416 autumn phenology is only determined by temperature ($f_A = f_T$).

417

418 We analyze species-specific temperature sensitivity of tree phenology at Harvard Forest
419 (section 4.3). Based on these results, we perform two additional sensitivity tests to
420 evaluate modeling uncertainties from the intraspecific variations. In the first run
421 (simulation 4), phenological parameters are derived based on records of species with the
422 lowest temperature sensitivity for both spring (Sweet Birch, *Betula Lenta*) and autumn
423 (Paper Birch, *Betula Papyrifera*). In the other run (simulation 5), parameters are derived
424 using records of species with the highest temperature sensitivity for spring (Striped
425 Maple, *Acer Pensylvanicum*) and autumn (Black Oak, *Quercus Velutina*). We applied the
426 derived parameters for the whole domain of U.S. by ignoring the realistic fractional
427 coverage of specific species, so as to estimate the maximum uncertainty of prediction due
428 to the intraspecific variations. We consider a change, trend, or correlation is significant if
429 $p < 0.05$, unless otherwise stated.

430

431 3 Results

432

433 3.1 Model evaluation

434

435 3.1.1 Site-level evaluation

436

Xu Yue 7/2/15 11:18 AM

Deleted: also

Xu Yue 7/2/15 11:18 AM

Deleted: *

Xu Yue 7/2/15 11:18 AM

Deleted: by setting $f_p = 1$. We consider a change, trend, or correlation is significant if $p < 0.05$, unless otherwise stated.

442 The five sites we select to calibrate and evaluate models are all located at Eastern U.S.,
 443 where >90% deciduous forests are located (Fig. 1). The site-level evaluations for 9 spring
 444 models and 4 autumn models are shown in Fig. 2 and summarized in Table S5. For the
 445 spring phenology, the alternating approach (S7-S9) has higher correlations and lower
 446 RMSE compared to parallel models (S3-S6). The sequential approach with thermal
 447 summation (S1) shows the largest correlations and lowest biases. However, it requires
 448 fitting 5 parameters, increasing its AIC value relative to the alternating models. The three
 449 alternating models have comparable correlations and RMSE. However, the modified
 450 alternating model (S9) has the lowest AIC, suggesting that fixing some parameters based
 451 on literature does not weaken the performance but can reduce model complexity. For the
 452 autumn phenology, no models predict correlations higher than 0.5, indicating that missing
 453 mechanisms, such as accidental frost, strong wind and rainfall, may be required to
 454 improve the current model structures (Richardson et al., 2006; Schuster et al., 2014). The
 455 ‘CDD-photoperiod’ scheme (A4) has comparable performance with that from Delpierre
 456 et al. (2009) (A3) based on correlation and RMSE, and has lower AIC than the latter due
 457 to the lower number of fit parameters (Table 5). As a result of the site-level evaluations,
 458 we select the spring model S9 and autumn model A4 (parameters listed in Table S3) as
 459 the state-of-art schemes for the regional simulations.

460

461 Site-level simulations with models S9 and A4 capture both the interannual variations and
 462 temporal trends of phenology at the validation sites (Fig. 1). Sites US-Ha1 and US-HB1
 463 provide >20 years of phenology records. The observation-simulation correlations for
 464 budburst dates are 0.7-0.8 at these sites. Model performance is poor for autumn
 465 phenology, with correlation coefficients between 0.2-0.4. Both observed and predicted
 466 budburst dates at US-Ha1 show significant advances of ~ 0.5 day yr^{-1} during 1992-2012.
 467 However, at US-HB1, the observed trend of -0.3 day yr^{-1} is not significant due to large
 468 interannual variations. In contrast, the dormancy start dates remains almost constant at
 469 US-Ha1, similar as that reported by Lee et al. (2003), but exhibits a significant delay of
 470 ~ 0.5 day yr^{-1} at US-HB1 in the past two decades, as reported by Keenan et al. (2014).

471

Xu Yue 7/2/15 11:18 AM

Deleted: 2.

Xu Yue 7/2/15 11:18 AM

Deleted: decreasing

Xu Yue 7/2/15 11:18 AM

Deleted: especially some physiological processes (e.g. synthesis, viscosity, and diffusion),

Xu Yue 7/2/15 11:18 AM

Deleted: Linkosalo

Xu Yue 7/2/15 11:18 AM

Deleted: 2008

479 Sites US-UMB and US-MMS have relatively short observations for 1999-2012. Missing
480 *in situ* forcing values limit the model's spring phenology performance compared to that
481 using MERRA reanalysis. With MERRA forcing, the model shows high correlations
482 (~0.8) and low biases (2-4 days) in the prediction of budburst dates. The simulated
483 autumn phenology again has lower correlations with observations at these sites. The
484 predicted dormancy start dates at US-UMB match the observed interannual variation
485 before 2010 but fail to capture the perturbations thereafter. The prediction at US-MMS
486 shows similar year-to-year variations as observations but with smaller magnitude. The
487 spring budburst dates show moderate changes at US-UMB but a significant advance at
488 US-MMS in the past decade. For the autumn phenology, both observations and
489 simulations show insignificant changes.

490

491 3.1.2 Continental-scale evaluation

492

493 Phenology has a distinctive spatial distribution over U.S. deciduous forests (Fig. 3).
494 Budburst occurs relatively later in the west of 105°W but earlier in the low latitudes of
495 the East (Fig. 3a). The area-weighted (based on cover fraction of deciduous forest)
496 budburst date for the western U.S. is May 4th or the 124th day of the year (DOY), with
497 higher values of > DOY 140 over the ridge of Rocky Mountains. In contrast, the mean
498 budburst date is April 15th (DOY 105) for the east of 105°W, with even earlier dates of
499 <DOY 100 at south of 40°N. At higher latitudes, such as the forests over New England
500 and Great Lakes, spring usually begins after DOY 125 due to the colder spring
501 temperatures. The simulated spatial pattern is consistent with phenology records from the
502 USA-NPN network, with a correlation coefficient of 0.77 over 46 sites (Fig. 3b).
503 However, the predicted budburst date at these sites is 5 days on average later than the
504 observations. Causes of such bias are unclear but might be related to the uneven spatial
505 distribution of network sites, the distinct nature of the protocols between calibration and
506 validation data, and the incompatibility between model parameters derived at regional
507 scale versus those for the continental scale. The distribution of autumn phenology shows
508 almost opposite pattern as that of spring phenology (Fig. 3c). At high latitudes and/or
509 altitudes, autumn phenology is sensitive to cold temperatures and as a result exhibits an

Xu Yue 7/2/15 11:18 AM
Deleted: 124

Xu Yue 7/2/15 11:18 AM
Deleted: DOY

Xu Yue 7/2/15 11:18 AM
Deleted: DOY

Xu Yue 7/2/15 11:18 AM
Deleted: DOY

Xu Yue 7/2/15 11:18 AM
Deleted: DOY

515 early dormancy onset. The validation against observations from 23 USA-NPN sites yields
516 a significant correlation coefficient of 0.80 for the simulated autumn phenology (Fig. 3d).
517 Similar to the spring budburst, predicted dormancy onset is later by 11 days than the
518 observations. The S9 model also reproduces year-to-year changes in spring phenology. In
519 2011, the area-weighted budburst date is DOY 117 (Fig. S12), which is advanced by 13
520 days in 2012 (Fig. S13). Such change follows the continental warming of spring (March-
521 May) temperature by ~3 °C in the latter year (not shown).

Xu Yue 7/2/15 11:18 AM

Deleted: DOY

522
523 We further evaluate the simulated year-to-year budburst dates with available long-term
524 records from NALN and USA-NPN network (Fig. 4). The correlations between modeled
525 and observed budburst dates are larger than 0.3 for 47 out of 59 sites, among which 26
526 are significant ($p < 0.1$), suggesting that the predicted interannual variation and long-term
527 trend of spring phenology are generally reasonable on the continental scale. The
528 insignificant correlations at 33 sites are in part attributed to the deviations in species
529 between model (PFT level) and measurements (species level). Furthermore, the record
530 length may also contribute to these biases because 28 out of the 33 sites with insignificant
531 correlations have records shorter than 8 years. The large interannual variability in the
532 spring phenology (see section 4.1) may affect the correlations especially for time series
533 with short record length. On the other hand, no long-term records are available to
534 evaluate the temporal variation of simulated autumn phenology on the continental scale.

Xu Yue 7/2/15 11:18 AM

Deleted: model-observation

Xu Yue 7/2/15 11:18 AM

Deleted: significantly positive

Xu Yue 7/2/15 11:18 AM

Deleted: 2) for most sites,

Xu Yue 7/2/15 11:18 AM

Deleted: However

536 3.2 Phenological change in U.S. deciduous forests

537

538 Driven with the MERRA forcing, the model simulates a significant advance of spring
539 budburst dates in central eastern U.S. during 1982-2012 (Fig. 5a). The largest advance of
540 0.42 day yr⁻¹ is predicted in the states of Illinois and Indiana. For eastern states covered
541 with >50% deciduous forests, such as Pennsylvania, West Virginia, and Virginia, the
542 budburst date is advanced by 0.34 day yr⁻¹. However, for deciduous forests in the
543 western, northern, northeastern, and southeastern U.S., the changes are either small or
544 insignificant. Two New England sites, Harvard Forest and Hubbard Brook, are located
545 within the same region but have different trends of spring phenology (Fig. 5a), consistent

Xu Yue 7/2/15 11:18 AM

Deleted:

552 with site-level evaluations for 1992-2012 (Fig. 1). On the other hand, the dormancy start
553 date is delayed by 0.20 day yr⁻¹ in the northern (Minnesota), 0.14 day yr⁻¹ in the
554 northeastern, and 0.16 day yr⁻¹ in the western forests (Fig. 5b). However, the autumn
555 phenology in central and southern U.S. does not show significant changes, consistent
556 with site-level evaluations at US-UMB and US-MMS (Fig. 1).

557

558 The spatial pattern of the trend in forest phenology follows spatial patterns of temperature
559 changes in the past 3 decades (Fig. [S14](#)). Both the reanalysis data and ground records
560 show a significant spring warming of 0.75 °C decade⁻¹ over central and eastern U.S.
561 while insignificant changes in the other portion of deciduous forest (Figs. [S14c-d](#)).
562 Meanwhile, the warmer winter may delay the spring budburst by reducing chilling days,
563 especially for forest in the northern U.S (Figs. [S14a-b](#)). On the other hand, autumn
564 warming in the northern, northeastern, and western forests (Figs. [S14e-f](#)) results in
565 delayed dormancy dates in those regions (Fig. 5b). However, autumn phenology in
566 central, eastern, and southern forests shows no significant change, due to either moderate
567 changes in temperature (Figs. [S14e-f](#)) or regulation through photoperiod. Based on the
568 synchronous phenological responses to temperature changes, we estimate long-term
569 temperature sensitivities of -3.3 days °C⁻¹ for spring budburst date and 2.2 days °C⁻¹ for
570 dormancy start date over U.S. deciduous forests. These values are close to the estimates
571 of -2.8 ± 0.3 days °C⁻¹ (spring) and 1.8 ± 0.8 days °C⁻¹ (autumn) based on observations
572 from five U.S. deciduous sites (Keenan et al., 2014). [The temperature sensitivity of](#)
573 [spring budburst date is also within the range from -2.05 to -7.48 days °C⁻¹ for different](#)
574 [species based on the field experiments performed by Vitasse et al. \(2009\).](#)

575

576 Advanced spring and delayed autumn together increased the length of the growing season
577 across the U.S. (Fig. [6](#)). Relative to the 1980s, the growing season in the 2000s extends
578 by 5.5 days (3.0%) in the eastern states with dense forest coverage (fraction > 50%). The
579 model predicts larger extension of 6.4 days (3.9%) in New England, 7.0 days (3.6%) in
580 states Illinois and Indiana, and 6.0 days (4.3%) in the upper Rocky Mountains forests
581 (Fig. [6](#)). This magnitude is comparable to the trend of 2.1-4.2 days per decade in Eurasian

Xu Yue 7/2/15 11:18 AM

Deleted: 6

Xu Yue 7/2/15 11:18 AM

Deleted: 6c

Xu Yue 7/2/15 11:18 AM

Deleted: 6a

Xu Yue 7/2/15 11:18 AM

Deleted: 6e

Xu Yue 7/2/15 11:18 AM

Deleted: 6e

Xu Yue 7/2/15 11:18 AM

Deleted: 7

Xu Yue 7/2/15 11:18 AM

Deleted: 7

589 and North American temperate forest estimated by other studies (Menzel et al., 2008;
590 Jeong et al., 2011).

591

592 3.3 Comparison with results from remote sensing

593

594 Most of up-to-date estimates of the changes in U.S. forest phenology are performed with
595 remote sensing data. [In a recent study, Buitenwerf et al. \(2015\) found an overall](#)
596 [extension of the growing season over boreal and temperate forests during 1981-2012](#)
597 [based on the Normalized Difference Vegetation Index \(NDVI\) from satellite data.](#)
598 [However, the exact phenological changes that underlie such overall greening differ](#)
599 [among regions. For U.S. forests, the longer growing season is primarily driven by later](#)
600 [leaf-off dates, though regional advance of spring is also observed. Our results are](#)
601 [generally consistent with their conclusions but with some deviations. For example, they](#)
602 [observed later autumn in almost all the eastern U.S., where we predict delays only in the](#)
603 [North and Northeast \(Fig. 5b\). Such discrepancies reflect prediction biases, and may also](#)
604 [be a consequence of satellite retrieval uncertainties \(Table 1\).](#)

605

606 We [further](#) compare our results to recent reports from the literature, selecting all studies
607 that [examine](#) phenological trends across the U.S. for at least 20 years (Table 1). All
608 selected studies use the [NDVI](#), however, they report different and even opposite trends.
609 Such discrepancies may be attributed to the differences in the definitions of phenological
610 dates (White et al., 2009) or the statistical algorithms in the extraction of the dates
611 (Keenan et al., 2014). Here, we summarize their results on Fig. [7](#) so as to conclude the
612 most robust changes for U.S. forest phenology in the past 2-3 decades. [Since the](#)
613 [definition of phenological events varies among different studies \(White et al., 2009\), we](#)
614 [qualitatively compare the simulations with the remote sensing retrievals so as to evaluate](#)
615 [the ensemble spatial distribution of phenological changes in the past decades.](#) For spring
616 phenology, [four](#) out of [seven](#) studies predict advanced budburst or greenup dates in the
617 East, while four predict delayed dates in the North (Fig. [7a](#)). There are no evident
618 phenological changes in the West, Northeast, and Southeast. Our results show similar
619 changes in spring phenology as the ensemble of the remote sensing studies, except that

Xu Yue 7/2/15 11:18 AM

Deleted: exam

Xu Yue 7/2/15 11:18 AM

Deleted: Normalized Difference Vegetation Index (NDVI) from satellite data

Xu Yue 7/2/15 11:18 AM

Deleted: 8

Xu Yue 7/2/15 11:18 AM

Deleted: 3.4 Modeling uncertainties .

... [1]

Xu Yue 7/2/15 11:18 AM

Deleted: three

Xu Yue 7/2/15 11:18 AM

Deleted: six

Xu Yue 7/2/15 11:18 AM

Deleted: 8a

631 we predict smaller delays in the northern states (Fig. 5a). In addition, our data-informed
 632 model simulates significant spring advances in the central U.S., while remote sensing
 633 studies largely disagree over this area. On the other hand, both the remote sensing studies
 634 and our results show that autumn phenology is significantly delayed in the West, [North](#),
 635 and Northeast (Figs. 5b and [7b](#)). However, the examined studies [also](#) exhibit significant
 636 delays in the [central](#) states, in contrast to our results. In other areas, the trends are
 637 insignificant (Southeast and East).

638

639

640 [4. Discussion](#)

641

642 [4.1 Impact of interannual variability](#)

643

644 Estimates of trends in phenology are sensitive to the length of the examined time frame
 645 due to relatively large internal climate variability (Badeck et al., 2004; Iler et al., 2013).
 646 Our analyses show that interannual variations may also cause large uncertainties in the
 647 estimated phenology trend, especially on short decadal time scales. For example, Keenan
 648 et al. (2014) estimated a large advance of 0.48 day yr^{-1} in the spring phenology in both
 649 the Harvard Forest and Hubbard Brook sites between 1990 and 2012, and across the
 650 Eastern U.S. temperate forest for 2000-2012. Our data-informed modeling approach
 651 estimated a similar change of 0.42 day yr^{-1} between 2000 and 2012 over the eastern US
 652 (Fig. [S15a](#)), but the trend was largely affected by the record-breaking advance of spring
 653 in 2012 (Jolly et al., 2005), especially over the central and eastern states (Figs. S12-S13).
 654 If we exclude this specific year, we achieve an average trend of only $-0.05 \text{ day yr}^{-1}$ for
 655 2000-2011, with delayed budburst dates in central and southern states (Fig. [S15b](#)). In
 656 addition, interannual variability may affect the significance of the derived trend. As
 657 shown in Fig. [S15a](#), the advance of spring phenology is not significant for 2000-2012,
 658 based on the linear regression, possibly because of the large year-to-year variations and
 659 the insignificant changes in air temperature (Fig. [S16](#)). A similar result is shown for
 660 autumn phenology (Fig. [S15d](#)). However, if we extend the analysis period to 1982-2011,
 661 the estimated trends and their significance are not affected by the anomalous phenology

Xu Yue 7/2/15 11:18 AM

Deleted: 8b

Xu Yue 7/2/15 11:18 AM

Deleted: do not

Xu Yue 7/2/15 11:18 AM

Deleted: northern

Xu Yue 7/2/15 11:18 AM

Deleted: either

Xu Yue 7/2/15 11:18 AM

Deleted:) or uncertain (Center).

Xu Yue 7/2/15 11:18 AM

Deleted: 3.4 Modeling uncertainty ... [2]

Xu Yue 7/2/15 11:18 AM

Deleted: S14a

Xu Yue 7/2/15 11:18 AM

Deleted: S14b

Xu Yue 7/2/15 11:18 AM

Deleted: S14a

Xu Yue 7/2/15 11:18 AM

Deleted: S15

Xu Yue 7/2/15 11:18 AM

Deleted: S14d

674 | change in the year 2012 (Figs. [S15c](#) and [S15f](#)), suggesting that the estimate of long-term
675 | trend is more robust compared to the short-term trend.

676

677 | [4.2 Impact of chilling requirement and photoperiod limit](#)

678

679 | We perform an additional sensitivity experiment (simulation 3) to examine the impact of
680 | model structure on the phenology prediction. For spring phenology, model validations
681 | have shown that the spring warming (1-phase) models are as efficient as chilling (2-
682 | phase) models (Vitasse et al., 2011; Fu et al., 2012a; Migliavacca et al., 2012). In the
683 | simulation 3, we remove the limit of chilling requirement on the forcing threshold F^* by
684 | defining a fixed forcing threshold. The site-level evaluation shows that this simulation
685 | has higher correlations at three out of four sites compared to that with chilling
686 | requirement (not shown). Driven with MERRA temperature, the simulation 3 (Fig. [S17a](#))
687 | predicts a similar spatial pattern for the trend of budburst date in the U.S. as that in
688 | simulation 2 (Fig. 5a), although the former estimates larger advances in central (0.52 day
689 | yr^{-1}) and eastern U.S (0.43 day yr^{-1}). Such stronger signal in the trend of spring
690 | phenology could be attributed to the omission of offset effects from the winter warming
691 | (Figs. [S14a-b](#)). In the simulation 3, we also remove the cap of photoperiod for autumn
692 | dormancy and achieve better correlations between simulations and observations at all
693 | sites, though this method tends to generate later dormancy, especially at warm sites (up to
694 | 20 days, not shown). Continental-scale simulation without photoperiod limit (Fig. [S17b](#))
695 | results in similar trend in autumn phenology as that with photoperiod (Fig. 5b),
696 | suggesting that the response to temperature dominates the phenological change in the
697 | U.S. deciduous forest.

698

699 | [Our investigation of the roles of chilling and photoperiod is sensitive to the model](#)
700 | [structure, climate variability, and data availability. First, the similar performance between](#)
701 | [spring warming and chilling models might also result from the inaccurate representation](#)
702 | [of chilling / photoperiod mechanisms. For example, the chilling units used in our](#)
703 | [parameterization are calculated based on daily average temperatures, while Piao et al.](#)
704 | [\(2015\) suggested that leaf unfolding dates during 1982-2011 are triggered by daytime](#)

Xu Yue 7/2/15 11:18 AM

Deleted: S14c

Xu Yue 7/2/15 11:18 AM

Deleted: S14f

Xu Yue 7/2/15 11:18 AM

Deleted: 3.

Xu Yue 7/2/15 11:18 AM

Deleted: S16a

Xu Yue 7/2/15 11:18 AM

Deleted: 6a

Xu Yue 7/2/15 11:18 AM

Deleted: S16b

Xu Yue 7/2/15 11:18 AM

Formatted: Font:Not Bold

Xu Yue 7/2/15 11:18 AM

Deleted: 3.4.3 Impact of species aggregation

... [3]

more than by nighttime temperature. The up-to-date autumn phenology model fails to capture interannual variability of dormancy onset (Fig. 2), suggesting that unknown processes are involved in the autumn leaf fall (Keenan and Richardson, 2015). It is unclear whether these processes are related to the variations of photoperiod. Second, the decadal changes in temperature may mask the role of chilling. The trend of winter warming is not significant for most areas in the U.S. (Fig. S14a), suggesting that chilling requirements have been fulfilled in the past 3 decades. However, it is unclear whether the winter warming will intensify in the future, which may slow the advancement of spring budburst. Third, we choose to calibrate the phenological parameterization at the level of plant functional type (PFT) because species-specific measurements are usually incomplete in time and uneven in space. Such incompleteness may influence the accuracy of derived decadal phenological records used for both model calibration and validation. At the same time, PFT-level parameterization may be too broad for the vegetation modeling because it fails to capture intraspecific variations (Van Bodegom et al., 2012; Reichstein et al., 2014). Observations at the community level suggest that the budburst of some species is sensitive to fall/winter and spring warming but with opposite signs (Cook et al., 2012). In the next subsection, we examine the records of 13 deciduous tree species at Harvard Forest.

4.3 Impact of species aggregation

Tree phenology and its responses to temperature changes have been shown to vary among species (Vitasse et al., 2009; Fu et al., 2012a; Archetti et al., 2013). In this study, however, we calibrate model parameters based on the long-term phenological cycle derived from LAI, which represents the mean growing seasonality averaged among species. We do not perform the species-specific simulation for the following three reasons. First, the species-level measurements are usually not available on the continental scale, which influences both model calibration and validation. Second, species-level modeling increases the complexity and computational costs while decreasing predictive reliability (Prentice et al., 2015). Third, investigations at both site level and continental

Xu Yue 7/2/15 11:18 AM

Deleted: are thought

Xu Yue 7/2/15 11:18 AM

Deleted: do not perform species-specific model calibration and validation, because we focus on phenology on the continental scale. We

Xu Yue 7/2/15 11:18 AM

Deleted: As a check, we

scale show similar temperature sensitivity of tree phenology between the species-specific and species-aggregation approaches.

We analyze the temperature sensitivity of tree phenology for 13 DBF species at Harvard Forest (Fig. 8). We calculate the ensemble phenology based on the basal area of each species (the dominant species are red oak (*Quercus Rubra*, 60% basal area) and red maple (*Acer Rubrum*, 23% basal area)) in order to represent the average phenology at Harvard Forest, which has been used in the site-level evaluation (Fig. 1). For spring phenology, the mean budburst dates vary by up to 3 weeks among different species, with the earliest being alternated-leaved dogwood (*Cornus Alterniflora*) and the latest white oak (*Quercus Alba*) (Fig. 8c). Two dominant species, red oak and red maple, have similar year-to-year variations, leading to a similar magnitude of ensemble phenology and the long-term trend (Fig. 8a). The 21-year average of the ensemble budburst date is DOY 126, very close to the DOY 125 derived from LAI (Table 3). Regressions against mean March and April temperature show similar sensitivity of budburst date for most species, especially for red oak ($-3.8 \text{ day } ^\circ\text{C}^{-1}$) and red maple ($-3.4 \text{ day } ^\circ\text{C}^{-1}$) (Fig. 8c). Such similarity also provides us the foundation to validate the simulated interannual variation of spring phenology with the lilac data (Fig. 4). For autumn phenology, the averaged dormancy onset date of red maple is 23 days earlier than that of red oak (Fig. 8d), leading to medium ensemble values (Fig. 8b). The 21-year average of the ensemble dormancy onset date is DOY 306, again close to the estimate of DOY 310 based on LAI (Table 3). The temperature sensitivity of autumn phenology is positive for all species, including similar magnitude of $2.6 \text{ day } ^\circ\text{C}^{-1}$ for red maple and $2.3 \text{ day } ^\circ\text{C}^{-1}$ for red oak (Fig. 8f), though the latter is insignificant due to the large year-to-year variations. The species-specific analyses show that calibration based on LAI may capture the representative phenology at deciduous forests, and is not affected by the large deviations among species. Since the eastern U.S. is dominated by oak and maple trees (http://www.nrs.fs.fed.us/atlas/tree/curr_fortypes.html), which we show have very similar temperature sensitivity for both the spring and autumn phenology, we expect that the species aggregation applied in this study may reasonably capture the temperature

Xu Yue 7/2/15 11:18 AM

Deleted: 9

Xu Yue 7/2/15 11:18 AM

Deleted: also

Xu Yue 7/2/15 11:18 AM

Deleted: -

... [4]

Xu Yue 7/2/15 11:18 AM

Deleted: 9c

Xu Yue 7/2/15 11:18 AM

Deleted: 9a

Xu Yue 7/2/15 11:18 AM

Deleted: 9c

Xu Yue 7/2/15 11:18 AM

Deleted: 9d

Xu Yue 7/2/15 11:18 AM

Deleted: 9b

Xu Yue 7/2/15 11:18 AM

Deleted: 9f

sensitivity of forest phenology on the continental scale, given that temperature is likely the dominant driver of phenology change for such deciduous forests (Fig. 5).

We perform two sensitivity runs to evaluate the modeling uncertainties due to intraspecific variations at the continental scale (Fig. S18). Simulations with either the lowest (simulation 4) or the highest (simulation 5) temperature sensitivity yield very similar phenological trends as that in the control simulation (simulation 2). In the East, simulation 4 predicts a spring advance by 0.33 day yr⁻¹ while simulation 5 predicts an advance by 0.35 day yr⁻¹, both of which are close to the 0.34 day yr⁻¹ from the control run. In the West and Northeast, both sensitivity runs predict autumn delay by 0.13-0.15 day yr⁻¹, lower than value of 0.14-0.16 day yr⁻¹ from control run, suggesting that site-level responses may not be necessarily consistent with responses at the continental scale. Both the similar temperature sensitivity at site level (Fig. 8) and the predicted phenological trends at continental scale (Fig. S18) support the concept of phenological modeling at the forest and PFT level, and corroborates the further investigation of phenology-climate interactions at the continental and global scale.

5 Conclusions

We performed model inter-comparison to identify the state-of-art scheme for predicting tree phenology of U.S. deciduous forests. An extensive database of ground measurements, including long-term records of phenological events at the site level and short-term records widely scattered on the national scale, was compiled to evaluate the models. The selected models with the lowest AIC values utilized the accumulative temperature summation, with additional constraints of winter chilling on spring phenology and photoperiod on autumn phenology. The 30-year phenology trend of U.S. deciduous forest was explored using the selected models. Consistent with an ensemble of remote-sensing studies, the continental simulation showed a significant advance of 0.34 day yr⁻¹ for spring budburst dates in the East with >50% coverage of deciduous forests during 1982-2012. However, no significant changes were found over the western,

Xu Yue 7/2/15 11:18 AM

Formatted: Font:Not Bold

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Deleted: -

... [5]

825 northern, northeastern, and southeastern U.S. On the other hand, the autumn dormancy
826 onset dates is delayed by 0.20 day yr⁻¹ in the northern, 0.14 day yr⁻¹ in the northeastern,
827 and 0.16 day yr⁻¹ in the western forests, but is not significant elsewhere.

828

829 Uncertainties in phenological predictions originate from drivers, parameters, and model
830 structures (Migliavacca et al., 2012). In this study, we minimize uncertainties from
831 meteorological forcings by utilizing an updated reanalysis product and validate the
832 gridded forcings with site-based observations. For the model parameters, we calibrate
833 model parameters with long-term average phenology at four deciduous sites with diverse
834 spatial distribution. This approach was chosen because a well-calibrated phenology
835 model based on a single dataset may have poor performance against external data sets
836 (Chuine et al., 1999; Richardson et al., 2006). The validation shows that the predicted
837 spatial pattern is reasonable and the long-term average matches observations within
838 sampling uncertainty (Figs. 3-4). However, due to the data scarcity, all the selected sites
839 are located in temperate areas ranging from 38°-46°N, suggesting that the model should
840 be used cautiously at other latitudes and parameters may require re-calibration. For model
841 structure, we perform sensitivity tests both with and without chilling requirements and
842 photoperiod limit and find that the predicted phenology and its change is not sensitive to
843 these constraints at least for the U.S. domain.

844

845 Our model inter-comparison does not show a distinct advantage for a specific spring
846 model, suggesting that the model formulation, such as sequential, parallel, and
847 alternating, is not a dominant source of uncertainty for estimates of spring phenology. On
848 the other hand, the evaluation of autumn phenology shows that models with cumulative
849 cold summation and photoperiod limits may better capture the trend of the dormancy
850 onset dates. However, the state-of-art autumn models still have large biases in capturing
851 year-to-year variations. Missing mechanisms, potentially including biotic (e.g. [tree age](#)
852 [\(Vitasse, 2013; Caldararu et al., 2014\)](#) and [species](#) (Vitasse et al., 2009)) and abiotic (e.g.
853 water stress (Jones et al., 2014), accidental frost (Schuster et al., 2014), strong wind, and
854 air pollution [\(Gallinat et al., 2015\)](#), and timing of spring flushing [\(Fu et al., 2014; Keenan](#)

Xu Yue 7/2/15 11:18 AM

Deleted: tree age (Caldararu et al., 2014) and species

Xu Yue 7/2/15 11:18 AM

Deleted:),

Xu Yue 7/2/15 11:18 AM

Deleted:), environmental (e.g.

Xu Yue 7/2/15 11:18 AM

Deleted: (Gallinat et al., 2015)) and physiological (e.g. rate of synthesis

Xu Yue 7/2/15 11:18 AM

Deleted: Schaber

and [Richardson, 2015](#)) factors, may jointly affect leaf fall in a process that is currently not well understood.

Given these uncertainties, our results [show](#) a significant advance of 0.34 day yr⁻¹ for spring budburst dates in the East of U.S. during 1982-2012, while a delay of 0.15 day yr⁻¹ for autumn dormancy onset dates in the Northeast and West. Such long-term changes in phenology are mainly attributed to the trends in temperature, as simulations without chilling requirement and photoperiod limit showed similar phenological changes. Due to either the advances in spring or delays in autumn, tree growth period extends by about 1 week (3-4%) at the 2000s relative to the 1980s, indicating prominent influences of climate change on the carbon cycle and ecological evolution of the U.S. deciduous forests.

Acknowledgements

Hubbard Brook phenology data were provided by A. Bailey at the USDA Forest Service, Northern Research Station, Hubbard Brook Experimental Forest. This project was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center. TFK acknowledges funding from a Macquarie University Research Fellowship.

Xu Yue 7/2/15 11:18 AM

Deleted: Badeck, 2003

Xu Yue 7/2/15 11:18 AM

Deleted: and timing of spring flushing (Keenan and Richardson, 2015)

Xu Yue 7/2/15 11:18 AM

Formatted: Font color: Auto

Xu Yue 7/2/15 11:18 AM

Deleted: showed

References

- Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium on Information Theory, Petrov, B. N. and Csaki, F. (Eds.), Akademiai Kiado, Budapest, 1973.
- Archetti, M., Richardson, A. D., O'Keefe, J., and Delpierre, N.: Predicting Climate Change Impacts on the Amount and Duration of Autumn Colors in a New England Forest, *Plos One*, 8, e57373, doi:10.1371/journal.pone.0057373, 2013.
- Badeck, F. W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J., and Sitch, S.: Responses of spring phenology to climate change, *New Phytol*, 162, 295-309, doi:10.1111/J.1469-8137.2004.01059.X, 2004.
- Bailey, A.: Routine Phenology Measurements. Hubbard Brook Data Archive, <http://hubbardbrook.org/data/dataset.php?id=51> (last access: October 13th, 2014), Durham, NH, 2014.
- Basler, D. and Korner, C.: Photoperiod sensitivity of bud burst in 14 temperate forest tree species, *Agr Forest Meteorol*, 165, 73-81, doi:10.1016/J.Agrformet.2012.06.001, 2012.
- Buitenwerf, R., Rose, L., and Higgins, S. I.: Three decades of multi-dimensional change in global leaf phenology, *Nat Clim Change*, 5, 364-368, 2015.
- Burnham, K. P. and Anderson, D. R.: Model selection and multimodel inference: A practical information-theoretic approach, Springer-Verlag, New York, 2002.
- Caldararu, S., Purves, D. W., and Palmer, P. I.: Phenology as a strategy for carbon optimality: a global model, *Biogeosciences*, 11, 763-778, doi:10.5194/Bg-11-763-2014, 2014.
- 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, doi:10.1046/J.1365-3040.1999.00395.X, 1999.
- Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A., and Schwartz, M. D.: Shifting plant phenology in response to global change, *Trends Ecol Evol*, 22, 357-365, doi:10.1016/J.Tree.2007.04.003, 2007.
- Cook, B. I., Wolkovich, E. M., and Parmesan, C.: Divergent responses to spring and winter warming drive community level flowering trends, *P Natl Acad Sci USA*, 109, 9000-9005, doi:10.1073/Pnas.1118364109, 2012.
- Delpierre, N., Dufrene, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J., and Francois, C.: Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France, *Agr Forest Meteorol*, 149, 938-948, doi:10.1016/J.Agrformet.2008.11.014, 2009.
- Delpierre, N., Vitasse, Y., Chuine, I., Guillemot, J., Bazot, S., Rutishauser, T., and Rathgeber, C. B. K.: Temperate and boreal forest tree phenology: from organ-scale processes to terrestrial ecosystem models, *Annals of Forest Science*, doi:10.1007/s13595-015-0477-6, 2015. doi:10.1007/s13595-015-0477-6, 2015.
- Dragoni, D. and Rahman, A. F.: Trends in fall phenology across the deciduous forests of the Eastern USA, *Agr Forest Meteorol*, 157, 96-105, doi:10.1016/J.Agrformet.2012.01.019, 2012.
- Dufrene, E., Davi, H., Francois, C., le Maire, G., Le Dantec, V., and Granier, A.: Modelling carbon and water cycles in a beech forest Part I: Model description and

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Deleted: 267-281,

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

uncertainty analysis on modelled NEE, *Ecol Model*, 185, 407-436, doi:10.1016/J.Ecolmodel.2005.01.004, 2005.

Easterling, D. R., Karl, T. R., Mason, E. H., Hughes, P. Y., and Bowman, D. P.: United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data, ORNL/CDIAC-87, NDP-019/R3, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 1996.

Fitter, A. H. and Fitter, R. S. R.: Rapid changes in flowering time in British plants, *Science*, 296, 1689-1691, doi:10.1126/Science.1071617, 2002.

Forrest, J. and Miller-Rushing, A. J.: Toward a synthetic understanding of the role of phenology in ecology and evolution, *Philos T R Soc B*, 365, 3101-3112, doi:10.1098/Rstb.2010.0145, 2010.

Fu, Y., Campioli, M., Deckmyn, G., and Janssens, I. A.: The Impact of Winter and Spring Temperatures on Temperate Tree Budburst Dates: Results from an Experimental Climate Manipulation, *Plos One*, 7, e47324, doi:10.1371/journal.pone.0047324, 2012a.

Fu, Y., Campioli, M., Van Oijen, M., Deckmyn, G., and Janssens, I. A.: Bayesian comparison of six different temperature-based budburst models for four temperate tree species, *Ecol Model*, 230, 92-100, doi:10.1016/J.Ecolmodel.2012.01.010, 2012b.

Fu, Y. S. H., Campioli, M., Vitasse, Y., De Boeck, H. J., Van den Berge, J., AbdElgawad, H., Asard, H., Piao, S. L., Deckmyn, G., and Janssens, I. A.: Variation in leaf flushing date influences autumnal senescence and next year's flushing date in two temperate tree species, *P Natl Acad Sci USA*, 111, 7355-7360, doi:10.1073/Pnas.1321727111, 2014.

Gallinat, A. S., Primack, R. B., and Wagner, D. L.: Autumn, the neglected season in climate change research, *Trends Ecol Evol*, 30, 169-176, doi:10.1016/J.Tree.2015.01.004, 2015.

Gordo, O. and Sanz, J. J.: Long-term temporal changes of plant phenology in the Western Mediterranean, *Global Change Biol*, 15, 1930-1948, doi:10.1111/J.1365-2486.2009.01851.X, 2009.

Gough, C. M., Vogel, C. S., Schmid, H. P., Su, H. B., and Curtis, P. S.: Multi-year convergence of biometric and meteorological estimates of forest carbon storage, *Agr Forest Meteorol*, 148, 158-170, doi:10.1016/J.Agrformet.2007.08.004, 2008.

Gu, L. H., Meyers, T., Pallardy, S. G., Hanson, P. J., Yang, B., Heuer, M., Hosman, K. P., Riggs, J. S., Sluss, D., and Wullschleger, S. D.: Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, *J. Geophys. Res.*, 111, D16102, doi:10.1029/2006jd007161, 2006.

Hartmann, D. L., Tank, A. M. G. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A.-R., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P.: Observations: Atmosphere and Surface. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Hurrell, J., Marengo, J., Tangang, F., and Viterbo, P. (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 159-254, 2013.

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Deleted: in Ecology and Evolution

Xu Yue 7/2/15 11:18 AM

Deleted: j.tree

983 | Iler, A. M., Hoyer, T. T., Inouye, D. W., and Schmidt, N. M.: Long-Term Trends Mask
984 | Variation in the Direction and Magnitude of Short-Term Phenological Shifts, *Am J*
985 | *Bot*, 100, 1398-1406, doi:10.3732/Ajb.1200490, 2013.

986 | Jeong, S. J., Ho, C. H., Gim, H. J., and Brown, M. E.: Phenology shifts at start vs. end of
987 | growing season in temperate vegetation over the Northern Hemisphere for the period
988 | 1982-2008, *Global Change Biol*, 17, 2385-2399, doi:10.1111/J.1365-
989 | 2486.2011.02397.X, 2011.

990 | Jolly, W. M., Nemani, R., and Running, S. W.: A generalized, bioclimatic index to
991 | predict foliar phenology in response to climate, *Global Change Biol*, 11, 619-632,
992 | doi:10.1111/J.1365-2486.2005.00930.X, 2005.

993 | Jones, M. O., Kimball, J. S., and Nemani, R. R.: Asynchronous Amazon forest canopy
994 | phenology indicates adaptation to both water and light availability, *Environ Res Lett*,
995 | 9, 124021, doi:10.1088/1748-9326/9/12/124021, 2014.

996 | Julien, Y. and Sobrino, J. A.: Global land surface phenology trends from GIMMS
997 | database, *Int J Remote Sens*, 30, 3495-3513, doi:10.1080/01431160802562255, 2009.

998 | Kaduk, J. and Heimann, M.: A prognostic phenology scheme for global terrestrial carbon
999 | cycle models, *Clim. Res.*, 6, 1-19, 1996.

1000 | Keenan, T. F., Gray, J., Friedl, M. A., Toomey, M., Bohrer, G., Hollinger, D. Y.,
1001 | Munger, J. W., O'Keefe, J., Schmid, H. P., SueWing, I., Yang, B., and Richardson, A.
1002 | D.: Net carbon uptake has increased through warming-induced changes in temperate
1003 | forest phenology, *Nat Clim Change*, 4, 598-604, doi:10.1038/Nclimate2253, 2014.

1004 | Keenan, T. F. and Richardson, A. D.: The timing of autumn senescence is affected by the
1005 | time of spring phenology: implications for predictive models, *Global Change Biol*,
1006 | 21, 2634-2641, doi:10.1111/gcb.12890, 2015.

1007 | Körner, C. and Basler, D.: Phenology Under Global Warming, *Science*, 327, 1461-1462,
1008 | doi:10.1126/Science.1186473, 2010.

1009 | Laube, J., Sparks, T. H., Estrella, N., and Menzel, A.: Does humidity trigger tree
1010 | phenology? Proposal for an air humidity based framework for bud development in
1011 | spring, *New Phytol*, 202, 350-355, doi:10.1111/Nph.12680, 2014.

1012 | Lee, D. W., O'Keefe, J., Holbrook, N. M., and Feild, T. S.: Pigment dynamics and
1013 | autumn leaf senescence in a New England deciduous forest, eastern USA, *Ecol Res*,
1014 | 18, 677-694, doi:10.1111/J.1440-1703.2003.00588.X, 2003.

1015 | Linkosalo, T., Lappalainen, H. K., and Hari, P.: A comparison of phenological models of
1016 | leaf bud burst and flowering of boreal trees using independent observations, *Tree*
1017 | *Physiol*, 28, 1873-1882, 2008.

1018 | Melaas, E. K., Richardson, A. D., Friedl, M. A., Dragoni, D., Gough, C. M., Herbst, M.,
1019 | Montagnani, L., and Moors, E.: Using FLUXNET data to improve models of
1020 | springtime vegetation activity onset in forest ecosystems, *Agr Forest Meteorol*, 171,
1021 | 46-56, doi:10.1016/J.Agrformet.2012.11.018, 2013.

1022 | Menzel, A., Estrella, N., Heitland, W., Susnik, A., Schleip, C., and Dose, V.: Bayesian
1023 | analysis of the species-specific lengthening of the growing season in two European
1024 | countries and the influence of an insect pest, *Int J Biometeorol*, 52, 209-218,
1025 | doi:10.1007/S00484-007-0113-8, 2008.

1026 | Menzel, A. and Fabian, P.: Growing season extended in Europe, *Nature*, 397, 659-659,
1027 | doi:10.1038/17709, 1999.

Xu Yue 7/2/15 11:18 AM

Deleted: Körner

- 1029 [Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kubler, K.,](#)
1030 [Bissolli, P., Braslavsky, O., Briede, A., Chmielewski, F. M., Crepinsek, Z., Curnel,](#)
1031 [Y., Dahl, A., Defila, C., Donnelly, A., Filella, Y., Jatza, K., Mage, F., Mestre, A.,](#)
1032 [Nordli, O., Penuelas, J., Pirinen, P., Remisova, V., Scheifinger, H., Striz, M., Susnik,](#)
1033 [A., Van Vliet, A. J. H., Wielgolaski, F. E., Zach, S., and Züst, A.: European](#)
1034 [phenological response to climate change matches the warming pattern, *Global*](#)
1035 [Change Biol, 12, 1969-1976, doi:10.1111/J.1365-2486.2006.01193.X, 2006.](#)
- 1036 [Migliavacca, M., Sonnentag, O., Keenan, T. F., Cescatti, A., O'Keefe, J., and Richardson,](#)
1037 [A. D.: On the uncertainty of phenological responses to climate change, and](#)
1038 [implications for a terrestrial biosphere model, *Biogeosciences*, 9, 2063-2083,](#)
1039 [doi:10.5194/Bg-9-2063-2012, 2012.](#)
- 1040 [Morin, X., Lechowicz, M. J., Augspurger, C., O'Keefe, J., Viner, D., and Chuine, I.: Leaf](#)
1041 [phenology in 22 North American tree species during the 21st century, *Global Change*](#)
1042 [Biol, 15, 961-975, doi:10.1111/J.1365-2486.2008.01735.X, 2009.](#)
- 1043 [Murray, M. B., Cannell, M. G. R., and Smith, R. I.: Date of Budburst of fifteen Tree](#)
1044 [Species in Britain Following Climatic Warming, *J Appl Ecol*, 26, 693-700,](#)
1045 [doi:10.2307/2404093, 1989.](#)
- 1046 [O'Keefe, J.: Phenology of Woody Species at Harvard Forest since 1990. Long Term](#)
1047 [Ecological Research Network,](#)
1048 [http://dx.doi.org/10.6073/pasta/b151c3eb552433a2a94c6f8de489740b, 2000.](#)
- 1049 [Piao, S. L., Cui, M. D., Chen, A. P., Wang, X. H., Ciais, P., Liu, J., and Tang, Y. H.:](#)
1050 [Altitude and temperature dependence of change in the spring vegetation green-up](#)
1051 [date from 1982 to 2006 in the Qinghai-Xizang Plateau, *Agr Forest Meteorol*, 151,](#)
1052 [1599-1608, doi:10.1016/J.Agrformet.2011.06.016, 2011.](#)
- 1053 [Piao, S. L., Tan, J. G., Chen, A. P., Fu, Y. H., Ciais, P., Liu, Q., Janssens, I. A., Vicca, S.,](#)
1054 [Zeng, Z. Z., Jeong, S. J., Li, Y., Myneni, R. B., Peng, S. S., Shen, M. G., and](#)
1055 [Penuelas, J.: Leaf onset in the northern hemisphere triggered by daytime](#)
1056 [temperature, *Nat Commun*, 6, 6911, doi:10.1038/Ncomms7911, 2015.](#)
- 1057 [Polgar, C. A. and Primack, R. B.: Leaf-out phenology of temperate woody plants: from](#)
1058 [trees to ecosystems, *New Phytol*, 191, 926-941, doi:10.1111/J.1469-](#)
1059 [8137.2011.03803.X, 2011.](#)
- 1060 [Prentice, I. C., Liang, X., Medlyn, B. E., and Wang, Y.-P.: Reliable, robust and realistic:](#)
1061 [the three R's of next-generation land-surface modelling, *Atmos. Chem. Phys.*, 15,](#)
1062 [5987-6005, doi:10.5194/acp-15-5987-2015, 2015.](#)
- 1063 [Reed, B. C.: Trend analysis of time-series phenology of North America derived from](#)
1064 [satellite data, *Gisci Remote Sens*, 43, 24-38, 2006.](#)
- 1065 [Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama,](#)
1066 [S. P. P., and Toure, A.: Assessment and Enhancement of MERRA Land Surface](#)
1067 [Hydrology Estimates, *J Climate*, 24, 6322-6338, doi:10.1175/Jcli-D-10-05033.1,](#)
1068 [2011.](#)
- 1069 [Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking](#)
1070 [plant and ecosystem functional biogeography, *P Natl Acad Sci USA*, 111, 13697-](#)
1071 [13702, doi:10.1073/Pnas.1216065111, 2014.](#)
- 1072 [Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G. S.,](#)
1073 [Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity,](#)
1074 [S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H.,](#)

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

Xu Yue 7/2/15 11:18 AM

Formatted: Justified

- 1075 Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto,
1076 D. M., Sahoo, A. K., Schaefer, K., Tian, H. Q., Vargas, R., Verbeeck, H., Xiao, J. F.,
1077 and Xue, Y. K.: Terrestrial biosphere models need better representation of vegetation
1078 phenology: results from the North American Carbon Program Site Synthesis, *Global*
1079 *Change Biol*, 18, 566-584, doi:10.1111/J.1365-2486.2011.02562.X, 2012.
- 1080 | Richardson, A. D., Bailey, A. S., Denny, E. G., Martin, C. W., and O'Keefe, J.:
1081 Phenology of a northern hardwood forest canopy, *Global Change Biol*, 12, 1174-
1082 1188, doi:10.1111/j.1365-2486.2006.01164.x, 2006.
- 1083 | Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey,
1084 M.: Climate change, phenology, and phenological control of vegetation feedbacks to
1085 the climate system, *Agr Forest Meteorol*, 169, 156-173, 2013.
- 1086 | Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E.,
1087 Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J. Y.,
1088 Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R.,
1089 Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson,
1090 F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-
1091 Era Retrospective Analysis for Research and Applications, *J Climate*, 24, 3624-3648,
1092 doi:10.1175/Jcli-D-11-00015.1, 2011.
- 1093 | Schmid, H. P., Grimmer, C. S. B., Cropley, F., Offerle, B., and Su, H. B.:
1094 Measurements of CO₂ and energy fluxes over a mixed hardwood forest in the mid-
1095 western United States, *Agr Forest Meteorol*, 103, 357-374, doi:10.1016/S0168-
1096 1923(00)00140-4, 2000.
- 1097 | Schuster, C., Kirchner, M., Jakobi, G., and Menzel, A.: Frequency of inversions affects
1098 senescence phenology of *Acer pseudoplatanus* and *Fagus sylvatica*, *Int J Biometeorol*,
1099 58, 485-498, doi:10.1007/S00484-013-0709-0, 2014.
- 1100 | Schwartz, M. D. and Crawford, T. M.: Detecting energy-balance modifications at the
1101 onset of spring, *Physical Geography*, 22, 394-409, 2001.
- 1102 | Schwartz, M. D. and Hanes, J. M.: Intercomparing multiple measures of the onset of
1103 spring in eastern North America, *Int J Climatol*, 30, 1614-1626,
1104 doi:10.1002/Joc.2008, 2010.
- 1105 | Schwartz, M. D. and Reiter, B. E.: Changes in North American spring, *Int J Climatol*, 20,
1106 929-932, doi:10.1002/1097-0088(20000630)20:8<929::Aid-Joc557>3.0.Co;2-5,
1107 2000.
- 1108 | Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O.,
1109 Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of
1110 ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ
1111 dynamic global vegetation model, *Global Change Biol*, 9, 161-185,
1112 doi:10.1046/J.1365-2486.2003.00569.X, 2003.
- 1113 | Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B.
1114 H., Milliman, T., O'Keefe, J., and Richardson, A. D.: Digital repeat photography for
1115 phenological research in forest ecosystems, *Agr Forest Meteorol*, 152, 159-177,
1116 doi:10.1016/J.Agrformet.2011.09.009, 2012.
- 1117 | Sparks, T. H. and Menzel, A.: Observed changes in seasons: An overview, *Int J Climatol*,
1118 22, 1715-1725, doi:10.1002/Joc.821, 2002.
- 1119 | Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K.,
1120 Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling CO₂ exchange

Xu Yue 7/2/15 11:18 AM

Deleted: Doi

Xu Yue 7/2/15 11:18 AM

Deleted: Schaber, J. and Badeck, F. W.:
Physiology-based phenology models for
forest tree species in Germany, *Int J*
Biometeorol, 47, 193-201,
doi:10.1007/S00484-003-0171-5, 2003.

1127 on timescales from hourly to decadal at Harvard Forest, *J. Geophys. Res.*, 112,
 1128 G02020, doi:10.1029/2006JG000293, 2007.
 1129 [Van Bodegom, P. M., Douma, J. C., Witte, J. P. M., Ordonez, J. C., Bartholomeus, R. P.,](#)
 1130 [and Aerts, R.: Going beyond limitations of plant functional types when predicting](#)
 1131 [global ecosystem-atmosphere fluxes: exploring the merits of traits-based approaches,](#)
 1132 [Global Ecol Biogeogr, 21, 625-636, doi:10.1111/J.1466-8238.2011.00717.X, 2012.](#)
 1133 [Vitasse, Y.: Ontogenic changes rather than difference in temperature cause understory](#)
 1134 [trees to leaf out earlier, New Phytol, 198, 149-155, doi:10.1111/Nph.12130, 2013.](#)
 1135 Vitasse, Y., Delzon, S., Dufrene, E., Pontailler, J. Y., Louvet, J. M., Kremer, A., and
 1136 Michalet, R.: Leaf phenology sensitivity to temperature in European trees: Do within-
 1137 species populations exhibit similar responses?, *Agr Forest Meteorol*, 149, 735-744,
 1138 doi:10.1016/J.Agrformet.2008.10.019, 2009.
 1139 Vitasse, Y., Francois, C., Delpierre, N., Dufrene, E., Kremer, A., Chuine, I., and Delzon,
 1140 S.: Assessing the effects of climate change on the phenology of European temperate
 1141 trees, *Agr Forest Meteorol*, 151, 969-980, doi:10.1016/J.Agrformet.2011.03.003,
 1142 2011.
 1143 White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O.
 1144 P., O'Keefe, J., Zhang, G., Nemani, R. R., van Leeuwen, W. J. D., Brown, J. F., de
 1145 Wit, A., Schaepman, M., Lin, X. M., Dettinger, M., Bailey, A. S., Kimball, J.,
 1146 Schwartz, M. D., Baldocchi, D. D., Lee, J. T., and Lauenroth, W. K.:
 1147 Intercomparison, interpretation, and assessment of spring phenology in North
 1148 America estimated from remote sensing for 1982-2006, *Global Change Biol*, 15,
 1149 2335-2359, doi:10.1111/J.1365-2486.2009.01910.X, 2009.
 1150 White, M. A., Thornton, P. E., and Running, S. W.: A continental phenology model for
 1151 monitoring vegetation responses to interannual climatic variability, *Global*
 1152 *Biogeochem Cy*, 11, 217-234, doi:10.1029/97gb00330, 1997.
 1153 Zhang, X. Y., Tarpley, D., and Sullivan, J. T.: Diverse responses of vegetation phenology
 1154 to a warming climate, *Geophys. Res. Lett.*, 34, L19405, doi:10.1029/2007gl031447,
 1155 2007.
 1156 Zhao, M. F., Peng, C. H., Xiang, W. H., Deng, X. W., Tian, D. L., Zhou, X. L., Yu, G.
 1157 R., He, H. L., and Zhao, Z. H.: Plant phenological modeling and its application in
 1158 global climate change research: overview and future challenges, *Environ Rev*, 21, 1-
 1159 14, doi:10.1139/Er-2012-0036, 2013.
 1160 Zhu, W. Q., Tian, H. Q., Xu, X. F., Pan, Y. Z., Chen, G. S., and Lin, W. P.: Extension of
 1161 the growing season due to delayed autumn over mid and high latitudes in North
 1162 America during 1982-2006, *Global Ecol Biogeogr*, 21, 260-271, doi:10.1111/J.1466-
 1163 8238.2011.00675.X, 2012.
 1164
 1165

Xu Yue 7/2/15 11:18 AM
 Formatted: Justified

1166
1167
1168
1169
1170
1171

Table 1. Summary of studies estimating phenology trend in U.S. for at least 20 years.

	Studies	Period	Data sets	Results
A	Reed (2006)	1982-2003	NDVI (AVHRR)	Spring: scattered trends towards advance and delay. Autumn: significant delay in the Northeast, and Northwest
B	Zhang et al. (2007)	1982-2005	NDVI (AVHRR)	Spring: advance in Center and East but delay in the North and Southeast.
C	Julien and Sobrino (2009)	1981-2003	NDVI (GIMMS)	Spring: advance in the West and East, no trend in the Northeast and Southeast, delay in the North. Autumn: advance almost everywhere.
D	White et al. (2009)	1982-2006	NDVI (AVHRR)	Spring: no evidence for time trends for most areas with significant delay in the North.
E	Jeong et al. (2011)	1982-2008	NDVI (AVHRR)	Spring: no evidence for time trends for most areas. Autumn: delay in the West, North, Northeast, and Southeast (except Center).
F	Dragoni and Rahman (2012)	1989-2008	NDVI (AVHRR)	Autumn: significant delay in Northeast but insignificant changes in the East and North
G	Zhu et al. (2012)	1982-2006	NDVI (GIMMS)	Spring: significant delay in the Center and East. Autumn: significant delay in the West but almost no changes in the East Spring: advance in the East and Southeast, but almost no changes in Northeast, North, and West.
H	Buitenwerf et al. (2015)	1981-2012	NDVI (GIMMS)	Autumn: delay in the East, Southeast, Northeast, and North, but almost no changes in the West.

1172
1173
1174
1175
1176
1177

NDVI: Normalized Difference Vegetation Index
AVHRR: Advanced Very High-Resolution Radiometers
GIMMS: Global Inventory Mapping and Monitoring Studies

Xu Yue 7/2/15 11:18 AM

Deleted: .

Xu Yue 7/2/15 11:18 AM

Deleted: Zhang et al. (2007)

Xu Yue 7/2/15 11:18 AM

Formatted Table

1180
1181
1182
1183
1184
1185
1186

1187
1188
1189
1190
1191

Table 2. Ground measurements of leaf area index (LAI) used to calibrate the phenology model. The location of these sites is denoted on Fig. 1.

Site	Name	Latitude	Longitude	Years	n^a	Reference
US-Ha1	Harvard Forest	42.54°N	72.17°W	1998-2008	68	Urbanski et al. (2007)
US-UMB	Univ. of Michigan Biological Station	45.56°N	84.71°W	1999-2007	116	Gough et al. (2008)
US-MMS	Morgan Monroe State Forest	39.32°N	86.41°W	1999-2010	207	Schmid et al. (2000)
US-MOz	Missouri Ozark	38.74°N	92.2°W	2006-2012	149	Gu et al. (2006)

^a n denotes the number of records

1192
1193
1194
1195
1196
1197
1198
1199

1200
1201
1202
1203

Table 3. Phenological and climatological parameters for four deciduous forest sites predicted by segmented regressions (Fig. S1) and the selected phenology models (S9+A4, refer to Table 5).

Sites	Annual Temp (°C)	Budburst (day of year)		Grow length (days)		Offset start (day of year)		Offset length (days)	
		LAI-derived	Model	LAI-derived	Model	LAI-derived	Model	LAI-derived	Model
US-Ha1	8.0	125	122 ± 5	30	47 ± 7	271	270 ± 10	39	33 ± 4
US-JMB	7.2	124	125 ± 4	42	45 ± 5	273	265 ± 21	34	34 ± 4
US-MMS	12.3	100	103 ± 5	51	39 ± 4	276	275 ± 6	35	40 ± 4
US-MOz	13.3	103	102 ± 4	41	35 ± 3	270	275 ± 5	45	42 ± 4

Xu Yue 7/2/15 11:18 AM

Formatted Table

Xu Yue 7/2/15 11:18 AM

Deleted: Reg.

Xu Yue 7/2/15 11:18 AM

Deleted: Reg.

Xu Yue 7/2/15 11:18 AM

Deleted: Reg.

Xu Yue 7/2/15 11:18 AM

Deleted: Reg.

1208
1209
1210
1211

1212 **Table 4.** Ground phenology measurements of deciduous trees used to validate the model.

1213

Site / network	Category	Duration	Sites	Species	Trees	<i>n</i>	Reference / Link
Harvard Forest	Dates	1990-2012	1	16	56	32393	O'Keefe (2000)
Hubbard Brook	Dates	1989-2012	1	3	27	1081	Bailey (2014)
	LAI	1999-2012	1	N/A	N/A	171	http://ameriflux.ornl.gov/
US-UMB	Dates	1999-2012	1	5	66	259	Gough et al. (2008)
	Photos	2005-2012	1	N/A	N/A	1265	Sonnentag et al. (2012)
	LAI	1999-2012	1	N/A	N/A	207	http://ameriflux.ornl.gov/
US-MMS	Dates	2000-2004	1	N/A	N/A	4	http://ameriflux.ornl.gov/
	Photos	2008-2012	1	N/A	N/A	1480	Sonnentag et al. (2012)
National Phenology Network 1 ^a	Dates	2011-2012	588	52	1986	29280	https://www.usanpn.org/
National Phenology Network 2 ^b	Dates	2004-2012	167	7	195	4231	https://www.usanpn.org/
North American Lilac Network	Dates	1982-2003	392	2	N/A	5072	Schwartz and Reiter (2000)

1214

1215 ^a Data used to evaluate spatial distribution of simulated phenology. Detailed species
1216 information is listed in Table S1.

1217 ^b Data used to evaluate temporal variation of simulated phenology. Detailed species
1218 information is listed in Table S2.

1219

1220

1221

Table 5. Summary of phenology models with fit parameters calibrated against the long-term phenology at four U.S. deciduous sites. The detailed parameters for the selected models, S9 and A4, are summarized in Table S3. [Optimized parameters for other models are summarized in Table S4.](#)

ID	Model Name	Category	Fixed Parameters	Fit Parameters
S1	Sequential CF1	Spring	0	$5 (t_1, T_f, T_c, C^*, F^*)$
S2	Sequential CF2	Spring	0	$4 (t_1, T_c, C^*, F^*)$
S3	Parallel1 CF1	Spring	0	$6 (t_1, T_f, T_c, C^*, a, b)$
S4	Parallel1 CF2	Spring	0	$5 (t_1, T_c, C^*, a, b)$
S5	Parallel2 CF1	Spring	0	$6 (t_1, T_f, T_c, C^*, a, b); t_2 = t_1$
S6	Parallel2 CF2	Spring	0	$5 (t_1, T_c, C^*, a, b); t_2 = t_1$
S7	Alternating CF1	Spring	0	$4 (t_1, T_c, a, b); T_f = T_c; t_2 = t_1$
S8	Alternating CF1 t1 fixed	Spring	1 (t_1)	$3 (T_c, a, b); T_f = T_c; t_2 = t_1$
S9	Alternating CF1 modified	Spring	3 (t_1, T_c, r)	$2 (a, b); T_f = T_c; t_2 = t_1$
A1	Jolly-2005 Origin	Autumn	4 (T_i, T_x, P_i, P_x)	0
A2	Jolly-2005 Adjusted	Autumn	0	$4 (T_i, T_x, P_i, P_x)$
A3	Delpierre-2009	Autumn	0	$6 (P_{\text{start}}, T_b, x, y, Y_{\text{crit}}, L_f)$
A4	CDD-photoperiod	Autumn	2 (t_3, T_b)	$4 (F_s, L_f, P_i, P_x)$

Figure captions

Figure 1. Simulation of spring and autumn phenology at four U.S. deciduous broadleaf forest (DBF) sites. The map shows the fraction of U.S. DBF derived from the Advanced Very High Resolution Radiometer (AVHRR). The area with >3% coverage is the domain for this study. Five triangles indicate the locations of sites whose long-term measurements of meteorology and phenology are used for the calibration and/or validation of the model: Harvard Forest (US-Ha1), Hubbard Brook Forest (US-HB1), Morgan-Monroe State Forest (US-MMS), University of Michigan Biological Station Forest (US-UMB), and Missouri Ozark Forest (US-MOz). Phenological dates are recorded at US-Ha1 and US-HB1 during 1992-2012. Measurements of leaf area index (LAI) and photos are used to derive phenology at US-UMB and US-MMS for 1999-2012. Derived phenological dates at US-MOz are used for model calibration but not validation and are not shown here. At each site, two simulations are performed with the spring model S9 and autumn model A4 (refer to Figure 2), driven by temperatures from either the *in situ* measurements (blue) or the Modern Era Retrospective-Analysis (MERRA) reanalysis (green). Trend of each time series (units: day yr⁻¹) is shown with colors indicating results from observations (red) and simulations (blue or green). Significant trends ($p<0.05$) are marked with asterisks.

Fig. 2. Comparison of model performance in the prediction of phenological dates at four U.S. DBF sites among (top) nine spring phenology models and (bottom) four autumn phenology models. The statistical metrics are correlation coefficient, root-mean-square error (RMSE), and the Akaike Information Criterion (AIC). Each point represents the mean values of the statistical metrics at four sites for one model. The error bar represents the range of the metrics. Each model uses the optimized parameters as summarized in Table 5 for the prediction. The red ones are the models used for the continental predictions. Detailed predictions at each site are shown in Figs. S4-S11. The values of correlation coefficients, RMSE, and AIC are summarized in Table S5.

Xu Yue 7/2/15 11:18 AM

Deleted: Two

Xu Yue 7/2/15 11:18 AM

Deleted: literature-based phenology model,

Xu Yue 7/2/15 11:18 AM

Deleted: .

Xu Yue 7/2/15 11:18 AM

Deleted: .

Xu Yue 7/2/15 11:18 AM

Formatted: Font:Bold

Xu Yue 7/2/15 11:18 AM

Deleted: ,

Xu Yue 7/2/15 11:18 AM

Deleted: .

Figure 3. Comparison of the simulated (a, b) budburst and (c, d) dormancy dates with *in situ* observations (colored circles) from the USA National Phenology Network for 2011-2012. Simulations are performed with the spring model S9 and autumn model A4. The number of the sites and the correlation coefficients are shown in the scatter plots. The separate evaluations in 2011 and 2012 are shown in Figs. S12 and S13. [The coverage of colored patches in \(a\) and \(c\) differs from that in Figure 1 because values at and beyond the low end of color scales have been shown in white. The number of the sites and the correlation coefficients are shown in the scatter plots. The number of sites shown in the plot is much fewer than the total of 588 because only a small portion of the sites has continuous records for both years and a stringent screening process is applied to derive phenological dates \(see Supplement\). The separate evaluations in 2011 and 2012 are shown in Figs. S12-S13.](#)

Figure 4. Correlations (circles) between the predicted budburst dates and observed first-bloom dates from the North American Lilac Network (circle) and first-leaf dates from the USA National Phenology Network (squares). Simulations are performed with the spring model S9. The correlation coefficients are calculated for individual trees with at least 6 years of observations during 1982-2012. Correlations with $p < 0.1$ are denoted with filled symbols.

Figure 5. Trend in the simulated (a) budburst and (b) dormancy dates for deciduous forests in the U.S. during 1982-2012. Simulations are performed with the spring model S9 and autumn model A4. The results are shown only for the grid squares where the fraction of deciduous forest is larger than 3%. Significant trends ($p < 0.05$) are denoted with dots.

Figure 6. The (a) difference and (b) its relative change in the growth length for U.S. deciduous forests between 2000s and 1980s. Significant changes ($p < 0.05$) are denoted with dots.

Xu Yue 7/2/15 11:18 AM

Deleted: and autumn model A4.

Xu Yue 7/2/15 11:18 AM

Deleted: 2

Xu Yue 7/2/15 11:18 AM

Deleted: Trend of surface air temperature for (a, b) January, (c, d) April, and (e, f) September over deciduous forest during 1982-2012. The temperature data are from (a, c, e) MERRA reanalyses and (b, d, f) USHCN Network. Significant trends ($p < 0.05$) are denoted with dots (a, c, e) or filled circles (b, d, f). The trends for 2000-2012 are presented in Fig. S15. [\[6\]](#)

Figure 7. Comparison of phenology trend over U.S. for (a) spring and (b) autumn estimated by different studies. The U.S. domain is divided into six patches to represent different geographic areas: west, north, northeast, center, east, and southeast. In each patch, different characters represent estimates from different studies over that area. A summary of all studies used for comparison is listed in Table 1. The color of a character indicate the sign of a trend as follows: red is positive, blue is negative, and black is zero or insignificant. A patch is hatched with the same color as the dominant trend if it is non-zero.

Figure 8. Interannual variations of phenological dates and their responses to temperature changes during 1992-2011 for each DBF species at Harvard Forest. The year-to-year (a) budburst and (b) dormancy onset dates are presented for species with observations available for at least 20 years. Values for red oak (*Quercus Rubra*, in red), red maple (*Acer Rubrum*, in green), and the ensemble average (in blue) based on basal area are highlighted in bold. The differences of species-specific dates relative to the ensembles are presented in (c) for budburst and (d) for dormancy onset. Temperature sensitivity of (e) budburst is calculated as the regressions between year-to-year budburst dates and March-April temperature. Similarly, regressions between dormancy onset dates and September temperature is calculated as the temperature sensitivity of (f) autumn phenology. For the middle and bottom panels, positive values are marked as red while negative ones are in blue. Significant ($p < 0.05$) temperature sensitivity in bottom panel is denoted with filled bar. Full names of species abbreviations are listed in Table S1, except for alternated-leaved dogwood (*Cornus Alterniflora* short as COAL) and Hawthorne (*Crataegus Sp* short as CRSP).

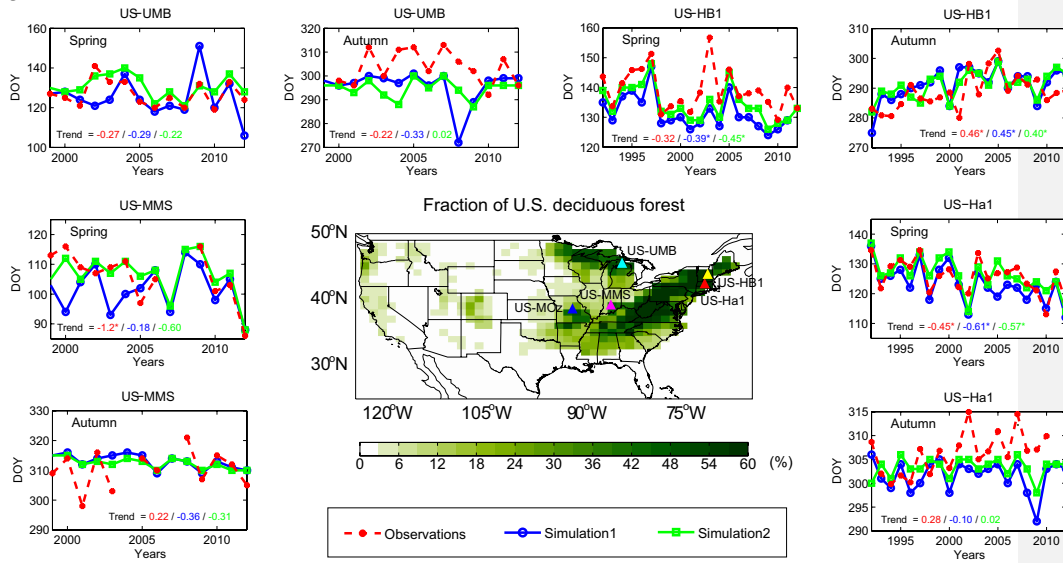
Xu Yue 7/2/15 11:18 AM

Deleted: 8

Xu Yue 7/2/15 11:18 AM

Deleted: 9

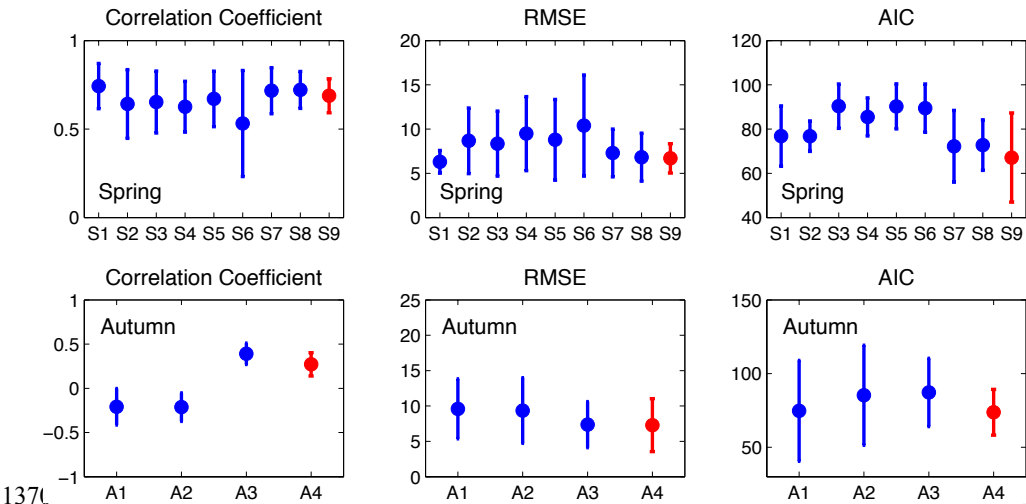
1342
1343
1344



1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365

Fig. 1 Simulation of spring and autumn phenology at four U.S. deciduous broadleaf forest (DBF) sites. The map shows the fraction of U.S. DBF derived from the Advanced Very High Resolution Radiometer (AVHRR). The area with >3% coverage is the domain for this study. Five triangles indicate the locations of sites whose long-term measurements of meteorology and phenology are used for the calibration and/or validation of the model: Harvard Forest (US-Ha1), Hubbard Brook Forest (US-HB1), Morgan-Monroe State Forest (US-MMS), University of Michigan Biological Station Forest (US-UMB), and Missouri Ozark Forest (US-MOz). Phenological dates are recorded at US-Ha1 and US-HB1 during 1992-2012. Measurements of leaf area index (LAI) and photos are used to derive phenology at US-UMB and US-MMS for 1999-2012. Derived phenological dates at US-MOz are used for model calibration but not validation and are not shown here. At each site, two simulations are performed with the spring model S9 and autumn model A4 (refer to Figure 2), driven by temperatures from either the *in situ* measurements (blue) or the Modern Era Retrospective-Analysis (MERRA) reanalysis (green). Trend of each time series (units: day yr⁻¹) is shown with colors indicating results from observations (red) and simulations (blue or green). Significant trends ($p < 0.05$) are marked with asterisks.

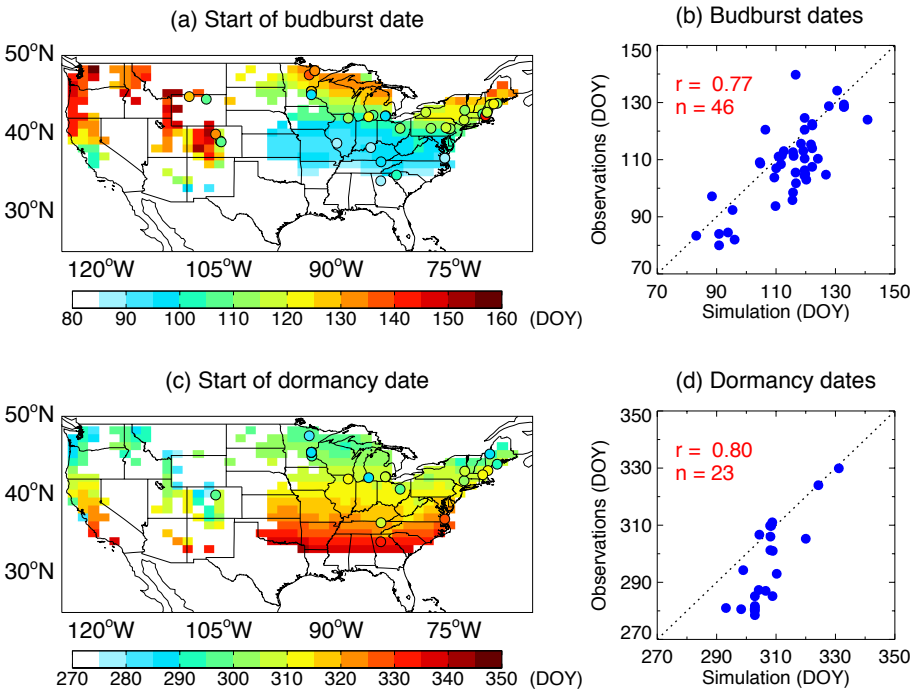
1366
1367
1368
1369



1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384

Fig. 2 Comparison of model performance in the prediction of phenological dates at four U.S. DBF sites among (top) nine spring phenology models and (bottom) four autumn phenology models. The statistical metrics are correlation coefficient, root-mean-square error (RMSE), and the Akaike Information Criterion (AIC). Each point represents the mean values of the statistical metrics at four sites for one model. The error bar represents the range of the metrics. Each model uses the optimized parameters as summarized in Table 5 for the prediction. The red ones are the models used for the continental predictions. Detailed predictions at each site are shown in Figs. S4-S11. The values of correlation coefficients, RMSE, and AIC are summarized in Table S5.

1385
1386
1387



1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400

Fig. 3 Comparison of the simulated (a, b) budburst and (c, d) dormancy dates with *in situ* observations (colored circles) from the USA National Phenology Network for 2011-2012. Simulations are performed with the spring model S9 and autumn model A4. The coverage of colored patches in (a) and (c) differs from that in Figure 1 because values at and beyond the low end of color scales have been shown in white. The number of the sites and the correlation coefficients are shown in the scatter plots. The number of sites shown in the plot is much fewer than the total of 588 because only a small portion of the sites has continuous records for both years and a stringent screening process is applied to derive phenological dates (see Supplement). The separate evaluations in 2011 and 2012 are shown in Figs. S12-S13.

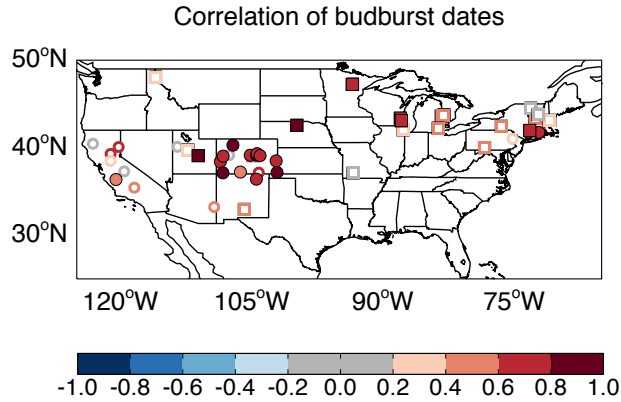
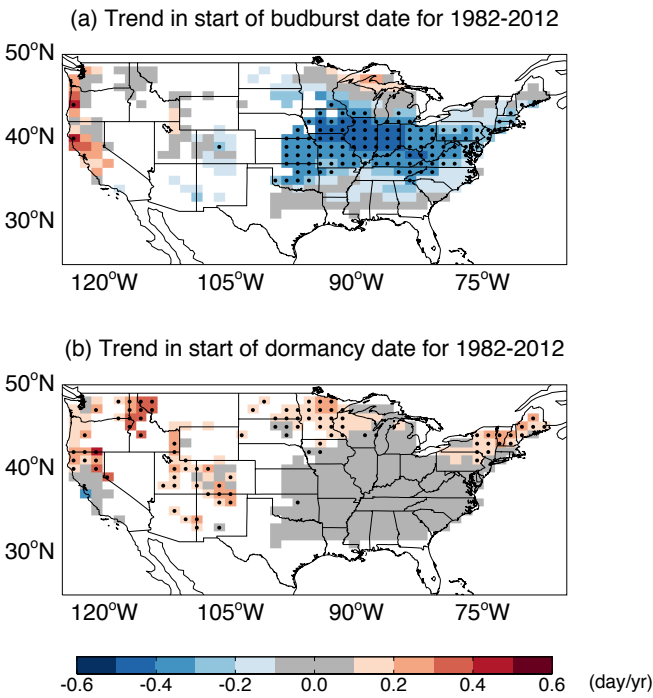


Fig. 4 Correlations (circles) between the predicted budburst dates and observed first-bloom dates from the North American Lilac Network (circle) and first-leaf dates from the USA National Phenology Network (squares). Simulations are performed with the spring model S9. The correlation coefficients are calculated for individual trees with at least 6 years of observations during 1982-2012. Correlations with $p < 0.1$ are denoted with filled symbols.

1433
1434



1435
1436
1437
1438
1439
1440
1441
1442

Fig. 5 Trend in the simulated (a) budburst and (b) dormancy dates for deciduous forests in the U.S. during 1982-2012. Simulations are performed with the spring model S9 and autumn model A4. The results are shown only for the grid squares where the fraction of deciduous forest is larger than 3%. Significant trends ($p<0.05$) are denoted with dots.

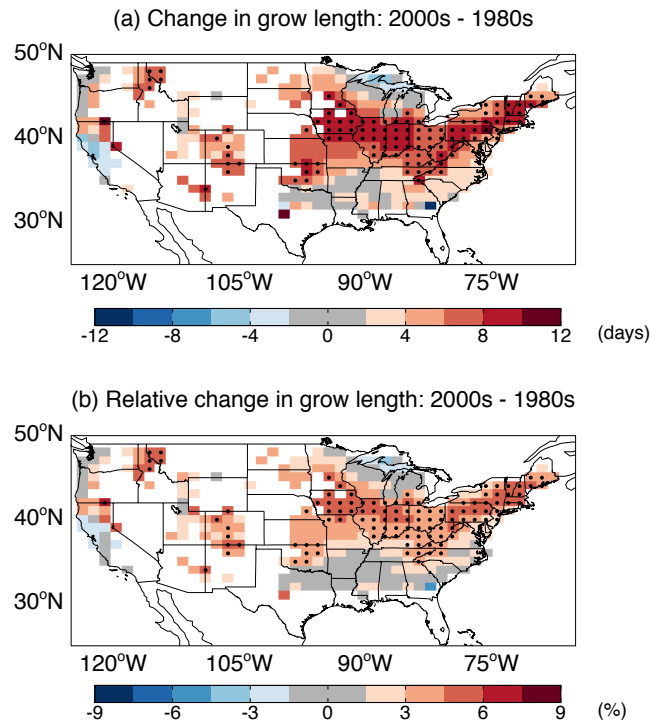


Fig. 6 The (a) difference and (b) its relative change in the growth length for U.S. deciduous forests between 2000s and 1980s. Significant changes ($p < 0.05$) are denoted with dots.

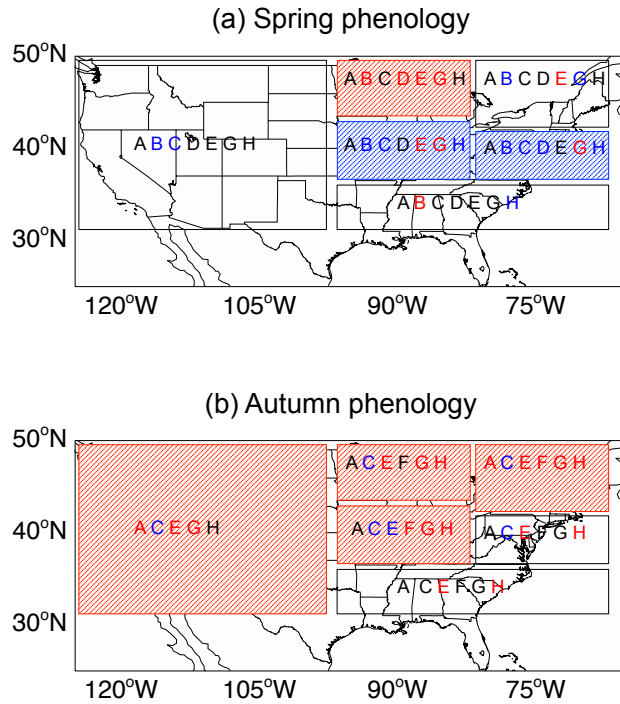


Fig. 7 Comparison of phenology trend over U.S. for (a) spring and (b) autumn estimated by different studies. The U.S. domain is divided into six patches to represent different geographic areas: west, north, northeast, center, east, and southeast. In each patch, different characters represent estimates from different studies over that area. A summary of all studies used for comparison is listed in Table 1. The color of a character indicate the sign of a trend as follows: red is positive, blue is negative, and black is zero or insignificant. A patch is hatched with the same color as the dominant trend if it is non-zero.

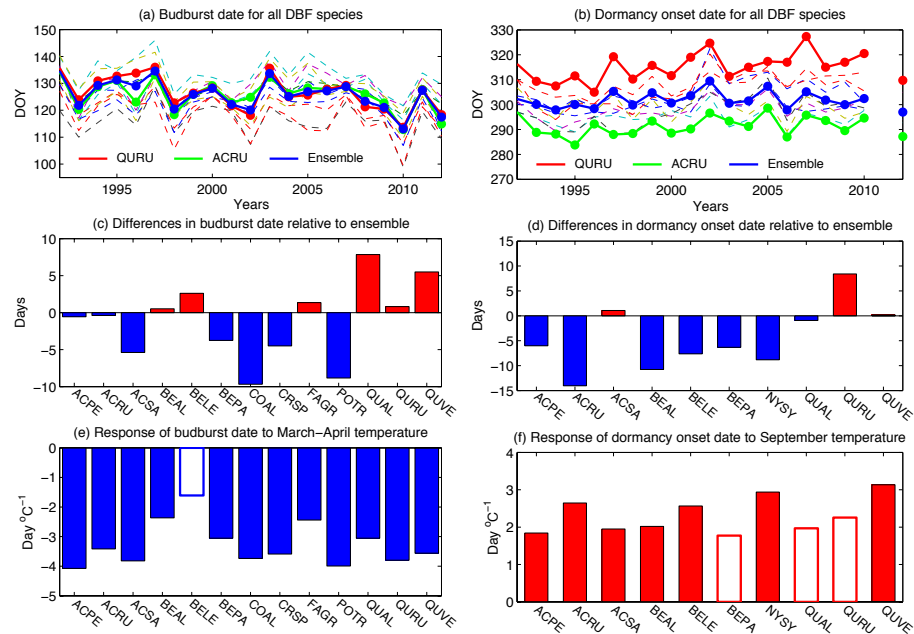


Fig. 8 Interannual variations of phenological dates and their responses to temperature changes during 1992-2011 for each DBF species at Harvard Forest. The year-to-year (a) budburst and (b) dormancy onset dates are presented for species with observations available for at least 20 years. Values for red oak (*Quercus Rubra*, in red), red maple (*Acer Rubrum*, in green), and the ensemble average (in blue) based on basal area are highlighted in bold. The differences of species-specific dates relative to the ensembles are presented in (c) for budburst and (d) for dormancy onset. Temperature sensitivity of (e) budburst is calculated as the regressions between year-to-year budburst dates and March-April temperature. Similarly, regressions between dormancy onset dates and September temperature is calculated as the temperature sensitivity of (f) autumn phenology. For the middle and bottom panels, positive values are marked as red while negative ones are in blue. Significant ($p < 0.05$) temperature sensitivity in bottom panel is denoted with filled bar. Full names of species abbreviations are listed in Table S1, except for alternated-leaved dogwood (*Cornus Alterniflora* short as COAL) and Hawthorne (*Crataegus Sp* short as CRSP).