A Comparison of Land Surface Phenology in the Northern Hemisphere Derived from Satellite Remote Sensing and the Community Land Model

XIAOLU LI,[a] ELI MELAAS,[b] CARLOS M. CARRILLO,[a] TOBY AULT,[a] ANDREW D. RICHARDSON,[c,d] PETER LAWRENCE,[e] MARK A. FRIEDL,[b] BIJAN SEYEDNASROLLAH,[c,d] DAVID M. LAWRENCE,[e] AND ADAM M. YOUNG[c,d]

[a] *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York*
[b] *Department of Earth and Environment, Boston University, Boston, Massachusetts*
[c] *School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona*
[d] *Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona*
[e] *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 28 August 2021, in final form 31 December 2021)

ABSTRACT: Large-scale changes in the state of the land surface affect the circulation of the atmosphere and the structure and function of ecosystems alike. As global temperatures increase and regional climates change, the timing of key plant phenophase changes are likely to shift as well. Here we evaluate a suite of phenometrics designed to facilitate an "apples to apples" comparison between remote sensing products and climate model output. Specifically, we derive day-of-year (DOY) thresholds of leaf area index (LAI) from both remote sensing and the Community Land Model (CLM) over the Northern Hemisphere. This systematic approach to comparing phenologically relevant variables reveals appreciable differences in both LAI seasonal cycle and spring onset timing between model simulated phenology and satellite records. For example, phenological spring onset in the model occurs on average 30 days later than observed, especially for evergreen plant functional types. The disagreement in phenology can result in a mean bias of approximately 5% of the total estimated Northern Hemisphere NPP. Further, while the more recent version of CLM (v5.0) exhibits seasonal mean LAI values that are in closer agreement with satellite data than its predecessor (CLM4.5), LAI seasonal cycles in CLM5.0 exhibit poorer agreement. Therefore, despite broad improvements for a range of states and fluxes from CLM4.5 to CLM5.0, degradation of plant phenology occurs in CLM5.0. Therefore, any coupling between the land surface and the atmosphere that depends on vegetation state might not be fully captured by the existing generation of the model. We also discuss several avenues for improving the fidelity between observations and model simulations.

KEYWORDS: Biosphere-atmosphere interaction; Carbon cycle; Climate variability; Remote sensing; Land surface model; Spring season; Vegetation; Vegetation-atmosphere interactions

1. Introduction

In temperate and boreal regions, plant phenology modulates the terrestrial carbon budget by governing the onset and duration of the growing season (e.g., Morisette et al. 2009; Richardson et al. 2009, 2010). Human-induced increases in temperature, as well as changes in precipitation, will likely modify plant phenology in the future, which in turn will affect the net storage of carbon on land (e.g., Morisette et al. 2009; Richardson et al. 2013) among a multitude of other changes (e.g., Schwartz 1992; Xu et al. 2020). For example, warmer average global temperatures could favor earlier spring onset dates and longer growing seasons, and hence lead to a larger terrestrial carbon sink. Alternatively, earlier springs and hotter, drier summers could also impose new stresses, including droughts, wildfires, and insect outbreaks (e.g., Guillevic et al. 2002; Peñuelas and Filella 2009), which would increase terrestrial carbon fluxes into the atmosphere.

Land surface phenology exhibits considerable year-to-year variability (Schwartz et al. 2006; White et al. 2009; Richardson et al. 2013) and characterizing those fluctuations accurately across space and through time requires multiyear large-scale observations of ecosystems. Long time series of high-resolution and internally consistent satellite data products are therefore essential tools for documenting and describing spring onset variations at large scales during the recent historical period.

In addition to remote sensing, land surface models (LSMs) provide insights into the historical coupled land–atmosphere processes that govern water, carbon, and energy fluxes through the climate system. Recently, several papers have examined both site-level and large-scale annual cycles of leaf area index (LAI) and net primary production (NPP), finding large disagreements between model simulations and observations (e.g., Richardson et al. 2012; Mahowald et al. 2016; Scholze et al. 2017; Albergel et al. 2018; Klosterman et al. 2018; Peano et al. 2021). Other studies have also adopted data assimilation approaches to improve simulations of the biogeophysical and

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-21-0169.s1.

Corresponding author: Xiaolu Li, xl552@cornell.edu

DOI: 10.1175/JHM-D-21-0169.1

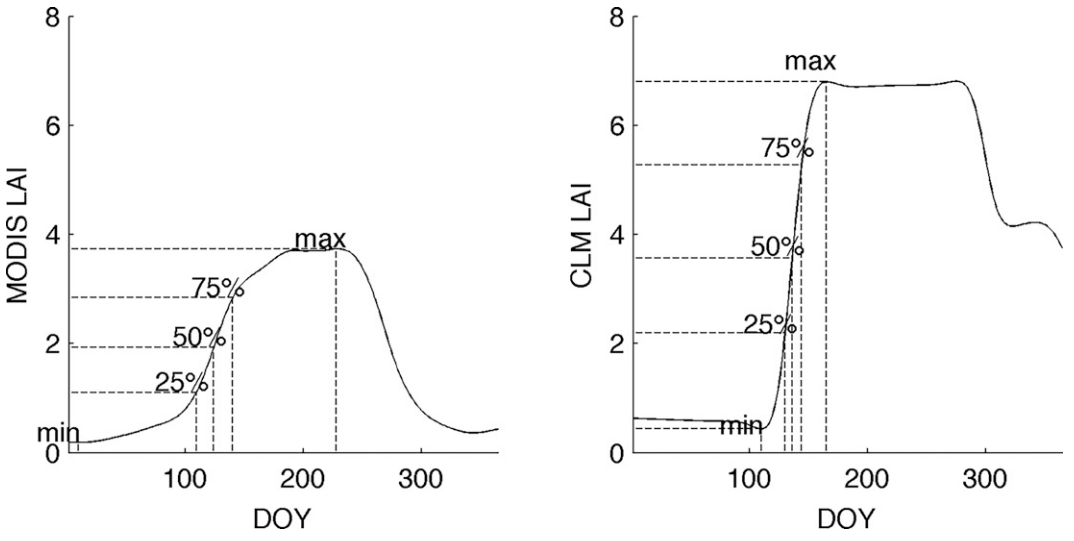

FIG. 1. Schematic diagram of how the threshold-based DOY indices are computed from (left) remote sensing and (right) CLM output.

biogeochemical processes (e.g., Sabater et al. 2008; Barbu et al. 2011; Albergel et al. 2017; Fox et al. 2018). However, most of these studies focused on small scales, a restricted number of plant functional types, or monthly averages.

The timing of spring onset and fall senescence in LSMs and their potential influences on land–atmosphere interactions have not yet been intensively analyzed, as additional obstacles complicate comparisons between model output and remote sensing data. For instance, the timing of spring onset varies by a few weeks from one year to the next, and identifying those variations requires vegetation data to be sampled at daily or weekly time intervals. Even when prognostic phenology routines are included in an Earth system model (ESM) simulation, the output from those routines is seldom archived at sufficiently high temporal resolution to compare day-of-year (DOY) metrics of spring onset with remote sensing products. Given these limitations, the *timing* of spring onset—as inferred from remote sensing and LSM simulations alike—has received relatively little attention in the literature to date.

The majority of previous studies evaluating phenology in the Community Land Model (CLM) and other land surface models have focused on comparisons at small spatial scales to develop parameters and characterize fundamental ecological processes, but large-scale comparisons are also important to understand model performance. Studies have used both in situ measurements and remote sensing at specific sites or for specific PFTs. For instance, Scholze et al. (2017) used a data assimilation approach at the site level and tuned parameterizations in land surface models to reproduce observed seasonal cycles. Dahlin et al. (2015, 2017) examined stress deciduous phenology in CLM in tropical drylands and found that a precipitation criterion is necessary to prevent rapid onset of growing seasons due to soil moisture fluctuations. Chen et al. (2016) implemented different spring onset triggers for seasonal deciduous trees in CLM and improved the model's simulation of productivity. While these studies are crucial for developing model parameters and evaluating ecological processes, accurately simulating the land surface is a critical component for future climate projections at the global scale. It is therefore important to evaluate the skill of CLM to simulate key aspects of spring "green up" across different climate zones at the hemispheric and continental scales that will be most important for future climate feedback and carbon sinks or sources.

In this study, we develop DOY indicators to characterize the progression of spring phenology in both remote sensing data and CLM simulations over the Northern Hemisphere. We then compare these indicators from MODIS to new runs from the CLM (Oleson et al. 2013; Lawrence et al. 2019). We ask 1) how well does CLM simulate the LAI seasonal cycle compared to MODIS estimates, 2) to what extent has the simulated phenology changed between CLM4.5 and CLM5.0 and what are the sources of these changes, and 3) how much do biases in CLM phenology affect the annual accumulated terrestrial carbon fluxes? By assessing the large-scale patterns of seasonal cycles of leaf phenology and its NPP influence over the Northern Hemisphere using metrics that can be readily computed from both data products, our approach therefore represents a framework for evaluating land surface phenology vegetation parameterizations and their response to atmospheric forcing.

2. Data and methods

a. LAI from remote sensing

The indices of primary interest here are all based on LAI, which is either derived from MODIS or calculated internally by CLM. We defined the dynamical range of LAI as the difference between minimum (winter) and maximum (summer) LAI each year. We then focused on the 25%, 50%, and 75% thresholds of the annual dynamical range of LAI (Fig. 1).

Using threshold-based indicators reduces the influence of land use change as well as differences in peak LAI from one year to the next [e.g., as in White et al. (2009), but applied to LAI instead of NDVI].

Our remote sensing data originate from the MODIS *Terra* MOD15A2H.v006 LAI product (Myneni et al. 2015), which is an 8-day composite dataset available at 500-m native spatial resolution. The algorithm uses a three-dimensional radiative transfer equation to estimate LAI based on surface reflectance and land cover type (Knyazikhin et al. 1999; Yan et al. 2016) and chooses the "best" pixel within the 8-day period from all *Terra* sensor measurements. We aggregated the raw MODIS LAI to the CLM finite-volume grid (nominally 1° × 1° latitude–longitude) in two steps. First, all "good quality" data points (i.e., points with a MODIS QA flag of "0") within a given CLM grid cell were averaged together to produce LAI fields sampled at the native 8-day time step. Second, a smoothing spline was fitted to the time series of each grid point for each year and used to interpolate the 8-day data to daily resolution. We then calculated the DOY time series of spring onset as indicated by the different LAI thresholds.

To characterize the agreement between CLM and MODIS spring onset timing, we used both simple (Pearson) correlation and scatterplots. For all simple correlation analysis, we adopted a 5% significance level and adjusted for false discovery (Benjamini and Hochberg 1995). We also used scatterplots and boxplots to assess how similar spring onset timing and other phenometrics are between different indices pairs.

b. *LAI from the Community Land Model*

CLM is the land component of the Community Earth System Model (CESM). It simulates comprehensive biogeophysical, biogeochemical, and hydrological processes at the land surface, permitting it to represent land–atmosphere fluxes of water, carbon, and energy (Oleson et al. 2013; Lawrence et al. 2019). It can be run in a "coupled" configuration, in which both CLM and the Community Atmosphere Model (CAM; Neale et al. 2010) exchange fluxes, or it can be forced with historical climate and meteorological reconstructions. Under this configuration, LAI varies through time as a function of meteorological forcing, parameterized plant phenology, and interactions with other processes modeled by CLM. Running CLM with historical boundary conditions requires a combination of gridded instrumental data products and reanalysis output to supply the requisite surface radiation, temperature, wind, and humidity fields, among others. Here, we used versions 4.5 and 5.0 of CLM (i.e., CLM4.5 and CLM5.0).

The phenology subroutines in CLM govern carbon and nitrogen fluxes for leaf development and litter fall for natural vegetation types (i.e., vegetation types that are not simulated by the crop module). These routines also partially regulate biogeophysical processes, like photosynthesis and canopy hydrology, by modifying LAI over the course of the year. There are three distinct phenology parameterizations in CLM—seasonal deciduous, stress deciduous, and evergreen—and each parameterization affects LAI in at least one of the model's 14 natural plant functional types (PFTs). For example, a growing degree-day (GDD) threshold triggers leaf emergence and growth in PFTs that use the seasonal deciduous phenology routine (White et al. 1997; Oleson et al. 2013; Lawrence et al. 2019). PFTs governed by stress deciduous phenology start growing only after their chilling requirements are met (except for tropical stress deciduous PFTs, e.g., broadleaf deciduous tropical trees, that do not have a chilling requirement) and in response to GDD thresholds and soil moisture (White et al. 1997; Oleson et al. 2013), as well as an antecedent precipitation requirement introduced in CLM5.0 (Dahlin et al. 2015; Lawrence et al. 2019). Evergreen phenology, in contrast, has a fixed background litter fall rate and no leaf onset/offset trigger (Oleson et al. 2013; Lawrence et al. 2019), so new leaf production is always occurring and depends on the current rate of photosynthesis and respiration. Area weight of the three different phenology plant functional types is shown in Fig. 2 and how the phenology types connect to the PFTs in CLM is listed in Table 1.

During each simulation year, PFTs using the deciduous phenology routine allocate some portion of their carbon and nitrogen for leaf development during the following year (Thornton and Zimmermann 2007; Oleson et al. 2013; Lawrence et al. 2019). Spring onset occurs when the environmental thresholds listed above are reached, and CLM begins to allocate stored carbon (and nitrogen) from the previous year to increase LAI over a fixed 30-day period. However, stress deciduous PFTs can have multiple growing seasons within one year, or a long growing season and no dormancy when conditions are favorable.

Usually, multiple PFTs will comprise the natural vegetation at a given grid cell, but the climate forcing used by each PFT will be uniform across that grid cell. Likewise, the changes in LAI at a given point will be the weighted average of all PFTs at that point (e.g., Fig. S1 in the online supplemental material). Importantly, grid cell LAI in CLM represents the average of multiple PFTs that are exposed to the exact same climate, whereas LAI in MODIS is the average over multiple land cover types, which may have geographical differences and varying microclimates. However, this variation within a grid cell is small relative to the total continental-to-global differences. We have also filtered out grid points with more than 50% crop PFTs as our focus here is on natural PFTs.

Interannual variability in the start of spring can be as large as 60 days, while the advancing trend of spring onset over the past few decades is on average 1.5 days per decade (e.g., Schwartz et al. 2006; Schwartz et al. 2013; Ault et al. 2015). As with MODIS, we therefore required daily LAI to accurately characterize spring onset during recent decades. By default, CLM only outputs monthly total leaf area index (TLAI) history files, therefore, we ran CLM4.5 and CLM5.0 and outputted daily TLAI and net primary production (NPP) values both as a grid cell average and for each PFT.

c. *Experimental design and atmospheric forcing*

We ran CLM (4.5 and 5.0) with the GSWP3 historical forcing dataset (Müller Schmied et al. 2016; http://hydro.iis.u-tokyo.ac.jp/GSWP3/exp1.html#boundary-conditions).

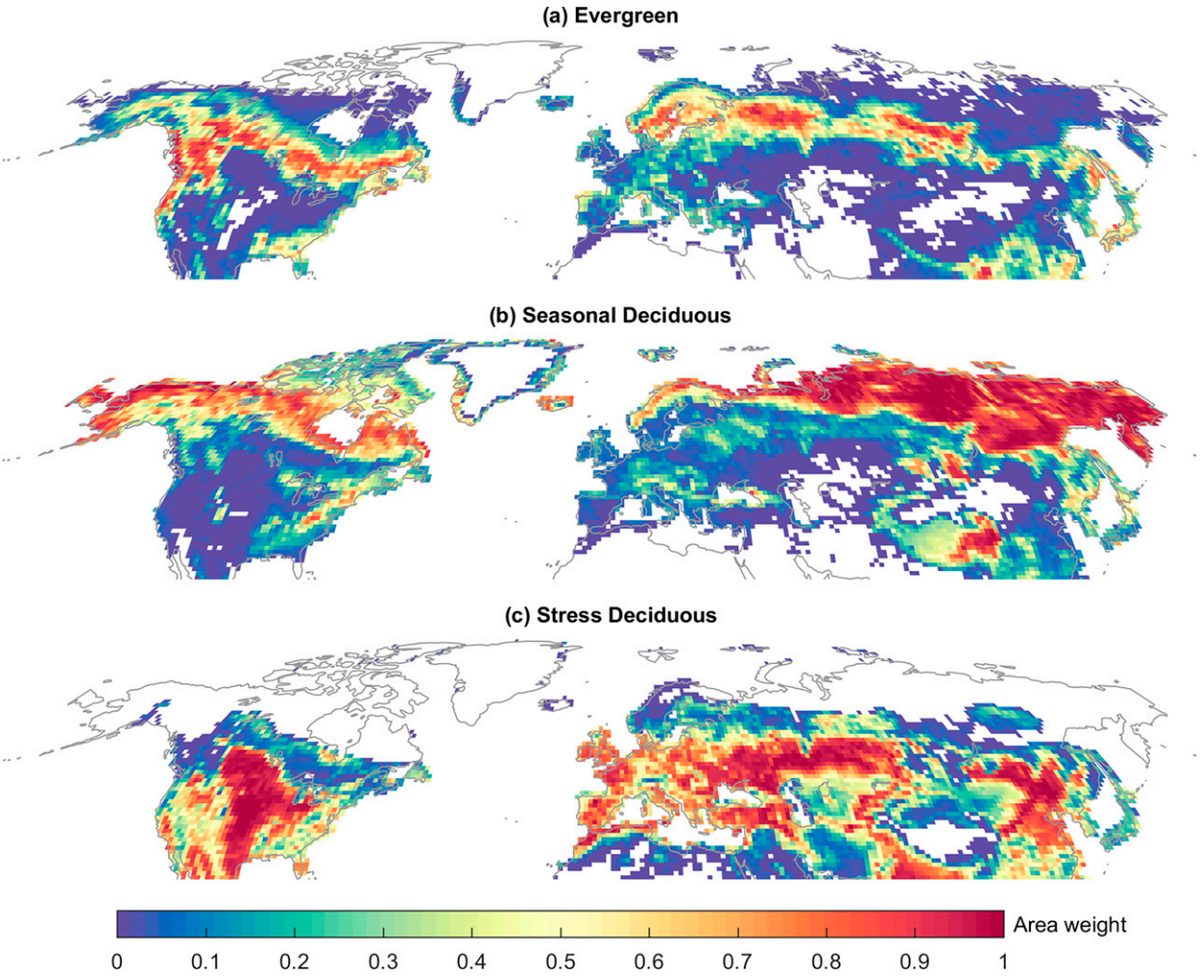

FIG. 2. Maps showing the fractional weight of different phenology types in CLM5.0 for the year 2000.

Briefly, GSWP3 data are dynamically downscaled from the Twentieth Century Reanalysis (20CR; Compo et al. 2011) and corrected using observations. Due to its temporal coverage, GSWP3 was used as boundary conditions over the period from 1 January 1970 through 31 December 2014. Although other boundary conditions have been used to examine the skill of CLM [e.g., CRUNCEP in Wang et al. (2014) and CRUJRA in the TRENDY project], our findings do not

TABLE 1. Summary information for the three different phenology subroutines used in CLM (first column) and the PFTs that invoke them (right columns).

	PFT			
Pheno PFT	Trees/forests	Shrub	Grassland	Cropland
Evergreen	Needleleaf evergreen temperate tree, needleleaf evergreen boreal tree, broadleaf evergreen tropical tree, broadleaf evergreen temperate tree	Broadleaf evergreen shrub		
Stress deciduous	Broadleaf deciduous tropical tree	Broadleaf deciduous temperate shrub	c3 nonarctic grass, c4 grass	c3 crop/crop1, c3 irrigated/crop2
Seasonal deciduous	Needleleaf deciduous boreal tree, broadleaf deciduous temperate tree, broadleaf deciduous boreal tree	Broadleaf deciduous boreal shrub	c3 arctic grass	

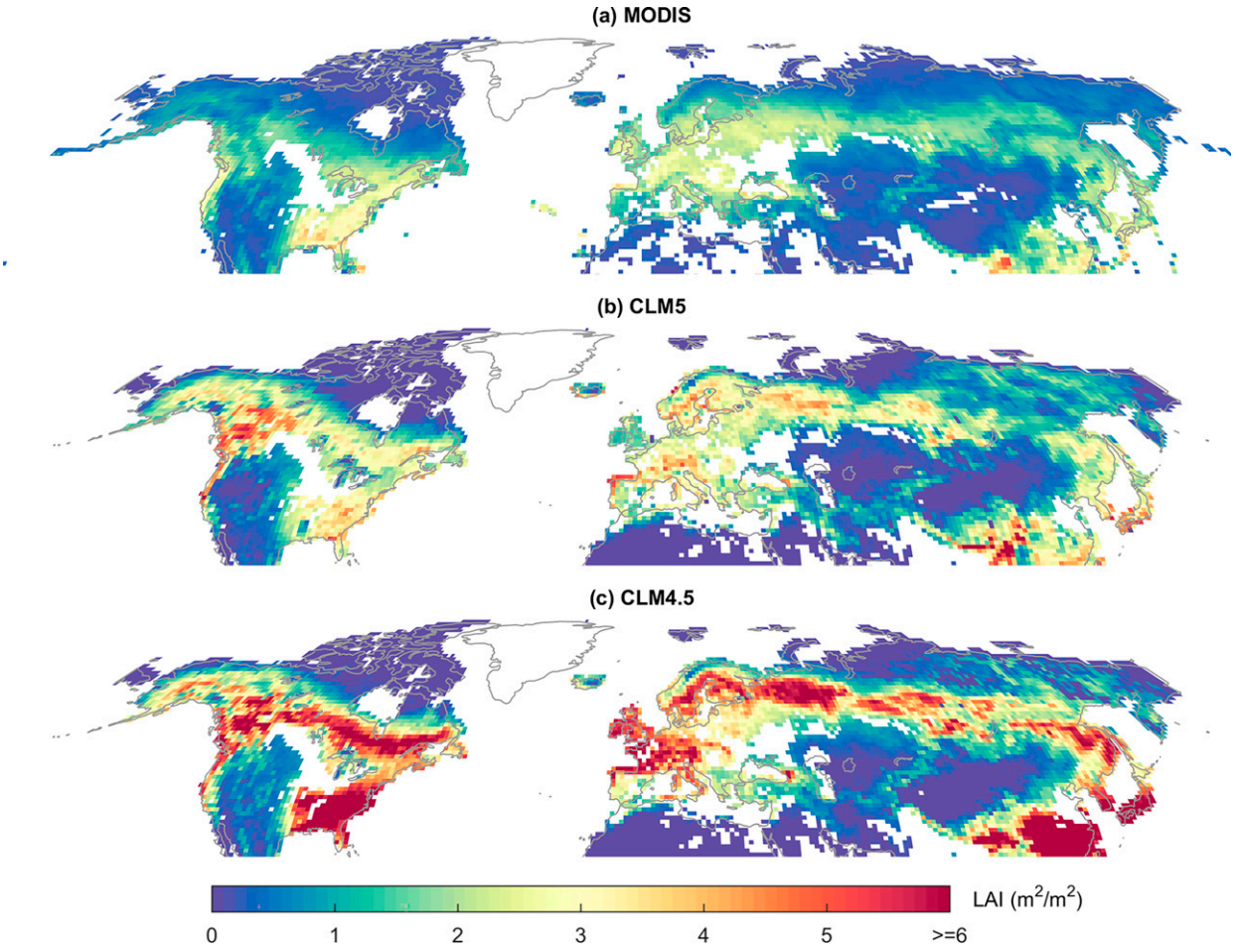

FIG. 3. Mean spring (April–June) LAI values in (a) MODIS, (b) CLM5.0, and (c) CLM4.5 averaged between 2003 and 2014.

depend strongly on the historical forcing data for the regions and variables of interest here. We discarded the first 33 years for spinup and used 2003–14 to compare CLM and MODIS over the Northern Hemisphere.

d. *Terrestrial productivity differences between MODIS and CLM*

To examine disagreement in the carbon cycle induced by the different leaf phenology in CLM and MODIS, we also computed the total net primary production (NPP) simulated by CLM during the differences in the duration of peak growing season between CLM and MODIS and named this indicator ΔNPP_{pheno}. We defined peak growing season as days within the year when LAI is above 75% of its annual dynamical range. We then calculated the difference between CLM and MODIS peak growing season and estimated the total NPP simulated by CLM during that difference window. CLM NPP is counted as positive when CLM LAI is within its peak growing season but MODIS is not, and negative vice versa. After computing the annual NPP difference induced by different peak growing windows in CLM and MODIS, we calculated ΔNPP_{pheno} by averaging across all years to characterize the potential influence of errors in modeled phenology on terrestrial carbon cycle simulations. We only examined NPP differences induced by differences in peak growing season length between CLM and MODIS to diminish the influence of LAI difference and focus on difference due to plant phenology. We also calculated how large ΔNPP_{pheno} is compared to the total annual NPP in CLM.

3. Results

a. *Spring onset timing*

Spring LAI values are generally lower in MODIS than in either version of CLM (Fig. 3 and Fig. S2). For example, boreal forests in the eastern North America typically exhibit LAI values between 2 and 3 $m^2\ m^{-2}$, as inferred from MODIS (Fig. 3a). In CLM4.5 these values are greater than 5 $m^2\ m^{-2}$ (Fig. 3c), while they are lower in CLM5.0 and closer to MODIS LAIs (about 4 $m^2\ m^{-2}$, Fig. 3b). Similar differences occur throughout the northern regions of North America and Eurasia, as well as in western Europe, the

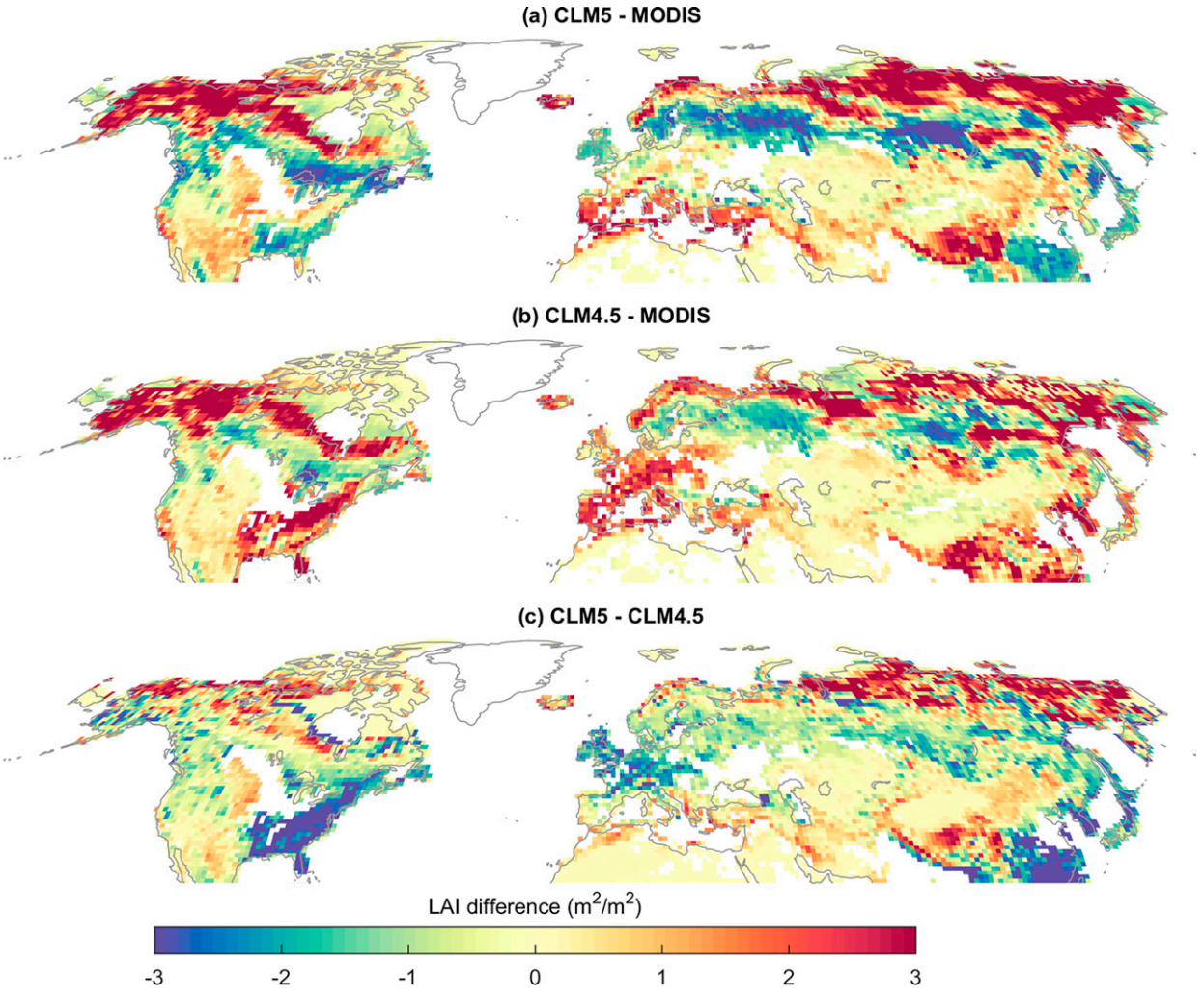

FIG. 4. Difference in mean annual range of LAI (annual maximum minus annual minimum) averaged between 2003 and 2014.

southeastern United States, and southeast China (Fig. 3). Although major improvements have been made from CLM4.5 to CLM5.0, especially over seasonal deciduous and evergreen PFTs, CLM spring LAIs are still larger than MODIS across phenology types in the Northern Hemisphere (Fig. S2). On average, difference between CLM and MODIS LAI is smaller over deciduous PFTs (less than 1 m^2 m^{-2} in CLM5.0 over seasonal deciduous PFTs) and larger over evergreen PFTs (larger than 2 m^2 m^{-2} in both CLM5.0 and CLM4.5; Fig. S2).

The annual dynamical range of LAI is also greater in CLM than in MODIS across most of the deciduous PFTs dominated regions (Figs. 2 and 4), except for seasonal deciduous PFTs in CLM5.0 (Fig. S3). For example, both CLM4.5 and CLM5.0 simulate a dynamical range that can be more than 3 m^2 m^{-2} larger than MODIS over northern Canada, northern Russia, and the Tibetan Plateau (Figs. 4a,b). Even though the overall magnitude decreases by around 1 m^2 m^{-2} in both spring LAI bias and deciduous LAI annual range bias between CLM4.5 and CLM5.0 (Figs. S2 and S3), the spatial pattern of bias in the dynamical range is similar for both model versions and may have even increased in some regions (blue colors in Fig. 4c). When compared to the annual amplitude in MODIS LAI, CLM4.5 also exhibits higher LAI amplitude in the Mediterranean, in southeast China, and in the southeastern United States (Fig. 4b). However, there are exceptions where the annual dynamical range in MODIS is greater than in CLM (blue regions in Figs. 4a,b), mostly in evergreen PFTs occupied regions and also over some seasonal deciduous PFTs (Fig. S3).

In general, the date when LAI reaches 50% of its annual amplitude occurs later in CLM than in MODIS (Figs. 5 and 6 and Figs. S4 and S5). For temperate and boreal regions, spring onset in CLM occurs 21.9 ± 26.8 days later in evergreen dominated regions and 8.9 ± 15.6 days later in seasonal deciduous dominated regions than in MODIS (Figs. 2 and 5 and Fig. S4). In the more southerly regions, which tend to be dominated by stress deciduous PFTs, CLM can reach the LAI 50% threshold 60 days (or more) later than MODIS (Figs. 5 and 6 and Fig. S5). Also, Mediterranean ecosystems on western coasts usually experience earlier leaf onset than

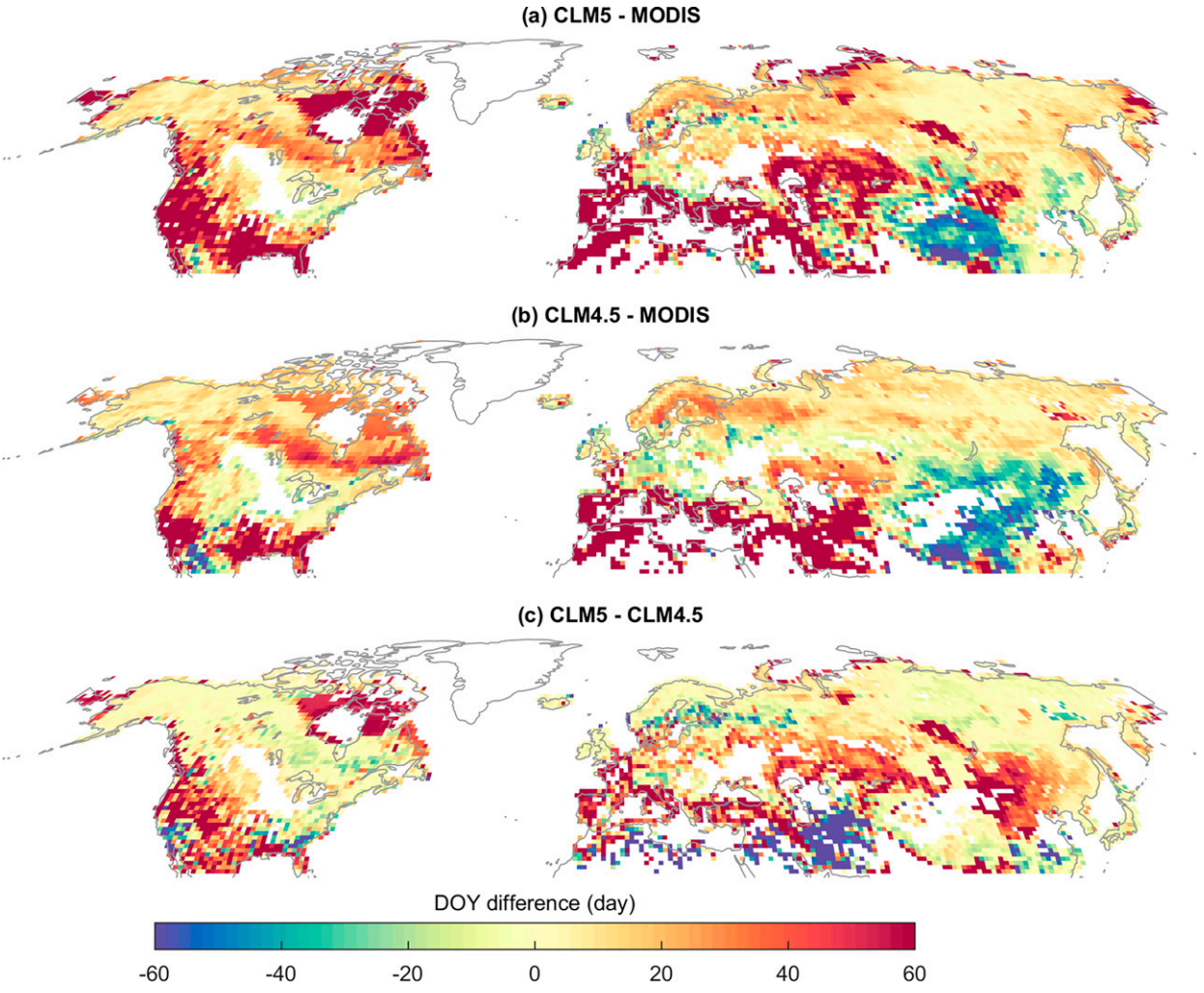

FIG. 5. Difference between mean DOY when LAI reaches the 50% threshold of the LAI annual dynamical range. The annual dynamical range of LAI is defined as the difference between minimum (winter) and maximum (summer) LAI each year.

eastern coasts at the same latitude in MODIS, but in CLM, these regions can be dominated by stress deciduous PFTs and experience not only later spring onset than in MODIS, but later start of spring than the eastern coasts in CLM (Fig. S4). Overall, agreements between CLM and MODIS DOY as well as between CLM4.5 and CLM5.0 increase northward (Fig. 6).

Outside of the subtropical and Mediterranean regions, both CLM and MODIS depict similar topographic and latitudinal gradients of spring onset at continental scales, with the earliest onset dates occurring in lower latitudes and lower elevations (Fig. S4). Yet despite these similarities in the gradient of spring onset, differences as large as 40 days (or more) emerge along the western edge of North America, the Rocky Mountains, northeastern Canada, Scandinavia, and western Russia (Fig. 5). In these areas, positive values (red shading in Figs. 5a,b) indicate that spring onset occurs more than 30 days later in CLM than in MODIS, and in general they are in regions where evergreen or stress deciduous PFTs dominate the grid cell (Figs. 2 and 5 and Table S1 and Fig. S5). Notably, spring onset is on average over 30 days later in CLM5.0 than in CLM4.5 over evergreen PFTs and over 20 days later over stress deciduous PFTs (Fig. S5).

Interannual variability in LAI 50% threshold DOY shows the strongest agreement between CLM and MODIS in high-latitude regions (north of 60°N) over Eurasia and North America (Fig. 7). In CLM4.5, midlatitude regions (around 40°–60°N) in North America and central Asia also exhibit strong (with correlation coefficients close to 0.5) and significant correlation with MODIS. In CLM5.0, although the absolute magnitude of LAI values is closer to those displayed in MODIS (Fig. 3 and Fig. S2), in low- to midlatitude regions (south of 60°N), correlations between LAI 50% threshold DOY from MODIS and CLM are mostly lower than 0.4 and not statistically significant, indicating that interannual variations in the start of spring disagree between these two datasets (Fig. 7a). Agreements between CLM5.0 and CLM4.5 are also high and significant in boreal regions (with correlation coefficients larger than 0.8), but only seasonal deciduous PFTs

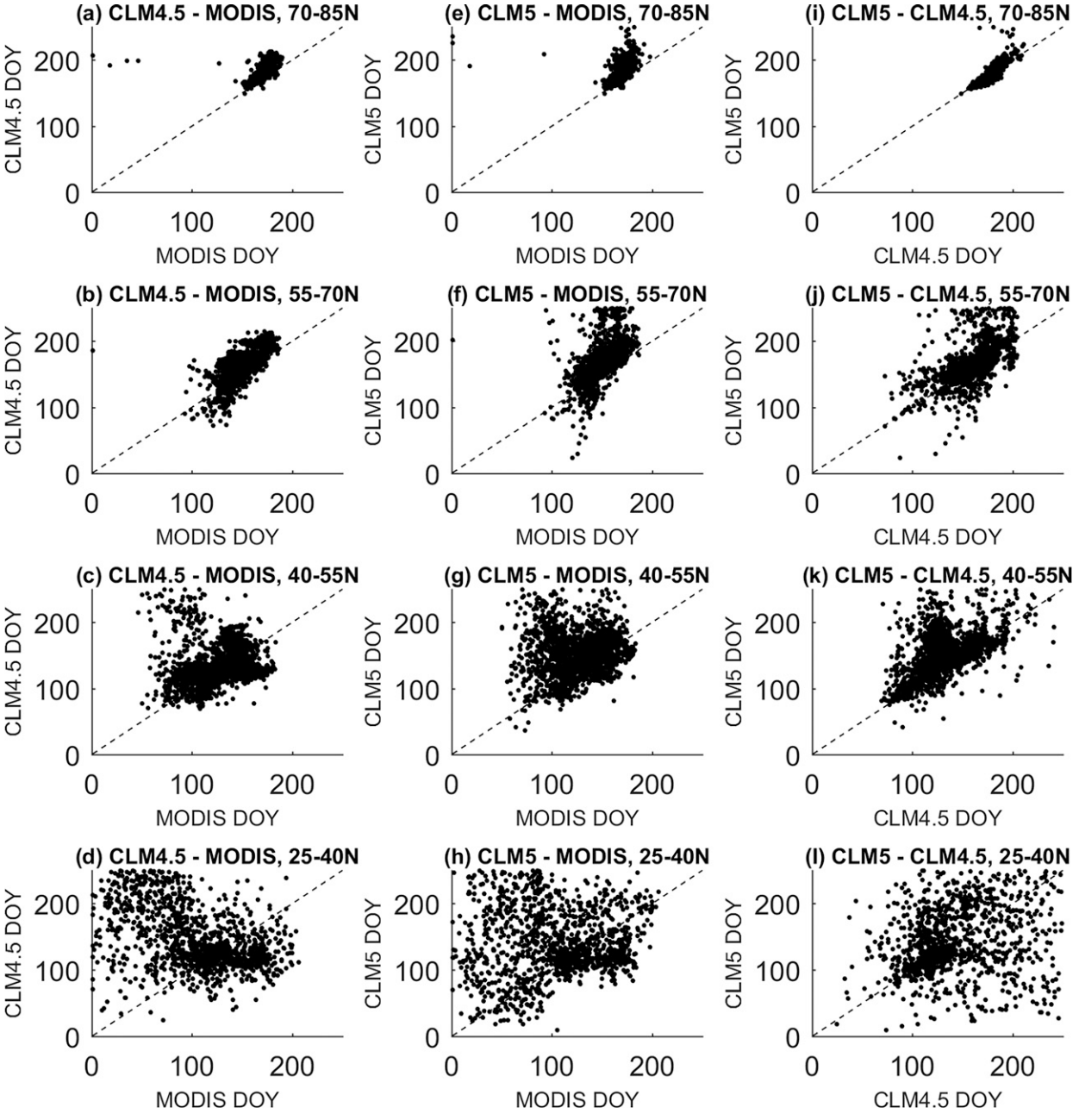

FIG. 6. Comparison of LAI 50% threshold DOYs derived from different datasets at different latitudes: (first row) 70°–85°N, (second row) 55°–70°N, (third row) 40°–55°N, and (fourth row) 25°–40°N (each band includes the upper/northern boundary and excludes the lower/southern boundary). (a)–(d) LAI 50% threshold DOYs derived from CLM4.5 and MODIS, averaged between 2003 and 2014, for each latitudinal band. (e)–(h) LAI 50% threshold DOYs derived from CLM5.0 and MODIS, averaged between 2003 and 2014, for each latitudinal band. (i)–(l) LAI 50% threshold DOYs derived from CLM4.5 and CLM5.0, averaged between 2003 and 2014, for each latitudinal band.

dominated locations in temperate regions show high and significant correlation (Fig. 7c).

Other than the added precipitation trigger in stress deciduous phenology, natural plant phenology schemes are largely the same in CLM4.5 and CLM5.0, therefore, we also investigated the soil properties that trigger spring onset. Soil temperature only exhibits small differences in subtropical and temperate regions from the GLDAS (Rodell et al. 2004; described in the supplementary material) and when forced with different datasets (Fig. S6), except over the Greenland and in high-altitude regions like the Tibetan Plateau. At high latitudes (north of 60°N), CLM4.5 exhibits warmer temperature than the GLDAS and CLM5.0 (Figs. S6a,d). The difference in soil moisture is much larger (Fig. S7). CLM4.5

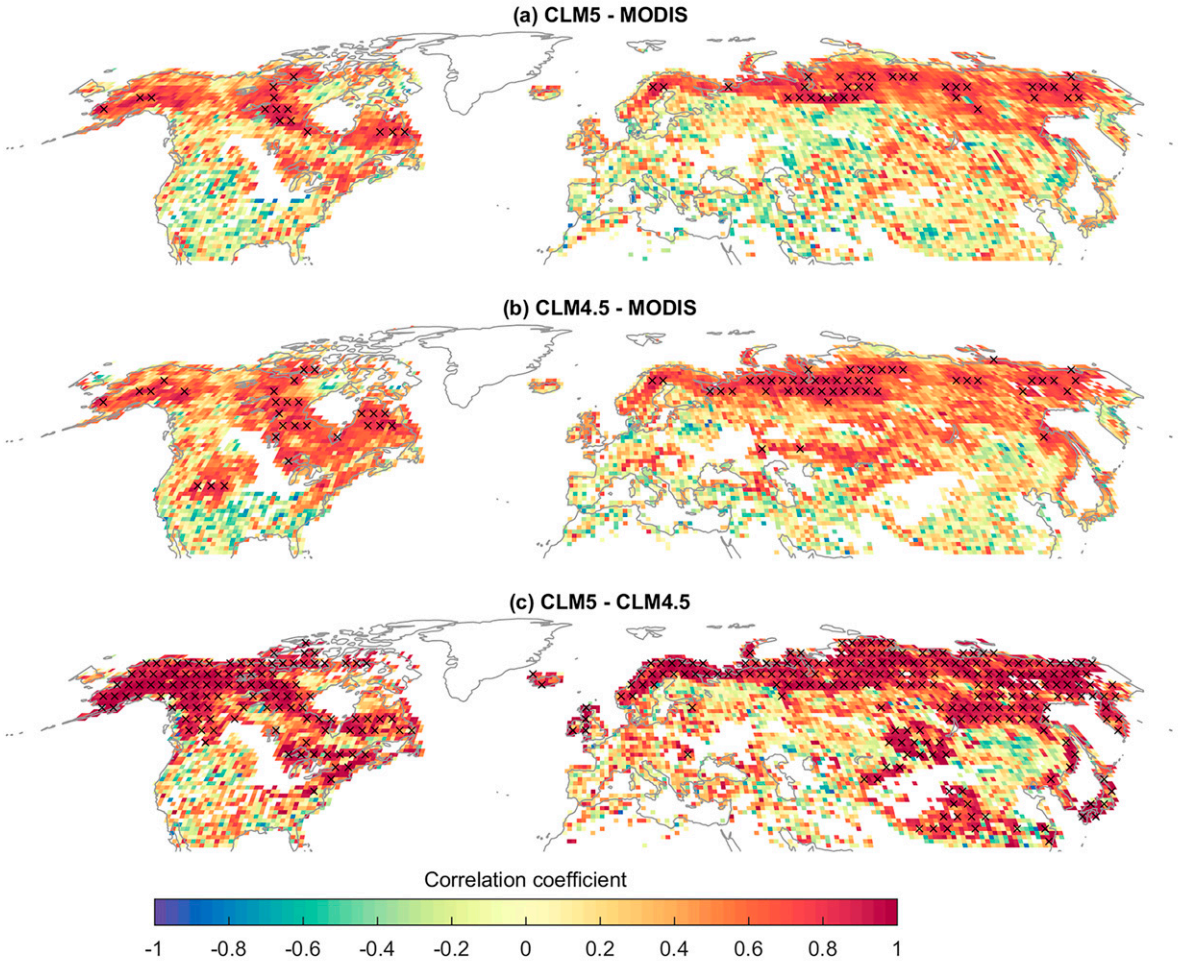

FIG. 7. Correlations between (a) CLM5.0 and MODIS, (b) CLM4.5 and MODIS, and (c) CLM5.0 and CLM4.5 LAI 50% threshold DOYs during 2003–14. Grid points exhibiting significant correlation after adjustment for false positive are marked with a black "x."

exhibits drier soil than both the GLDAS and CLM5.0, except in boreal regions in northern Russia and northern Canada (Figs. S7a,d). CLM5.0 mostly exhibits wetter soil than the GLDAS, except for a few regions in northern Canada, southeast and western China, and in northern Russia (Fig. S7b). On average, CLM5.0 soil moisture in the top 10cm of soil is more than 50% lager than CLM4.5 in low- to midlatitude regions (more than 30% drier in Fig. S7d). Although there are large differences between the amplitude of soil moisture, both soil temperature and soil moisture show significant correlation between model simulations and the GLDAS (Figs. S8 and S9).

b. *Terrestrial productivity*

To evaluate plant carbon uptake, we also investigated the seasonal window of "peak growth," as indicated by LAI values over 75% of the annual dynamical range (Fig. 8 and Figs. S10 and S11). Across the Northern Hemisphere, peak growing season length decreases northward (Fig. 8). In general, CLM has a longer peak growing season than MODIS (Figs. S10a,b and S11). For instance, peak growing season can be more than 80 days longer in CLM than in MODIS over the high-elevation regions in Asia and 50 days longer in boreal regions in North America and Eurasia (Figs. S10a,b). In CLM5.0, evergreen forest dominated boreal regions (i.e., southern coast of the Hudson Bay and northern and eastern Europe) exhibit longer peak growing season than in both CLM 4.5 and MODIS (Figs. S10a,c). On average, difference between CLM and MODIS peak growing season length is the largest over stress deciduous PFTs and is more than 50 days in CLM4.5 and more than 35 days in CLM5.0 (Fig. S11).

To characterize how discrepancies in peak growing season length between CLM and MODIS potentially influence modeled estimates of the terrestrial carbon budget, we calculated CLM NPP during that period (ΔNPP_{pheno}, Fig. 9) and how large the mismatch is compared against annual total NPP (Fig. 10). Over the Northern Hemisphere, NPP is consistently overestimated in CLM at high latitudes, in eastern and

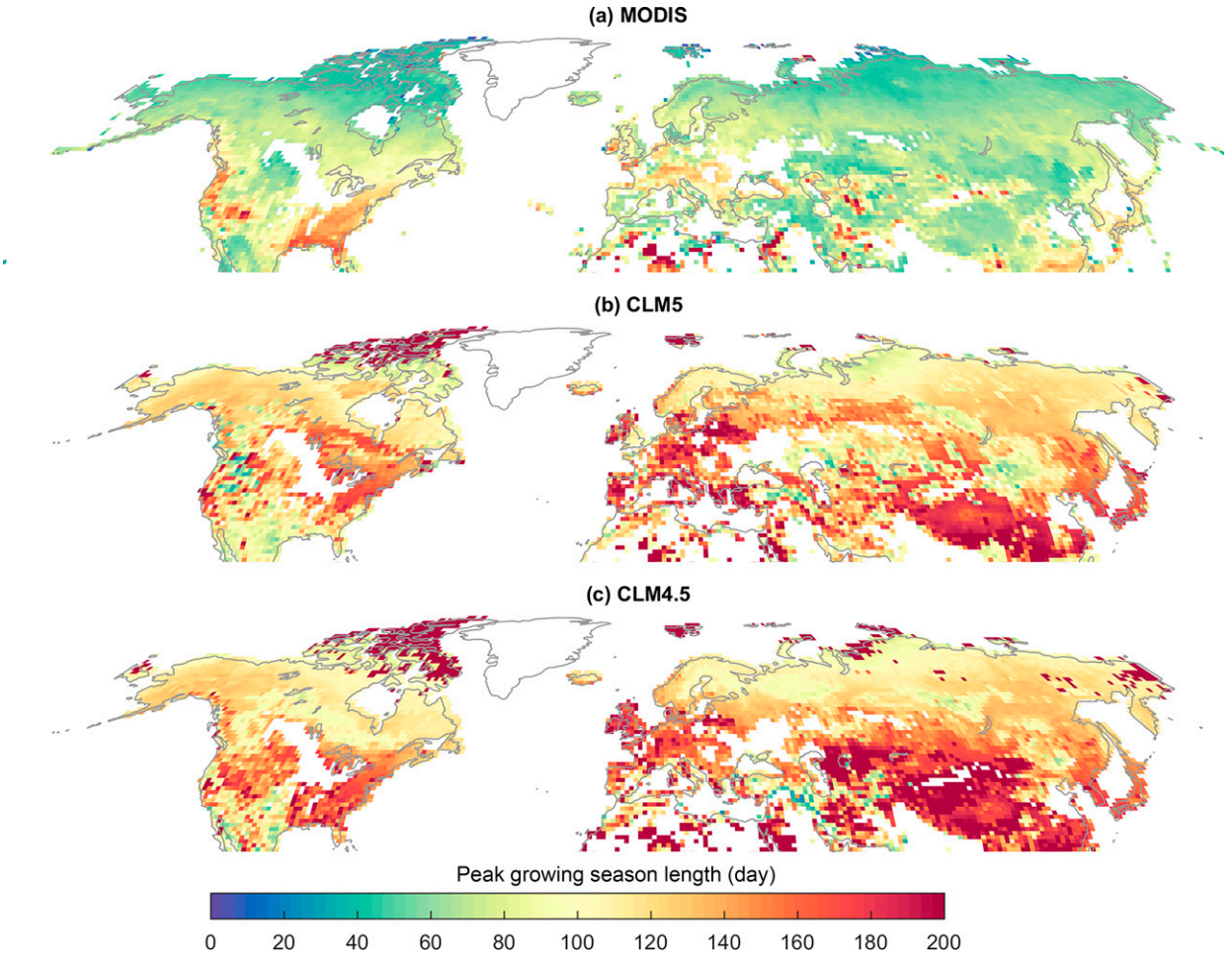

FIG. 8. Peak growing season length as indicated by LAI 75% threshold averaged between 2003 and 2014.

southwestern United States, and in eastern and southeastern Asia. For instance, in the southwestern United States, CLM5.0 estimates 0.05 ± 0.02 kgC m^{-2} NPP in the extended growing season, consisting of 32% ± 13% of total annual net productivity in CLM (Figs. 9 and 10). Likewise, in eastern Asia, CLM5.0 estimates 0.2 ± 0.05 kgC m^{-2} more NPP per year in the peak growing season, accounting for 40% ± 10% of total annual net productivity in that region in CLM. Compared to CLM5.0, CLM4.5 estimates an even larger NPP in the low to midlatitudes of eastern and southeastern Asia, Europe, and the United States, but simulates lower annual NPP in high-latitude regions. Overall, disagreements between CLM and MODIS peak growing season produce larger NPP in CLM over the Northern Hemisphere, except for a few evergreen forest and grassland-dominated locations (Fig. 10).

4. Discussion

Appreciable differences are present in both LAI seasonal cycle and spring onset timing between CLM simulated phenology and MODIS estimates. The timing of spring onset derived from satellite remote sensing and CLM exhibits similar topographic and latitudinal gradients at continental scales across the Northern Hemisphere (Fig. 5 and Fig. S4). However, CLM spring onset generally occurs later than MODIS across most temperate and boreal regions. In CLM, the start of spring can be delayed for more than 30 days in high-latitude regions where land cover is dominated by evergreen needleleaf trees and more than 60 days in arid midlatitude regions dominated by grasslands and shrublands (Fig. 5). The reasons for the delayed spring onset in CLM may be that CLM onset flag for seasonal deciduous was calibrated using 50% threshold in remote sensing records, i.e., the critical thresholds of growing degree days or soil moisture that determines spring onset was originally tuned to the timing when 50% threshold of NDVI is reached in remote sensing, and the 30-day fixed onset period after that (White et al. 1997; Lawrence et al. 2019), as well as the lack of seasonality in evergreen phenology routine in CLM. Although spring onset timing is later in CLM than in MODIS at high latitudes, interannual variability of LAI thresholds shows the highest correlation in high-latitude regions (Figs. 6 and 7). In addition, although CLM5.0 represents the seasonal amplitude of LAI better than CLM4.5 (Fig. 4 and Figs. S2 and S3), CLM5.0 also

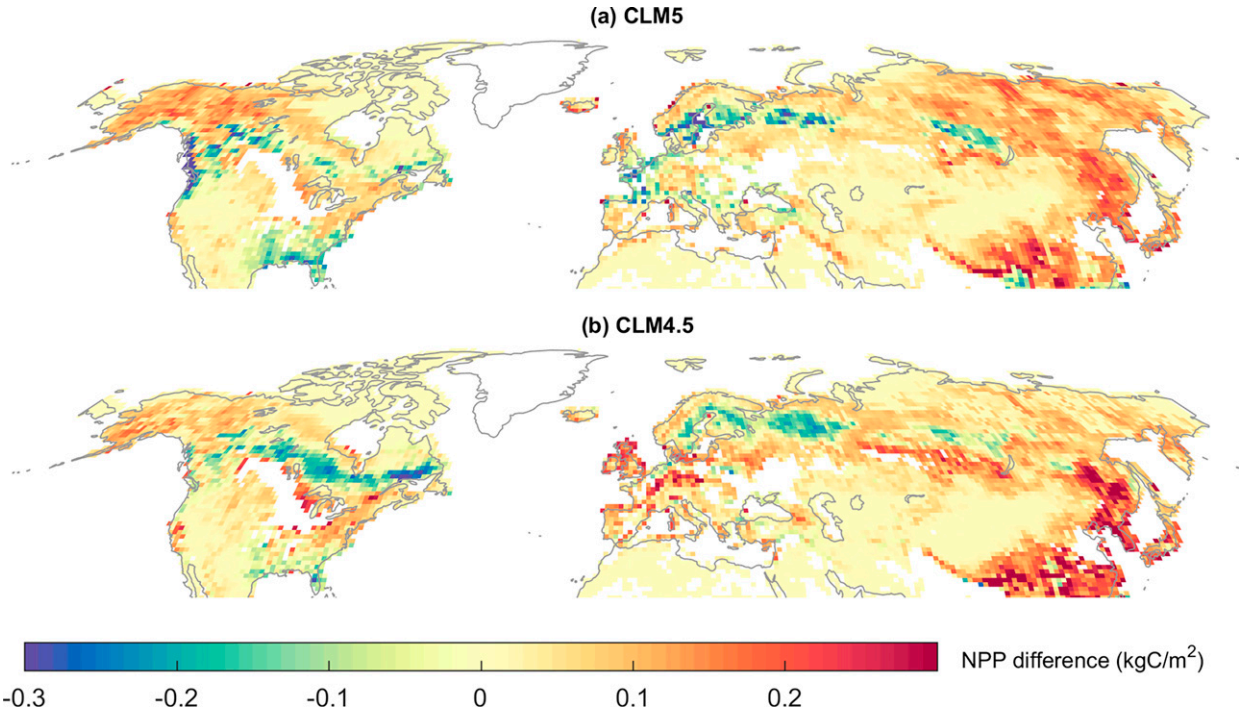

FIG. 9. Annual net primary production produced during the difference window between MODIS peak growing season as indicated by LAI 75% threshold and that of (a) CLM5.0 and (b) CLM4.5 (i.e., ΔNPP_{pheno}), averaged between 2003 and 2014.

displays a larger difference in mean DOY of LAI thresholds and weaker correlations when compared against MODIS LAI (Fig. 7 and Fig. S5).

The phenology routines of CLM affect the start and duration of terrestrial productivity across a wide range of PFTs. Mismatches between the start and end dates of the growing season therefore can translate into large potential discrepancies between the CLM carbon budget and MODIS carbon budget. Differences between peak growing season lengths average around 40 days each year, resulting in a large difference in terrestrial net primary productivity and hence the carbon cycle. Over the Northern Hemisphere, the total difference in modeled NPP due to an excessively long peak growing season length is 2.17 ± 0.79 Pg carbon per year in CLM5.0 and 2.85 ± 1.37 Pg carbon per year in CLM4.5, accounting for approximately 5% of total estimated Northern Hemisphere NPP (Li et al. 2017). We note that the bias in peak growing season length is about 33% in relative terms (on average 40 days versus a mean peak growing season length of 120 days); which is substantially larger than the 5% bias in NPP. One explanation for this is that "extra" days in autumn are less productive than those in summer, because the reduced amount of available sunlight, and dry soils, both serve to limit plant productivity.

Despite the large differences in the start of spring, phenology schemes were largely the same in CLM4.5 and CLM5.0, except for the added precipitation trigger in stress deciduous phenology. Accordingly, the changes in phenology across model versions as diagnosed here must primarily be due to secondary impacts of new or revised CLM processes that affect phenological triggers. For example, changes in processes that affect soil moisture will have a direct influence on modeled phenological transitions in stress deciduous ecosystems. As a result, while we have seen major improvements for a range of states and fluxes from CLM4.5 to CLM5.0 (Lawrence et al. 2019; Wieder et al. 2019), plant phenology shows poorer agreement with MODIS even though the phenology routines have not changed considerably. Because phenology simulation can change without direct modification of the phenology parameterizations and plant LAIs have important implications for productivity and the carbon cycle, it is important to evaluate not only LAI amplitudes but season cycles of leaf activity using indicators of seasonal transitions when assessing model outputs or comparing model performance.

Indeed, interactions among model components increase the overall challenge of improving CLM phenology, and we acknowledge that improving model performance for simulating the seasonality of LAI could have unforeseen ripple effects that negatively impact model performance for other key variables or metrics, some of which may be more important than phenology itself. However, we argue that improving modeled phenology should, ultimately, lead to overall better model performance and enhanced realism. Improving phenological routines in CLM could be achieved in a variety of ways. The basic elements of CLM's phenology routines are little-changed since the pioneering work of White et al. (1997). But, in the last two decades, the quality of satellite phenology data products has dramatically increased. The high

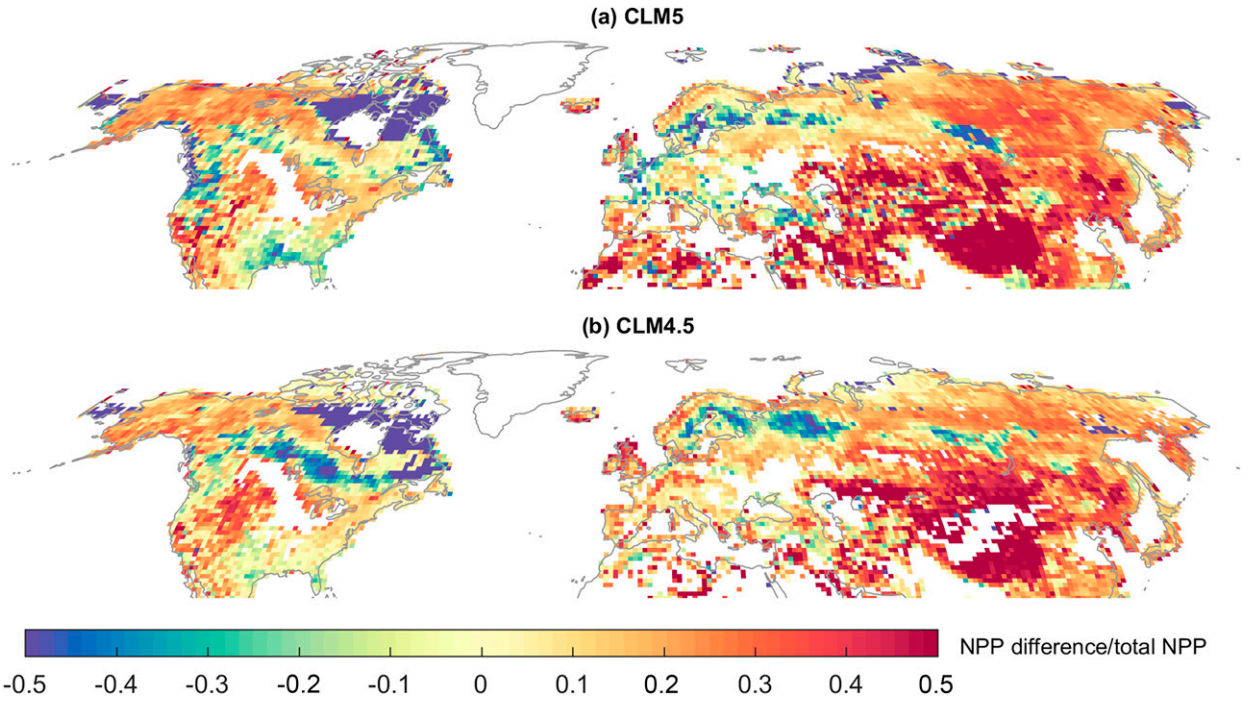

FIG. 10. Fraction of total annual net primary production (ΔNPP_{pheno}/total NPP) that is produced during the difference window between MODIS peak growing season as indicated by LAI 75% threshold and that of (a) CLM5.0 and (b) CLM4.5, averaged between 2003 and 2014.

quality of MODIS and VIIRS phenological metrics are well established (Moon et al. 2019; Richardson et al. 2018), and newer high-resolution data products from the Harmonized Landsat Sentinel-2 project should further reduce observational uncertainties. Additionally, near-surface remote sensing now provides consistent phenological data, at high spatial and temporal resolution, across a wide range of biomes (Richardson 2019). These data can be leveraged for phenological model development and independent validation (e.g., Melaas et al. 2016; Hufkens et al. 2016).

A variety of reasons may be causing these differences in CLM and MODIS plant phenology. One possible source of differences in the observed and simulated annual cycle of LAI is a mismatch between CLM PFTs and MODIS land cover types due to both land cover heterogeneity and the comparatively coarse spatial resolution of CLM. Another potential explanation is the different procedures of how leaf area index is estimated from MODIS products and simulated in CLM. For example, evergreen phenology in land surface models is usually designed to have a negligible annual cycle because a substantial portion of foliage is retained from one year to the next (Kim et al. 2015; Lawrence and Chase 2010). By comparison, MODIS LAI is derived from assumptions about land cover type and observed reflectance, which may be affected by snow on the ground particularly in evergreen systems. Therefore, differences between CLM and MODIS LAI in evergreen-dominant boreal regions likely stem, at least in part, from observational uncertainties in MODIS, which tends to overestimate the magnitude of the winter dormancy signal due to snow cover and hence overestimate the annual dynamical range (Myneni et al. 2015; Yan et al. 2016). At the same time, constant production by evergreens in CLM is also not physiologically accurate; the phenology routines governing evergreen LAI in CLM constrain the timing and magnitude of phenology as there is no onset and offset trigger and growth can happen all year around (as shown by the evergreen PFTs in Fig. S12). Uncertainties can also rise from the curve fitting approach we adopted to interpolate the 8-day composite to daily LAI in MODIS (Klosterman et al. 2014). In addition, other problems in the phenology subroutines in CLM may be causing the differences between LAI observations and simulations. For instance, temperate and boreal evergreen forests can maintain high LAI values and exhibit low intra-annual variation, while grasslands at low- to midlatitude regions experience multiple growing seasons within one calendar year (Fig. S12).

Simulating terrestrial ecosystems globally with one single land surface model can be challenging. Large heterogeneity is present within terrestrial ecosystems, and yet only limited observations are available to derive the parameters from and constrained computational power is assigned to the land surface components in Earth system models. CLM has full parameter sets for 14 different natural plant functional types and fitting the entire terrestrial natural ecosystems with 14 sets of parameters can be hard. LAI biases have improved from CLM4.5 to CLM5.0 and a few previously simplified biogeophysical and biogeochemical processes have been added into CLM5.0, such as fixation and uptake

of nitrogen and evaporative influences on plant hydrology. However, compared to CLM4.5, CLM5.0 shows larger disagreements with MODIS start of spring dates (Figs. 5 and 6), and leaf phenology modifies the simulated carbon cycle in the model.

Model improvement can be envisioned at a variety of levels. First, if the underlying theory is sound, improvement might only require better model parameters (characterizing, for example, temperature, photoperiod and moisture sensitivities or thresholds) that are more representative at the global scale. These parameter estimates can be conditioned on observational data to minimize the mismatch between observations and model simulations. Second, if the observational data indicate that the existing process representation is inadequate, additional processes or functional representations of processes can be incorporated in the model for specific phenological strategies or PFTs (Chen et al. 2016; Dahlin et al. 2015). Finally—and more ambitiously—the wealth of observational data now available should allow for careful evaluation of whether additional phenological strategies or PFTs need to be included, or whether phenological model parameters and/or structure need to vary within existing PFTs. Data assimilation (e.g., Stöckli et al. 2008; Ling et al. 2019) and machine learning (Czernecki et al. 2018) methods may find use in these efforts. In addition to mean and peak LAI, future work should also compare seasonal variability of plant phenology in LSMs with different phenology schemes or parameters to assess the phenology simulations and their potential influences on the carbon cycle. In this study, we present a suite of indicators that facilitate the comparison of phenology between models and observations and identify potential biases induced by nonphenological changes in the model. We believe that through rigorous model–data fusion (Williams et al. 2009) it should be possible to develop improved phenological routines for CLM that are robust in both time and space.

5. Conclusions

Spring onset indicators based on LAI thresholds are critical for identifying sources of mismatch/bias in timing that would be otherwise difficult to see in seasonal mean values at monthly scales. Our results reveal large discrepancies in the start of spring dates between MODIS and CLM as well as between CLM4.5 and CLM5.0. Across the Northern Hemisphere, CLM generally simulates later spring onset timing and longer growing season than presented in MODIS. Although CLM5.0 provides better representation of LAI absolute values and seasonal amplitudes than CLM4.5, the start of spring timing shows less agreement between CLM5.0 and MODIS. These disagreements in the start of spring and growing season length can result in large biases in terrestrial productivity simulations (approximately 5% of total estimated Northern Hemisphere NPP). In addition, although the phenology routines have not changed considerably from CLM4.5 to CLM5.0 and despite the broad improvements for a range of states and fluxes including LAI amplitudes, CLM5.0 phenology shows large differences from CLM4.5 and a poorer agreement with MODIS in the timing of the seasonal cycle.

Temporal correlations between CLM and MODIS LAI thresholds at each grid point are mostly not statistically significant except in high-latitude regions. These findings suggest that there are fundamental differences in the environmental sensitivity of phenology derived from MODIS and CLM, especially for evergreen and grassland plant functional types in the Northern Hemisphere. These differences need to be resolved in order to increase confidence in model-based simulations both of land–atmosphere coupling, and phenological responses to climate change and climate variability. Model improvement can be guided by the wealth of satellite and near-surface remote sensing data now available at high spatial resolution and temporal frequency, leveraging state-of-the-art parameter estimation and data assimilation techniques that have been developed and applied to land surface models over the last decade. Improving phenological routines in CLM should enable better representation of seasonal patterns of land–atmosphere carbon, water, and energy fluxes and hence land–atmosphere feedbacks more generally.

Acknowledgments. This work is supported by NSF Macrosystems Biology Award (DEB-1702551) and NSF Career Award (AGS-1751535). We thank the reviewers for their helpful comments that significantly improved the paper. We would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory. The authors declare no conflicts of interest with this publication.

Data availability statement. The MODIS LAI data are publicly available online through the USGS website: https://lpdaac.usgs.gov/products/mod15a2hv006/. LAI and NPP simulations from the CESM experiments are publicly available on figshare (doi: 10.6084/m9.figshare.16441608).

REFERENCES

Albergel, C., and Coauthors, 2017: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8. 0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geosci. Model Dev.*, **10**, 3889–3912, https://doi.org/10.5194/gmd-10-3889-2017.

——, E. Dutra, S. Munier, J. C. Calvet, J. Munoz-Sabater, P. D. Rosnay, and G. Balsamo, 2018: ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better? *Hydrol. Earth Syst. Sci.*, **22**, 3515–3532, https://doi.org/10.5194/hess-22-3515-2018.

Ault, T. R., M. D. Schwartz, R. Zurita-Milla, J. F. Weltzin, and J. L. Betancourt, 2015: Trends and natural variability of spring onset in the coterminous United States as evaluated by a new gridded dataset of spring indices. *J. Climate*, **28**, 8363–8378, https://doi.org/10.1175/JCLI-D-14-00736.1.

Barbu, A. L., J. C. Calvet, J. F. Mahfouf, C. Albergel, and S. Lafont, 2011: Assimilation of soil wetness index and leaf area index into the ISBA-A-GS land surface model:

Grassland case study. *Biogeosciences*, **8**, 1971–1986, https://doi.org/10.5194/bg-8-1971-2011.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57B**, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

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

Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https://doi.org/10.1002/qj.776.

Czernecki, B., J. Nowosad, and K. Jabłońska, 2018: Machine learning modeling of plant phenology based on coupling satellite and gridded meteorological dataset. *Int. J. Biometeor.*, **62**, 1297–1309, https://doi.org/10.1007/s00484-018-1534-2.

Dahlin, K. M., R. A. Fisher, and P. J. Lawrence, 2015: Environmental drivers of drought deciduous phenology in the Community Land Model. *Biogeosciences*, **12**, 5061–5074, https://doi.org/10.5194/bg-12-5061-2015.

——, D. Del Ponte, E. Setlock, and R. Nagelkirk, 2017: Global patterns of drought deciduous phenology in semi-arid and savanna-type ecosystems. *Ecography*, **40**, 314–323, https://doi.org/10.1111/ecog.02443.

Fox, A. M., and Coauthors, 2018: Evaluation of a data assimilation system for land surface models using CLM4. 5. *J. Adv. Model. Earth Syst.*, **10**, 2471–2494, https://doi.org/10.1029/2018MS001362.

Guillevic, P., R. D. Koster, M. J. Suarez, L. Bounoua, G. J. Collatz, S. O. Los, and S. P. P. Mahanama, 2002: Influence of the interannual variability of vegetation on the surface energy balance—A global sensitivity study. *J. Hydrometeor.*, **3**, 617–629, https://doi.org/10.1175/1525-7541(2002)003<0617:IOTIVO>2.0.CO;2.

Hufkens, K., and Coauthors, 2016: Productivity of North American grasslands is increased under future climate scenarios despite rising aridity. *Nat. Climate Change*, **6**, 710–714, https://doi.org/10.1038/nclimate2942.

Kim, Y., P. R. Moorcroft, I. Aleinov, M. J. Puma, and N. Y. Kiang, 2015: Variability of phenology and fluxes of water and carbon with observed and simulated soil moisture in the Ent Terrestrial Biosphere Model (Ent TBM version 1.0.1.0.0). *Geosci. Model Dev.*, **8**, 3837–3865, https://doi.org/10.5194/gmd-8-3837-2015.

Klosterman, S. T., and Coauthors, 2014: Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, **11**, 4305–4320, https://doi.org/10.5194/bg-11-4305-2014.

——, K. Hufkens, and A. D. Richardson, 2018: Later springs green-up faster: The relation between onset and completion of green-up in deciduous forests of North America. *Int. J. Biometeor.*, **62**, 1645–1655, https://doi.org/10.1007/s00484-018-1564-9.

Knyazikhin, Y., and Coauthors, 1999: MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD 15). Algorithm Theoretical Basis Doc., version 4.0, 126 pp., https://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf.

Lawrence, D. M., and Coauthors, 2019: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Model. Earth Syst.*, **11**, 4245–4287, https://doi.org/10.1029/2018MS001583.

Lawrence, P. J., and T. N. Chase, 2010: Investigating the climate impacts of global land cover change in the community climate system model. *Int. J. Climatol.*, **30**, 2066–2087, https://doi.org/10.1002/joc.2061.

Li, P., and Coauthors, 2017: Quantification of the response of global terrestrial net primary production to multifactor global change. *Ecol. Indic.*, **76**, 245–255, https://doi.org/10.1016/j.ecolind.2017.01.021.

Ling, X. L., C. B. Fu, Z. L. Yang, and W. D. Guo, 2019: Comparison of different sequential assimilation algorithms for satellite-derived leaf area index using the Data Assimilation Research Testbed (version Lanai). *Geosci. Model Dev.*, **12**, 3119–3133, https://doi.org/10.5194/gmd-12-3119-2019.

Mahowald, N., F. Lo, Y. Zheng, L. Harrison, C. Funk, D. Lombardozzi, and C. Goodale, 2016: Projections of leaf area index in Earth system models. *Earth Syst. Dyn.*, **7**, 211–229, https://doi.org/10.5194/esd-7-211-2016.

Melaas, E. K., M. A. Friedl, and A. D. Richardson, 2016: Multiscale modeling of spring phenology across deciduous forests in the eastern United States. *Global Change Biol.*, **22**, 792–805, https://doi.org/10.1111/gcb.13122.

Moon, M., X. Zhang, G. M. Henebry, L. Liu, J. M. Gray, E. K. Melaas, and M. A. Friedl, 2019: Long-term continuity in land surface phenology measurements: A comparative assessment of the MODIS land cover dynamics and VIIRS land surface phenology products. *Remote Sens. Environ.*, **226**, 74–92, https://doi.org/10.1016/j.rse.2019.03.034.

Morisette, J. T., and Coauthors, 2009: Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century. *Front. Ecol. Environ.*, **7**, 253–260, https://doi.org/10.1890/070217.

Müller Schmied, H., and Coauthors, 2016: Impact of climate forcing uncertainty and human water use on global and continental water balance components. *Proc. IAHS*, **374**, 53–62, https://doi.org/10.5194/piahs-374-53-2016.

Myneni, R., Y. Knyazikhin, and T. Park, 2015: MOD15A2H MODIS/Terra leaf area index/FPAR 8-day L4 global 500 m SIN grid V006. NASA EOSDIS Land Processes DAAC, accessed 2 July 2018, https://doi.org/10.5067/MODIS/MOD15A2H.006.

Neale, R. B., and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+ STR, 268 pp., www.cesm.ucar.edu/models/cesm1.1/cam/docs/description/cam5_desc.pdf.

Oleson, K. W., and Coauthors, 2013: Technical description of version 4.5 of the Community Land Model (CLM). NCAR Tech. Note NCAR/TN-503+STR, 420 pp., https://doi.org/10.5065/D6RR1W7M.

Peano, D., and Coauthors, 2021: Plant phenology evaluation of CRESCENDO land surface models–Part 1: Start and end of the growing season. *Biogeosciences*, **18**, 2405–2428, https://doi.org/10.5194/bg-18-2405-2021.

Peñuelas, J., and I. Filella, 2009: Phenology feedbacks on climate change. *Science*, **324**, 887–888, https://doi.org/10.1126/science.1173004.

Richardson, A. D., 2019: Tracking seasonal rhythms of plants in diverse ecosystems with digital camera imagery. *New Phytol.*, **222**, 1742–1750, https://doi.org/10.1111/nph.15591.

——, D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, and J. O'Keefe, 2009: Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England

forests. *Tree Physiol.*, **29**, 321–331, https://doi.org/10.1093/treephys/tpn040.

——, and Coauthors, 2010: Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philos. Trans. Roy. Soc.*, **B365**, 3227–3246, https://doi.org/10.1098/rstb.2010.0102.

——, and Coauthors, 2012: Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program Site Synthesis. *Global Change Biol.*, **18**, 566–584, https://doi.org/10.1111/j.1365-2486.2011.02562.x.

——, T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey, 2013: Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteor.*, **169**, 156–173, https://doi.org/10.1016/j.agrformet.2012.09.012.

——, K. Hufkens, T. Milliman, and S. Frolking, 2018: Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1. 0 and MODIS satellite remote sensing. *Sci. Rep.*, **8**, 5679, https://doi.org/10.1038/s41598-018-23804-6.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, https://doi.org/10.1175/BAMS-85-3-381.

Sabater, J. M., C. Rüdiger, J. C. Calvet, N. Fritz, L. Jarlan, and Y. Kerr, 2008: Joint assimilation of surface soil moisture and LAI observations into a land surface model. *Agric. For. Meteor.*, **148**, 1362–1373, https://doi.org/10.1016/j.agrformet.2008.04.003.

Scholze, M., M. Buchwitz, W. Dorigo, L. Guanter, and S. Quegan, 2017: Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences*, **14**, 3401–3429, https://doi.org/10.5194/bg-14-3401-2017.

Schwartz, M. D., 1992: Phenology and springtime surface-layer change. *Mon. Wea. Rev.*, **120**, 2570–2578, https://doi.org/10.1175/1520-0493(1992)120<2570:PASSLC>2.0.CO;2.

——, R. Ahas, and A. Aasa, 2006: Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biol.*, **12**, 343–351, https://doi.org/10.1111/j.1365-2486.2005.01097.x.

——, T. R. Ault, and J. L. Betancourt, 2013: Spring onset variations and trends in the continental United States: Past and regional assessment using temperature-based indices. *Int. J. Climatol.*, **33**, 2917–2922, https://doi.org/10.1002/joc.3625.

Stöckli, R., and Coauthors, 2008: Use of FLUXNET in the Community Land Model development. *J. Geophys. Res. Biogeosci.*, **113**, G01025, https://doi.org/10.1029/2007JG000562.

Thornton, P. E., and N. E. Zimmermann, 2007: An improved canopy integration scheme for a land surface model with prognostic canopy structure. *J. Climate*, **20**, 3902–3923, https://doi.org/10.1175/JCLI4222.1.

Wang, X., and Coauthors, 2014: A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, **506**, 212–215, https://doi.org/10.1038/nature12915.

White, M. A., P. E. Thornton, and S. W. Running, 1997: A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochem. Cycles*, **11**, 217–234, https://doi.org/10.1029/97GB00330.

——, and Coauthors, 2009: Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biol.*, **15**, 2335–2359, https://doi.org/10.1111/j.1365-2486.2009.01910.x.

Wieder, W. R., and Coauthors, 2019: Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochem. Cycles*, **33**, 1289–1309, https://doi.org/10.1029/2018GB006141.

Williams, M., and Coauthors, 2009: Improving land surface models with FLUXNET data. *Biogeosciences*, **6**, 1341–1359, https://doi.org/10.5194/bg-6-1341-2009.

Xu, X., W. J. Riley, C. D. Koven, G. Jia, and X. Zhang, 2020: Earlier leaf-out warms air in the north. *Nat. Climate Change*, **10**, 370–375, https://doi.org/10.1038/s41558-020-0713-4.

Yan, K., and Coauthors, 2016: Evaluation of MODIS LAI/FPAR product collection 6. Part 1: Consistency and improvements. *Remote Sens.*, **8**, 359, https://doi.org/10.3390/rs8050359.