Intercomparison and evaluation of spring phenology products using National Phenology Network and AmeriFlux observations in the contiguous United States

**Dailiang Peng^a, Xiaoyang Zhang^b, Chaoyang Wu^{c*}, Wenjiang Huang^{a*},
Alemu Gonsamo^d, Alfredo R. Huete^e, Kamel Didan^f, Bin Tan^g, Xinjie Liu^a,
Bing Zhang^a**

^aKey Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

^bGeospatial Sciences Center of Excellence, South Dakota State University, Brookings, South Dakota, 57007, USA

^cState Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, 100101, China

^dDepartment of Geography and Planning, University of Toronto, 100 St. George St., Toronto, ON, M5S 3G3, Canada

^eClimate Change Cluster (C3), University of Technology Sydney, NSW 2007, Australia

^fDepartment of Electrical and Computer Engineering, The University of Arizona, 1230 E. Speedway, Tucson, AZ 85721, USA

^gEarth Resources Technology Inc. at NASA Goddard Space Flight Center, Code 614.5, Greenbelt, MD 20771, USA

* Corresponding author:

E-mail address: hefery@163.com (C.Wu); huangwj@radi.ac.cn (W. Huang)

Abstract

Many remote sensing based spring phenology products have been developed to monitor and study vegetation phenology at regional and global scales. It is important to understand how these products perform relative to each other and to ground observations. In this study, we extracted spring green-up onset dates (GUD) over the contiguous United States (CONUS) from six major land surface phenology (LSP) products: (1) Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Dynamics Phenology (MCD12Q2); (2) Vegetation Index and Phenology Multi-sensor Phenology (VIPPHENEVI2); (3) Global Long-Term Climate Modeling Grid Land Surface Phenology (CMGLSP); (4 and 5) North American Carbon Program (NACP) Phenology (MOD09Q1PEVI and MOD15PHN); and (6) USGS/EROS advanced very high resolution radiometer (AVHRR) phenology (AVHRRP). We characterized and compared the GUD data in these LSP products, and evaluated their accuracy using ground-based phenology observations [i.e., human observations of first leaf and sensor readings of gross primary productivity (GPP)] from the USA National Phenology Network (USA-NPN) and AmeriFlux. The results revealed the

consistencies and discrepancies of GUD estimates among LSP products.

Intercomparison of the six products indicated that the root mean square error (RMSE) of these products range from 17.8 days to 31.5 days, whereas AVHRRP GUD has the lowest correlation and largest RMSE (~30 days) relative to other products. When compared to ground observations, GUD estimates in six LSP products generally have RMSE values of ~20 days and significant correlations ($p < 0.001$). For the products (MCD12Q2, AVHRRP, MOD09Q1PEVI, and MOD15PHN) available for comparisons in the short-term period (from 2001–2007), AVHRRP GUD presented **relatively weaker correlations and a lower** index of agreement (IOA), however, MCD12Q2 GUD showed overall slightly better consistencies with ground observations. In the two long-term products (CMGLSP and VIPPHENEVI2 from 1982–2013), CMGLSP exhibited stronger correlations, lower RMSE, and higher IOA with ground observations of the first leaf dates than VIPPHENEVI2 did. To our knowledge, our study provides the first comprehensive evaluation of phenology products using two independent ground-based datasets.

Keywords: remote sensing; ; ; , first leaf dates, land surface phenology, green-up onset date, evaluation.

1. Introduction

Plant phenology is widely used as an independent measure and powerful indicator of how ecosystems are responding to climate change (Badeck et al., 2004;

Friedl et al., 2006; Parmesan, 2006; White et al., 2009; Ma et al., 2013). It is linked to primary productivity, crop yields, insect emergence, and bird migration (Parmesan, 2006; Rosenzweig et al., 2007; Ault et al., 2015). Shifts in the timing of spring arrival can change the time of seeding in agricultural systems (Knudson, 2012), disrupt ecological interactions between plants and animals (Burkle et al., 2013; Kellermann et al., 2015), and contribute to changes in species distributions (Chuine & Beaubien, 2001). Earlier spring onset, in particular, can alter the surface energy balance, accelerate transpiration (Wilson & Baldocchi, 2007; Piao et al., 2015), enhance mid-summer droughts, and increase the incidence of severe wildfires (White & Nemani, 2003; Angert et al., 2005; Westerling et al., 2006).

Land surface phenology (LSP), the study of the timing of recurring seasonal pattern of variation in vegetated land surfaces observed from satellite sensors, has been widely used in detecting and understanding climate changes and ecosystem dynamics (Gonsamo et al., 2012; Friedl et al., 2006; Parmesan, 2006). Various studies have tended to find that spring phenology has occurred earlier in recent decades, especially across the Northern Hemisphere, as a result of rising global temperatures (McCabe et al., 2012; Allstadt et al., 2015). However, not all documented trends in spring phenology have been consistent, whether detected by satellite-based sensors, ground-based sensors, or human observations (Menzel et al., 2006; Reed, 2006; Piao et al., 2007; White et al., 2009). Because the differences in phenological trends influence the forecasts of climate change, crop development, and ecological dynamics, it is critical to understand the nature of the inconsistencies.

Over the last decades, a number of LSP products have been developed from satellite data, which have different spatial and temporal characteristics. The products include (Table 1): (1) the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Dynamics Product (MCD12Q2) (Zhang et al., 2006; Ganguly et al., 2010); (2) Global Vegetation Index and Phenology Multi-sensor Phenology (VIPPHENEVI2) (White et al., 2009; Didan et al., 2010; Didan et al., 2016), which uses data from the advanced very high resolution radiometer (AVHRR) and MODIS data based on daily two-band enhanced vegetation index (EVI2); (3) Global Long-Term Climate Modeling Grid Land Surface Phenology (CMGLSP) (Zhang et al., 2014; Zhang, 2015); (4 and 5) the MODIS-based products (MOD09Q1PEVI and MOD15PHN) in support of the North American Carbon Program (NACP), which were produced over the contiguous United States (CONUS) using MOD09 reflectance and MOD15 leaf area index (LAI), respectively (Morisette et al., 2009; Tan et al., 2008; Tan et al., 2011); and (6) the United States Geological Survey/ Earth Resources Observation Systems (USGS/ EROS) Phenology Product (AVHRRP) over the CONUS (Reed et al., 1994). In addition, agencies and researchers have developed many other specialized and regional products (Verstraete et al., 2008; Hargrove et al., 2009; Dash et al., 2010; Lacaze et al., 2011; Gonsamo & Chen, 2016).

Current LSP products could contain considerable discrepancies and uncertainties because there are no standardized methods and criteria to quantify the LSP metrics (de Beurs & Henebry, 2010; White et al., 2009). After comparing 10 methods for estimating green-up onset dates (GUD) from the same satellite data source, White et

al. (2009) found that estimates differed by more than one month depending on the method used. Unfortunately, we know of no studies that have compared and validated a range of LSP products, particularly which use different satellite data sources, vegetation indices, and analytical methods to estimate spring phenology. Thus, direct validation and intercomparison of existing LSP products are essential for further applications that rely on accurate estimates of phenological metrics. These applications include climate models, agricultural models, and estimates of shifts in species performance and distribution.

This study compares GUD in the six major LSP products (Table 1) that are widely used and regularly updated by their developers. The main goals are to answer several crucial questions as follows: (1) what are the consistencies and discrepancies of GUD estimates across the products? (2) which product provides the best consistency when compared to ground-based observations? (3) what is the temporal and spatial retrieval efficiency in each product?

1 Table 1. Characteristics of MCD12Q2, VIPPHENEVI2, CMGLSP, MOD09Q1PEVI, MOD15PHN and AVHRRP products

Product Name	Satellite/ Vegetation Index Source	Production Period	Spatial resolution	Main methods	References
MODIS collection 5 phenology product (MCD12Q2)	Combined TERRA and AQUA platforms, MODIS 8-day NBAR-EVI	2001–2013	500 m	Moving- window linear interpolation, Savitzky-Golay filter, and piecewise logistic model	Zhang et al., 2003; Zhang et al., 2006,
Vegetation Index and Phenology Multi-sensor Phenology (VIPPHENEVI2)	Daily AVHRR (1981–1999) and MODIS C5 EVI2 (2000–2014)	1981–2014	0.05 deg.	Two-step filtering approach of the input data and the use of a new per pixel continuity algorithm at a monthly step; 3-year moving window average, and modified half- max method.	White et al., 1997, 2009; Didan et al., 2010; Didan et al., 2016
Global Long- Term Climate Modeling Grid	Daily AVHRR (1982–1999) and MODIS C5	1982–2013	0.05 deg.	Background EVI2 determination,	Zhang, 2003; Zhang et

Land Surface Phenology (CMGLSP)	EVI2 (2000–2013)			Savitzky-Golay filter, and hybrid piecewise logistic model.	al., 2014; Zhang, 2015
North American Carbon Project Phenology Product (MOD09Q1PEVI and MOD15PHN)	MOD09EVI, MCD15A2LAI	2001–2013	250 m and ~ 500 m for EVI, and 1 km for LAI	Spatially gap-filled and temporally smoothed method, and enhanced TIMESAT algorithm based on the asymmetric Gaussian smoothing function.	Tan et al., 2011
USGS/EROS AVHRR phenology product (AVHRRP)	AVHRR (NDVI)	1989–2010	1 km	Weighed, least-squares linear regression approach, and curve derivative method.	Reed et al., 1994.

2. Methods

Our study focused on the intercomparison and validation of GUD in LSP products across the CONUS. It was due to the fact that the well-organized ground reference measures were available, which were spatially distributed across the entire CONUS and included a number of ecosystems and land cover types. These data allowed us to compare the performance of land surface phenology products across a range of conditions.

2.1 Intercomparison of green-up onset dates among LSP products

We conducted intercomparisons in two different periods based on the availability of LSP products. The first was the comparison of short-term products that covered the period from 2001–2007 because most MODIS products were produced from 2001; MOD15PHN was only available from 2001–2007; and MOD09PHN from 2008–2013 was produced using a different approach (Tan et al., 2008, 2011). Over this period, we extracted GUD data from LSP products of MCD12Q2, VIPPHENEVI2, CMGLSP, MOD09Q1PEVI, MOD15PHN, and AVHRRP (Table 1). We then conducted following three intercomparisons. (1) Spatial difference of average GUD. These products were temporally averaged based on the valid GUD values at a pixel scale from 2001–2007 for the comparison of spatial performance. The GUD values in each LSP product were also averaged separately based on the International Geosphere–Biosphere Program (IGBP) classifications derived from the MODIS land cover product (Friedl et al., 2010) that was recorded in the MOD09Q1PEVI (Fig. 1). (2) Linear correlations of GUD among LSP products. For this comparison, the

temporally-averaged GUD values (from 2001–2007) in MCD12Q2, MOD09Q1PEVI, MOD15PHN, and AVHRRP were first spatially aggregated to a spatial resolution of 0.05 degrees using an average approach to match the pixel size in the VIPPHENEVI2 and CMGLSP products. The linear regression of pixel-based GUD was then calculated between any two LSP products. (3) Trend analysis. Interannual trends in GUD were examined separately for all six products in the period of 2001–2007. The trend only calculated for the pixels where GUD was determined during all seven years, otherwise it was indicated as missing observations. Note that the purpose of this trend analysis of the short-term observations was to examine the differences and similarities among the LSP products instead to investigate climate associated variation, which requires long-term observations. Therefore, no significant level was provided.

The second was long-term LSP comparison for the period from 1982–2013, which included GUD derived from AVHRR data (1982–1999) and MODIS data (2000–2013) in the VIPPHENEVI2 and CMGLSP products. Similar to short-term comparison, the GUD temporal trends and spatial patterns were analyzed and compared.

2.2 Ground observations of spring phenology

The USA National Phenology Network (USA-NPN) is a consortium of individuals and organizations that collect, share, and use phenology data. The organization aims to develop and distribute derived phenological information (Denny et al., 2014). The USA-NPN has developed standardized methods for professionals

and volunteers to observe and report the phenology of plants and animals, and has established a suite of quality assurance and quality control methods (Schwartz et al., 2012; Denny et al., 2014). Phenological observations from USA-NPN have been widely used for calibrating and evaluating LSP products, and for developing models to predict the timing of spring leaf emergence (Jeong et al., 2013; Chapman et al., 2014; Melaas et al., 2015; Gallinat et al., 2015; Gerst et al., 2015; Mazer et al., 2015).

The USA-NPN provides observations of first bloom and first leaf of lilac (including honeysuckle) from 1960s across the CONUS (Rosemartin et al., 2015; USA National Phenology Network, 2015; Schwartz et al., 2006). This study collected first leaf dates of lilac for the validation of GUD in six LSP products. It is due to the fact that (1) the first leaf is better than other spring phenological events (such as leaf budburst and first flowers) for the assessment of remotely sensed GUD (White et al., 2009); (2) the first leaf dates from the USA-NPN provide good estimates of phenological status at the species level across the CONUS with high spatial density of observations (Kellerman et al., 2015; Piao et al., 2015); (3) lilac observations have a long time period to temporally match the selected LSP products. Therefore, we acquired 7390 observations of the first leaf dates of lilacs from 1268 sites in the period from 1982–2013. These observations were spatially matched GUD pixels where the size varied with LSP products. If there were more than one lilac observations within a pixel in a given year, multiple observations were averaged. Only the lilac first leaf dates were chosen if the corresponding GUD values were valid for all the LSP products. Further, the samples were removed if the difference between

the LSP GUD and the lilac first leaf dates was larger than two months, which were likely associated with different drivers between the lilac dates and LSP GUD estimates. This was due to the fact that (1) lilac first leaf dates are sensitive to spring temperature (Schwartz et al., 2006), but large amount of lilac samples in the western CONUS (arid and semiarid regions) do not necessarily reflect the vegetation green-up controlled by water availability; and (2) lilac first leaf dates unlikely represent well the LSP GUD in very heterogeneous ecosystems and croplands. As a result, the selected observations were 249 from 2001–2007 in 92 sites and 4256 observations from 1982–2013 in 1099 sites for the validation of LSP GUD.

Further, we derived an additional 81 ground-based phenological observations during 2001–2007 from measurements of daily gross primary production (GPP) at 23 AmeriFlux sites (Fig. 1). Daily GPP ($\text{g C m}^{-2} \text{d}^{-1}$) was estimated from net ecosystem production and ecosystem respiration (Baldocchi et al., 2001). We used a negative exponential model, which was based on the polynomial regression and weights computed from the Gaussian density function, to derive smoothed curves for daily GPP observations in each site (Wu et al., 2013). Then, we calculated GUD as the dates at which GPP reached 10% of its seasonal maximal value across the smoothed curves of daily GPP observations (Wu et al., 2013).

2.3 Statistical analysis for evaluating green-up onset dates among LSP products using ground observations

We validated LSP GUD against ground observations for two periods: 2001–2007 for four short-term LSP products (MCD12Q2, MOD09PHN, MOD15PHN, and

AVHRRP), and 1982–2013 for two long-term LSP products (CMGLSP and VIPPHENEVI2). Statistical analysis included the number of observations, the correlation R and its associated p -value, the root mean square error (RMSE, Eq. 1) with its systematic (RMSE_s, Eq. 2) and unsystematic (RMSE_u, Eq. 3) components (Willmott, 1981), and the index of agreement (IOA) (Eq. 4). RMSE, RMSE_s, and RMSE_u provide gross error, the distance between this linear fit and the 1:1 line, and the dispersion around the best linear fit, respectively (Willmott, 1981; Delbart et al., 2015). RMSE_s calculates the systematic error, which is caused by the methods used to detect GUD. The methods mainly include the selecting vegetation indices, smoothing the time series, and detecting phenological metrics. Currently, there is no widely accepted or agreed definition of GUD from satellite data (White et al., 1997, 2009; Zhang et al., 2003; Didan et al., 2016). For example, the GUD estimates from the mid-point approach in VIPPHENEVI2 correspond to much higher vegetation index values in the area without snow cover relative to those from other approaches (Zhang et al., 2003; Tan et al., 2011; Reed et al., 1994). RMSE_u is a measure of the proportion of the discrepancy between satellite estimates and ground observations, which is due to random processes or influences other than the methods used to estimate GUD.

$$RMSE = \left[\sum_{i=1}^n \frac{(P_i - Q_i)^2}{n} \right]^{0.5} \quad (1)$$

where P_i and Q_i represent LSP GUD and ground observations, respectively, and n represents the number of ground observations.

$$RMSE_s = [\sum_{i=1}^n \frac{(\hat{P}_i - Q_i)^2}{n}]^{0.5} \quad (2)$$

$$RMSE_u = [\sum_{i=1}^n \frac{(P_i - \hat{P}_i)^2}{n}]^{0.5} \quad (3)$$

where $\hat{P}_i = a + bQ_i$, a and b are separately the intercept and slope of a least-square regression between the LSP GUD and ground observations.

IOA condenses the differences between LSP GUD estimates and ground observations into one statistical quantity. It is the ratio of the total RMSE to the sum of two differences – between each LSP GUD and the observed mean, and each observation and the observed mean. IOA has a theoretical range of 0 to 1, where the latter score suggests perfect agreement (Willmott, 1981; Zhang, 2015).

$$IOA = 1 - \frac{\sum_{i=1}^n (P_i - Q_i)^2}{\sum_{i=1}^n (|P_i'| + |Q_i'|)^2} \quad (4)$$

where $P_i' = P_i - \bar{Q}$, $Q_i' = Q_i - \bar{Q}$, P_i and Q_i represent separately the LSP GUD estimates and ground observations, and $\bar{Q}$ is the mean of the ground observations.

3. Results

3.1 Spatial consistency of green-up onset dates among LSP products

Fig. 2 shows the geographical distributions of GUD (day-of-year, DOY) averaged for 2001–2007 from each LSP product. All products showed relatively early GUD in southern locations and later GUD in northern locations, as expected. In the eastern CONUS, all products showed comparable GUD (~100 DOY) (Fig. 2).

The LSP products differed substantially in their estimates of GUD. MCD12Q2 GUD was generally earlier (Fig. 2a) while CMGLSP GUD was later (Fig. 2d). The products also provided spatially distinct patterns of GUD in specific regions (Fig. 2). For example, in the Upper-Midwest region, CMGLSP, MOD15PHN, and VIPPHENEVI2 estimated a later GUD (~130 DOY) relative to MCD12Q2 and AVHRRP (~100 DOY) (Fig. 2). In New Mexico (southwestern USA), GUD estimates varied substantially among LSP products, which were 50–90 DOY in MCD12Q2, ~100 DOY in MOD09Q1PEVI, 90–160 DOY in CMGLSP, MOD15PHN, and VIPPHENEVI2, and 120–170 DOY in AVHRRP. The AVHRRP GUD along coastal areas and in New Mexico and Florida was noticeably later than that in the northern regions (Fig. 2f).

The GUD estimates among LSP products differed across land cover types (Table 2). The most consistent GUD was found in deciduous broadleaf forests with a standard deviation (Sd) of 5.7 days. It was followed by mixed forests, croplands, and cropland/natural vegetation with a standard deviation less than 10 days. In contrast, the standard deviation was larger than 15 days in evergreen forests and shrublands, and as large as 32 days in barren or sparsely vegetated lands.

Table 2. Average green-up onset dates reported as day of year (DOY) for the period (2001–2007) across different land cover types. The acronym for land surface phenology products and land cover types is the same as defined in Table 1 and Fig. 1, respectively. Sd represents standard deviation (days).

MCD12Q2	MOD09Q1PEVI	MOD15PHN	CMGLSP	VIPPHENEVI2	AVHRRP	Sd
---------	-------------	----------	--------	-------------	--------	----

ENF	80.5	100.1	110.7	124.0	94.6	100.8	14.7
EBF	62.6	70.2	58.0	89.3	46.4	106.3	22.0
DNF	78.5	95.2	102.6	111.5	100.8	101.4	11.1
DBF	92.5	91.6	95.6	106.8	100.4	99.7	5.7
MF	89.8	91.8	93.5	110.2	100.6	101.0	7.6
SHR	56.2	99.9	105.3	96.4	98.3	85.9	17.9
SAV	59.9	84.8	89.3	78.3	91.6	82.1	11.4
GRA	68.4	86.2	97.2	98.6	89.0	83.3	11.0
WET	50.9	92.8	84.1	86.6	94.3	84.1	15.9
CRO	87.9	98.8	104.4	110.1	108.0	98.4	8.1
URB	60.2	77.4	79.4	87.7	86.2	87.1	10.4
CNV	74.0	78.8	85.9	96.4	90.2	95.7	9.1
BSV	42.0	110.2	97.4	97.3	34.6	65.6	31.8

Except for AVHRRP, the GUD estimates from other LSP products were strongly correlated, with R -values ranging from 0.50 to 0.72, and RMSE values from 17.8 to 31.5 days (Fig. 3). Similar to the spatial pattern shown in Fig. 2, the correlation indicated that MCD12Q2 GUD was generally earlier than estimates from other LSP products (Figs. 3a–e). The correlation between MOD09Q1PEVI and MOD15PHN was highest ($R = 0.72$) with a RMSE of 19.7 days largely because they used the same retrieval algorithm (Fig. 3f). AVHRRP GUD was most different from other LSP products (Figs. 3e, i, l, n, o). It had the lowest correlation ($R < 0.27$), largest intercept (> 68 days) and RMSE (~ 30 days) relative to other LSP products.

3.2 Completeness of LSP detections

All of the LSP products produced missing values for some parts over the CONUS, mostly in the Southwest regions, for at least one year in the time series (2001–2007 and 1982–2013) (Fig. 4). In the short-term period of 2001–2007, the CONUS land area with missing GUD detections that occurred at least once was largest (54%) in MCD12Q2 (Fig. 4a), while it was smallest (6.0%) in VIPPHENEVI2

(Fig. 4e). The area was very similar among MOD09Q1PEVI (22.8%), CMGLSP (23.5%), and AVHRRP (21.9%) (Figs. 4b, d, f). MOD15PHN detections was missing 33% (Fig. 4c) because of a missing tile image in 2007. For the long-term period of 1982–2013 (Fig. 5), missing CMGLSP GUD appeared at least one year in almost all of the southern CONUS (34% of total land area) (Fig. 5a), while VIPPHENEVI2 GUD was missing for only 8% of the land area (Fig. 5b).

3.3 Comparison of trends in green-up onset dates from LSP products

3.3.1 Short-term trends (2001–2007)

After excluding areas with missing data, analysis indicated that spring GUD delayed in 57.5% and 55.1% of the area at an average rate of 0.83 and 0.74 days per year in MCD12Q2 and AVHRRP, respectively. In contrast, spring GUD experienced advances in 55% of the area for VIPPHENEVI2, and about 64% for other three products (MOD09Q1PEVI, CMGLSP, and MOD15PHN) (Fig. 4). The average advance trends in days per year were 0.69 in MOD09Q1PEVI, and 0.84 in CMGLSP, and 0.2 in both MOD15PHN and VIPPHENEVI2. Rates of change in some pixels were larger than 5 days per year (Fig. 4). Five products (all except MCD12Q2) showed that spring was occurring earlier over time in central and northern regions of the CONUS. In contrast, analysis of MCD12Q2, CMGLSP, VIPPHENEVI2, and AVHRRP indicated that GUD was occurring later over time in the eastern CONUS (~1 days per year) (Fig. 4).

3.3.2 Long-term trends (1982–2013)

For the long time series (1982–2013), analysis of CMGLSP and VIPPHENEVI2 showed that GUD tended to occur earlier over time (< 0.2 days per year) in eastern and northeastern areas, and tended to occur later over time (< 0.3 days per year) in other regions (Fig. 5). However, CMGLSP indicated trends toward later GUD (< 0.5 days per year) in Midwest, western, and northern regions, in contrast with trends toward earlier GUD (< 0.3 day per year) indicated in those regions by VIPPHENEVI2.

3.4 Evaluation of green-up onset dates in LSP products using ground-based observations

3.4.1 Correlations

The GUD estimates from the four short-term (2001–2007) LSP products were significantly correlated ($p < 0.001$) with the lilac first leaf observations if all selected samples were considered (Fig. 6a). At the land cover level, however, their correlations were all insignificant in evergreen needle-leaf forests but significant in other land cover types except for AVHRRP GUD. Overall, the first leaf observations were most significantly correlated with the GUD estimates from MOD15PHN, followed by MOD09Q1PEVI, MCD12Q2, and AVHRRP.

The GUD values from the four LSP products were significantly correlated ($p < 0.1$) to the GPP GUD observations from all AmeriFlux sites (Fig. 6b). However, the correlations were all insignificant in mixed forests and croplands. Overall, the GPP GUD was most significantly correlated to the GUD estimates from MCD12Q2, followed by AVHRRP, MOD15PHN, and MOD09Q1PEVI.

For the long-term (1982–2013) LSP products, the GUD values from either CMGLSP or VIPPHENEVI2 were significantly correlated ($p < 0.001$) with the first leaf observations across all land cover types except in barren or sparsely vegetated area. Generally, CMGLSP GUD was more strongly correlated with ground-based observations than VIPPHENEVI2 GUD (Fig. 6c).

3.4.2 Root mean square errors

Comparison of the GUD from four short-term LSP products (MOD09Q1PEVI, MOD15PHN, MCD12Q2, and AVHRRP) with the first leaf dates showed an overall RMSE of ~20 days (Fig. 7a). The RMSE value, however, varied substantially among land cover types. According to individual comparisons of LSP products and land cover types, the largest RMSE was 29 days for both MOD09Q1PEVI GUD and MOD15PHN GUD in evergreen needle-leaf forests (Fig. 7a), whereas the smallest RMSE was 12.3 days for MCD12Q2 GUD in cropland/natural vegetation mosaics. Among the land cover types, the overall smallest RMSE values for all LSP products occurred in deciduous broadleaf forests (~15 days), followed by mixed forests and

cropland/natural vegetation mosaics; whereas, the largest RMSE appeared in evergreen needle-leaf forests and followed by savannas. Among the LSP products, the overall RMSE was smallest for AVHRRP, followed by MCD12Q2, MOD15PHN, and MOD09Q1PEVI.

RMSE was ~23 days when comparing the LSP GUD estimates with ground-based measurements of GPP GUD data across all Ameriflux sites (Fig. 7d). The RMSE values were particularly high for AVHRRP (43 days) in cropland, followed by MOD15PHN (39 days), MCD12Q2 (37 days), and MOD09Q1PEVI (30 days) in evergreen needle-leaf forests (Fig. 7d). In contrast, MOD09Q1PEVI had the smallest RMSE (11 days) in grassland. Like the comparison with the first leaf dates, the smallest RMSE (16 days) occurred in deciduous broadleaf forests for all LSP products in comparisons with GPP GUD.

The $RMSE_s$ values in MOD09Q1PEVI, MCD12Q2, and MOD15PHN were generally lower than $RMSE_u$ values when compared to the first leaf dates. In contrast, $RMSE_s$ values in AVHRRP across most land cover types and MOD15PHN in evergreen needle-leaf forests were larger than $RMSE_u$ values (Figs. 7b, c). MOD09Q1PEVI showed the smallest $RMSE_s$ values (~5 days) in most land cover types, except for grassland (17 days). Overall, smallest $RMSE_s$ with largest $RMSE_u$ values appeared in MOD09Q1PEVI, and followed by MOD15PHN, MCD12Q2, and AVHRRP.

When compared to GPP GUD at AmeriFlux sites (2001–2007), MOD09Q1PEVI generally showed smaller values for $RMSE_s$ than $RMSE_u$, except for croplands (Figs. 7e, f). Moreover, $RMSE_s$ was particularly larger than $RMSE_u$ for MOD15PHN and MCD12Q2 in evergreen needle-leaf forests, and AVHRRP had larger $RMSE_s$ than $RMSE_u$, especially for croplands. The overall pattern of $RMSE_s$ and $RMSE_u$ was very similar to that in the comparison with the first leaf dates.

Comparison of long-term (1982–2013) LSP GUD to the first leaf dates indicated that RMSE was slightly smaller in CMGLSP than VIPPHENEVI2 (Fig. 7g). In both LSP products, RMSE was approximately 30 days for evergreen needle-leaf forests, and generally fewer than 20 days for other land cover types. RMSE were particularly small (16 days) for deciduous broadleaf forests. Both CMGLSP and VIPPHENEVI2 showed relatively smaller $RMSE_s$ than $RMSE_u$ (Figs. 7h, i). However, $RMSE_s$ for CMGLSP was larger than that for VIPPHENEVI2, whereas $RMSE_u$ displayed an opposite pattern.

3.4.3 Indices of agreement

Short-term (2001–2007) LSP products generally showed good agreement with ground-based observations ($IOA > 0.5$), especially for deciduous broadleaf forests. Comparisons with the first leaf dates showed that IOA was similar for MOD09Q1PEVI, MOD15PHN and MCD12Q2 although IOA differed slightly across various land cover types (Fig. 8a). Relatively, AVHRRP GUD generally resulted in

lower IOA. All the four LSP products displayed low IOA (< 0.5) in evergreen needle-leaf forests in comparing with the first leaf dates (Fig. 8a), and in evergreen needle-leaf forests, mixed forests, and croplands in comparing with AmeriFlux GPP GUD data (Fig. 8b).

Long-term (1982–2013) LSP products showed good consistency with the first leaf dates with an IOA of 0.78 (CMGLSP) and 0.71 (VIPPHENEVI2) for all samples (Fig. 8c). Moreover, IOA values were higher for CMGLSP than VIPPHENEVI2 in each land cover type, indicating CMGLSP GUD was more strongly related to the ground observations than VIPPHENEVI2 GUD.

3.4.4 Evaluation of long-term LSP GUD estimates with trend disagreements

In the areas where GUD values from CMGLSP and VIPPHENEVI2 showed opposite trend directions (earlier vs later) (Fig. 5), their LSP GUD estimates exhibited considerably different consistencies with ground-based data. Based on 1280 site-years of the first leaf observations across the disagreement regions, it was revealed that GUD estimates from CMGLSP were more strongly correlated with ground observations ($R = 0.66$, $p < 0.001$) with higher IOA (0.78) and smaller RMSE (20.2 days) values relative to the values from VIPPHENEVI2 (Fig. 9).

4. Discussion

4.1 Evaluating LSP GUD estimates using ground observations

Reconciling ground and satellite-based phenological observations is very challenging because of the mismatch in scales. Ground observations often describe phenological events in individual trees or stands, whereas satellites capture landscape-level phenology (Zhang & Goldberg, 2011; Ganguly et al., 2010; Liang & Schwartz, 2009; Morisette et al., 2009; Schwartz & Hanes, 2009; White et al., 2009; Zhang et al., 2006). Thus, the point observations of the first leaf dates in USA-NPN unlikely represent well the landscape phenology characterized in LSP products although averaging multiple observations within a pixel-year in this study likely reduced the scale influence to some extent on the comparisons between ground and satellite observations. Moreover, this study only used the first leaf dates from honeysuckle or lilac, which to some extent limits the effectiveness of ground-based reference. It is expected to include as many observations from USA-NPN as possible in assessing LSP data in future. Measurements of GPP at AmeriFlux sites are less vulnerable to the spatial mismatch because the footprint of AmeriFlux is between 1.1 km² to 5.0 km² that cover about 1–20 pixels in the LSP products used in this study. Because of relatively large footprint and regular daily measurements, GPP measurements at AmeriFlux sites have been widely used to validate LSP in previous studies (Baldocchi et al., 2001; Baldocchi et al., 2005; Gonsamo et al., 2012; Melaas et al., 2013; Wu et al., 2016). However, the drawbacks of using GPP to validate LSP products include the

limited number of sites, the variation of the footprint areas with the AmeriFlux sites and the wind strength and direction, and the indirect linkage between GPP and greening or presence of leaves on trees (Chen et al., 2011). Our results indicated that LSP GUD estimates were generally correlated to field observations well across the CONUS, including the significant correlations ($p < 0.001$) (Fig. 6) and good IOA values (> 0.6) (Fig. 8) but relatively high RMSE (~ 20 days) (Fig. 7).

Although the spring phenology could vary considerably among vegetation species within a local area or satellite footprint, the coexistence of various species is believed to lead the similar spatial variation patterns in their phenological dates across a region, which are continuous and local-climate dependent (Westerling et al., 2006; Piao et al., 2007; Zhang & Goldberg, 2011; Mazer et al., 2015; Piao et al., 2015; Delbart et al., 2015; Wu et al., 2016). Thus, the differences between field observations and LSP estimates across the CONUS in multiple years could be relatively scale-invariant (Wu et al., 2014; Allstadt et al., 2015; Zhang, 2015). As a result, comparison of the LSP GUD estimates with observations from the USA-NPN and AmeriFlux revealed the temporal and spatial consistencies and discrepancies within a LSP product and among various LSP products. Of course, the comparison results could be to some extent affected by different spatial resolutions among the LSP products and such comparison was unable to provide the exact accuracy or absolute error of LSP GUD estimates. To validate LSP products appropriately, field observations from individual species (such as USA-NPN), canopy cameras (such as PhenoCam), and flux towers should be temporally and spatially scaled up to match satellite pixels

across various ecosystems and geographical regions (Liang & Schwartz, 2009). In particular, with the widely available observations from the PhenoCam network (Richardson et al. 2009), LSP could be more properly validated, which has been demonstrated in several studies (Keenan & Richardson 2015; Klosterman et al. 2014; Rodriguez-Galiano et al. 2015).

4.2 Comparing green-up onset dates among LSP products

Because of the complexity in LSP and the spatial mismatch between field observations and LSP detections, a single parameter is not able to quantify the quality of a LSP product. Comparing with ground observations, a LSP product could show a good agreement based on one comparison parameter but a poor agreement for another parameter. Moreover, a LSP product could perform well in one ecosystem or land cover type but poorly in another. Based on five comparison parameters (p -value, RMSE, RMSE_s, RMSE_u, and IOA) across various land cover types, the spatial and temporal consistencies of LSP products were revealed. Among the short-term (2001–2007) LSP products, the GUD estimates from MCD12Q2, MOD09Q1PEVI, and MOD15PHN had similar correlations with ground observations of the first leaf dates and GPP GUD data while the correlations were relatively poor for AVHRRP GUD. Specially, RMSE was ~20 days with slightly small RMSE_u and agreement index was 0.7–0.8 in the comparisons of the GUD estimates from MCD12Q2, MOD09Q1PEVI, and MOD15PHN with ground

observations. Although AVHRRP GUD presented poor quality in p -value and IOA, it showed relatively low unsystematic error (**RMSE_u**). Clearly, the four LSP products had their advantages in some land cover types and individual parameters. However, close examinations of all the parameters showed that the MCD12Q2 GUD had overall higher p -values, better IOA, lower RMSE, and lower RMSE_u for most land cover types, which suggests that the MCD12Q2 GUD has overall higher consistence with ground observations.

Between the two long-term (1982–2013) LSP products of CMGLSP and VIPPHENEVI2, the CMGLSP GUD appeared to match ground-based observations better as evidenced by stronger significant correlations (p -value), lower RMSE, lower RMSE_u, and higher agreement index across various land cover types. The better correlation also appeared in the pixels where GUD trends from CMGLSP and VIPPHENEVI2 were opposite. This suggests that CMGLSP GUD is able to characterize long-term GUD variation.

Intercomparison the GUD estimates among the six LSP products revealed that highest correlation coefficient appeared in MOD09Q1PEVI vs MOD15PHN, followed by CMGLSP vs MOD15PHN, and CMGLSP vs MCD12Q2, whereas the lowest RMSE occurred in CMGLSP vs MOD15PHN, followed by CMGLSP vs MCD12Q2, and CMGLSP vs MOD09Q1PEVI. The high correlation coefficient with low RMSE suggest these LSP products are valuable for various applications although one product could be better than others in a specific location or ecosystem. In contrast, AVHRRP was poorly correlated to all other LSP products with high RMSE.

It is likely high uncertainty exists in AVHRRP GUD. Moreover, the trend analyses indicated that the similar interannual variations among all the products only occurred in local regions while discrepancies were distributed across the CONUS.

The spatial and temporal distributions of missing values in LSP products could limit their applications. Overall missing GUD detections occurred over large areas, particularly in southwestern areas, in MCD12Q2 and CMGLSP products. In contrast, MOD09Q1PEVI and VIPPHENEVI2 almost produced wall-to-wall coverage every year. This difference raises a question if LSP GUD should be produced regardless the actual vegetation seasonality. This is particularly true in the following cases. In the southwestern CONUS, barren or sparsely vegetated pixels are dominant, where vegetation could grow well in wet years while no green vegetation appears in drought years. In evergreen forests, seasonal variation in vegetation indices could be very subtle, which could be as small as the noise of the vegetation index. In the southern CONUS, consecutive cloud contamination in satellite observations could be longer than a month, which obstructs the reconstruction of reliable trajectories of vegetation growth. As a result, missing values in a LSP product does not necessarily represent poor detections. Although applications of LSP products could have different requirements, it would be reasonable for a LSP product to detect all possible phenological events accompanying with well-defined quality of LSP detections.

Further, to investigate the discrepancies and consistencies of GUD estimates from various LSP products, it is necessary to understand the methods of a LSP product development. This is particularly important because LSP GUD is currently defined

using different criteria in the temporal variation of vegetation indices and temporal data processing approaches (Zhang et al., 2014, Tan et al., 2011; Reed et al., 1994; White et al., 2009). Indeed, the GUD estimates from MOD09Q1PEVI and MOD15PHN were retrieved using the same data processing algorithms (Tan et al., 2011) so that the GUD difference was resulted from the vegetation indices (EVI for MOD09Q1PEVI and LAI for MOD15PHN). In contrast, CMGLSP and VIPPHENEVI2 products were both developed using the same vegetation index (daily EVI2). However, relative to VIPPHENEVI2 GUD, CMGLSP GUD presented the higher consistencies with ground observations (Figs. 6–8). This suggests that the CMGLSP algorithm is more robust to estimate GUD.

4.3 Assessing green-up onset dates in LSP products across land cover types

The quality of LSP detections is strongly influenced by ecosystem processes and vegetation types (White et al., 2009; Gonsamo et al., 2012; Hmimina et al., 2013; Melaas et al., 2015). As revealed in previous studies, LSP GUD is best estimated in deciduous forests because of the strong seasonal changes in greenness, while the detection is difficult in evergreen forests where seasonal changes in greenness are very subtle (Baldocchi et al., 2005; Zhang et al., 2006; Kamel et al., 2008; Garrity et al., 2011; Guyon et al., 2011; Gonsamo et al., 2012; Hufkens et al., 2012; Hmimina et al., 2013; Melaas et al., 2013; Ryu et al., 2014; Wu et al., 2014; Shen et al., 2014; Liu et al., 2016). Our results further bear this out. The agreements between LSP GUD and

ground-based observations varied considerably across land cover types (Figs. 6–8).

The LSP GUD estimates appeared best in deciduous broadleaf forests as evidenced by relatively strong correlations and low RMSE values with ground-based observations (Figs. 6–8). In contrast, the LSP GUD estimates varied most in western and southeastern parts of the CONUS (Fig. 2), where land cover was dominated by evergreen forests or barren or sparse vegetation (Fig. 1).

5. Conclusions

Validation of LSP products has been highly expected to quantify the quality or accuracy (Delbart et al., 2015; White et al., 2009; Zhang et al., 2006; Liang et al., 2011). However, it is currently challenging because *in-situ* data that spatially match satellite footprints are barely available. On the other hand, widely acceptable criteria to define the LSP GUD were not available. Thus, a high quality LSP product should be self-consistent spatially and temporally.

In this study, we investigated GUD in six regularly updated LSP products by intercomparing their consistencies and discrepancies, evaluating them with ground observations, and identifying the retrieval efficiency. Our main findings are accordingly summarized as follows. First, GUD estimates from LSP products were generally strongly correlated each other, except for AVHRRP, which had the largest RMSE relative to other products. In western and southeastern areas, mostly covered by evergreen forest or sparse vegetation, the GUD estimates varied substantially

among six land surface phenology products. Second, for the short-term (2001–2007) LSP products, **MCD12Q2, MOD09Q1PEVI, MOD15PHN** show the similar **RMSE values and correlations** when compared to ground-based observations, while **AVHRRP** has relatively weak correlation and agreement with ground-based estimates. Considering all the comparison parameters, the **MCD12Q2** GUD estimates showed slightly better consistencies with ground observations. For the long-term (1982–2013) LSP products, CMGLSP had a relatively higher consistency with ground observations in all the comparison parameters than VIPPHENEVI2 did. Third, the spatial and temporal distributions of the missing values occurred in most LSP products. Relatively, missing values were less in **MOD09Q1PEVI** and **VIPPHENEVI2** while they were large in MCD12Q2 and CMGLSP. Because missing values could reflect the absence of vegetation seasonality or the consecutive cloud contaminations in temporal satellite observations, detections of phenology for all the pixels do not necessarily provide better quality of a LSP product.

Acknowledgements

The authors acknowledge Jake Weltzin and Theresa Crimmins (USA National Phenology Network) for their useful comments and the following data support: Lilac/honeysuckle data were provided by the USA National Phenology Network, Joseph M. Caprio, Mark D. Schwartz, and all contributors to past U.S. Dept. of Agriculture regional phenology projects; MCD12Q2 from Land Processes Distributed Active Archive Center (LP DAAC); Long Term Land Surface Phenology Record

from ftp://bruin.sdstate.edu/ftp/XYZ/AVHRR_EVI2/ in South Dakota State University; Vegetation Index and Phenology multi-sensor Phenology product from Vegetation Index and Phenology Lab, The University of Arizona; MOD09PHN and MOD15PHN from Level 1 and Atmosphere Archive and Distribution System (LAADS), which is part of the Terrestrial Information Systems Laboratory at NASA Goddard Space Flight Center in Greenbelt, MD; AVHRR phenology product from USGS/ EROS. This study was supported by the National Key Research and Development Program of China [2016YFB0501501], the Strategic Priority Research Program (B) of the Chinese Academy of Sciences [XDB03030406], National Natural Science Foundation of China [41571423], Youth Innovation Promotion Association, Chinese Academy of Sciences [Y4YR1300QM], the key Research Program of Frontier Sciences, CAS to C. Wu [QYZDB-SSW-DQC011], and Director Foundation of Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences [Y6SJ020CX].

Reference

Allstadt, A. J., Vavrus, S. J., Heglund, P. J., Pidgeon, A. M., Thogmartin, W. E., & Radeloff, V. C. (2015). Spring plant phenology and false springs in the conterminous US during the 21st century. *Environmental Research Letters*, 10 (10), 104008.

- Angert, A., Biraud, S., Bonfils, C., Henning, C. C., Buermann, W., Pinzon, J., ... & Fung, I. (2005). Drier summers cancel out the CO₂ uptake enhancement induced by warmer springs. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31), 10823–10827.
- Ault, T. R., Schwartz, M. D., Zurita-Milla, R., Weltzin, J. F., & Betancourt, J. L. (2015). Trends and natural variability of spring onset in the coterminous United States as evaluated by a new gridded dataset of spring indices. *Journal of Climate*, 28(21), 8363–8378.
- Badeck, F. W., Bondeau, A., Böttcher, K., Doktor, D., Lucht, W., Schaber, J., & Sitch, S. (2004). Responses of spring phenology to climate change. *New Phytologist*, 162 (2), 295–309.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., ... & Fuentes, J. (2001). FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11), 2415–2434.
- Baldocchi, D. D., Black, T. A., Curtis, P. S., Falge, E., Fuentes, J. D., Granier, A., ... & Valentini, R. (2005). Predicting the onset of net carbon uptake by deciduous forests with soil temperature and climate data: a synthesis of FLUXNET data. *International Journal of Biometeorology*, 49(6), 377–387.

- Burkle, L. A., Marlin, J. C., & Knight, T. M. (2013). Plant–pollinator interactions over 120 years: loss of species, co–occurrence, and function. *Science*, 339 (6127), 1611–1615.
- Chapman, D. S., Haynes, T., Beal, S., Essl, F., & Bullock, J. M. (2014). Phenology predicts the native and invasive range limits of common ragweed. *Global Change Biology*, 20(1), 192–202.
- Chen, B., Coops, N. C., Fu, D., Margolis, H. A., Amiro, B. D., Barr, A. G., ... & Lafleur, P. M. (2011). Assessing eddy–covariance flux tower location bias across the Fluxnet–Canada Research Network based on remote sensing and footprint modelling. *Agricultural and forest meteorology*, 151(1), 87–100.
- Chuine, I., & Beaubien, E. G. (2001). Phenology is a major determinant of tree species range, *Ecology Letters*, 4(5), 500–510.
- Dash, J., Jeganathan, C., & Atkinson, P. M. (2010). The use of MERIS Terrestrial Chlorophyll Index to study spatio–temporal variation in vegetation phenology over India. *Remote Sensing of Environment*, 114 (7), 1388–1402.
- de Beurs, K.M., & Henebry, G.M. (2010). Spatio–temporal statistical methods for modeling land surface phenology. In I.L. Hudson, & M.R. Keatley (Eds.), *Phenological Research – Methods for Environmental and Climate Change Analysis* (pp. 177–208). New York: Springer.
- Didan, K. (2010). Multi–satellite earth science data record for studying global vegetation trends and changes. In *Proceedings of the 2010 international*

geoscience and remote sensing symposium, Honolulu, HI, USA (Vol. 2530, p. 2530).

Didan, K. Barreto A.M, Miura T., Tseng–Ayush J., Zhang X., Friedl M., ... & Meyer D. (2016). Multi–Sensor Vegetation Index and Phenology Earth Science Data Records: Algorithm Theoretical Basis Document and User Guide Version 4.0. (https://vip.arizona.edu/VIP_ATBD_UsersGuide.php, last accessed March 22nd, 2016)

Delbart, N., Beaubien, E., Kergoat, L., Le & Toan, T. (2015). Comparing land surface phenology with leafing and flowering observations from the PlantWatch citizen network. *Remote Sensing of Environment*, 160, 273–280.

Denny, E. G., Gerst, K. L., Miller-Rushing, A. J., Tierney, G. L., Crimmins, T. M., Enquist, C. A., ... & Weltzin, J. F. (2014). Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International journal of biometeorology*, 58(4), 591–601.

Friedl, M., Henebry, G., Reed, B., Huete, A., White, M., Morisette, J., Nemani, R., Zhang, X., & Myneni, R. (2006). Land Surface Phenology. A Community White Paper requested by NASA. April 10. Available online at ftp://zeus.geog.umd.edu/Land_ESDR/Phenology_Friedl_whitepaper.pdf.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X. (2010). MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114 (1),

168–182.

- Gallinat, A. S., Primack, R. B., & Wagner, D. L. (2015). Autumn, the neglected season in climate change research. *Trends in Ecology & Evolution*, 30(3), 169–176.
- Ganguly, S., Friedl, M. A., Tan, B., Zhang, X., & Verma, M. (2010). Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, 114(8), 1805–1816.
- Garrity, S. R., Bohrer, G., Maurer, K. D., Mueller, K. L., Vogel, C. S., & Curtis, P. S. (2011). A comparison of multiple phenology data sources for estimating seasonal transitions in deciduous forest carbon exchange. *Agricultural and Forest Meteorology*, 151 (12), 1741–1752.
- Gerst, K. L., Kellermann, J. L., Enquist, C. A., Rosemartin, A. H., & Denny, E. G. (2015). Estimating the onset of spring from a complex phenology database: trade-offs across geographic scales. *International journal of biometeorology*, 1–10.
- Gonsamo, A., Chen, J. M., Wu, C., & Dragoni, D. (2012). Predicting deciduous forest carbon uptake phenology by upscaling FLUXNET measurements using remote sensing data. *Agricultural and forest meteorology*, 165, 127–135.
- Gonsamo, A., & Chen, J. M. (2016). Circumpolar vegetation dynamics product for global change studies. *Remote Sensing of Environment*, 182, 13–26.

- Guyon, D., Guillot, M., Vitasse, Y., Cardot, H., Hagolle, O., Delzon, S., & Wigneron, J.P. (2011). Monitoring elevation variations in leaf phenology of deciduous broadleaf forests from SPOT/VEGETATION time-series. *Remote Sensing of Environment*, 115, 615–627.
- Hargrove, W. W., Spruce, J. P., Gasser, G. E., & Hoffman, F. M. (2009). Toward a national early warning system for forest disturbances using remotely sensed canopy phenology. *Photogrammetric Engineering & Remote Sensing*, 75(10), 1150–1156.
- Hufkens, K., Friedl, M., Sonnentag, O., Braswell, B. H., Milliman, T., & Richardson, A. D. (2012). Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment*, 117, 307–321.
- Hmimina, G., Dufrêne, E., Pontailier, J. Y., Delpierre, N., Aubinet, M., Caquet, B., ... & Gross, P. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment*, 132, 145–158.
- Jeong, S. J., Medvigy, D., Shevliakova, E., & Malyshev, S. (2013). Predicting changes in temperate forest budburst using continental-scale observations and models. *Geophysical Research Letters*, 40(2), 359–364.
- Kamel, S., Guerric, le M., Eric, D., Christophe, F., Nicolas, D., Erwin, U., & Sébastien, C. (2008). Evaluation of the onset of green-up in temperate deciduous

- broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sensing of Environment* 112 (5), 2643–2655.
- Kellermann, J. L., Enquist, C. A., Humple, D. L., Seavy, N. E., Rosemartin, A., Cormier, R. L., & Barnett, L. (2015). A Bird’s–Eye View of the USA National Phenology Network. *Phenological Synchrony and Bird Migration: Changing Climate and Seasonal Resources in North America*, 47, 47.
- Keenan, T.F., & Richardson, A.D. (2015). The timing of autumn senescence is affected by the timing of spring phenology: implications for predictive models. *Global Change Biology*, 21, 2634-2641.
- Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M.A., & Richardson, A.D. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, 11, 4305-4320
- Knudson, W. A. (2012). The economic impact of this spring’s weather on the fruit and vegetable sectors. *Working Pap. 01*, 52012.
- Liang, L., & Schwartz, M. (2009). Landscape phenology: An integrative approach to seasonal vegetation dynamics. *Landscape Ecology*, 24(4), 465–472.
- Liang, L.A., Schwartz, M.D., Fei, S.L. (2011). Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. *Remote Sensing of Environment*, 115, 143–157.

- Lacaze, R., Baret, F., Camacho, F., d'Andrimont, R., Freitas, S. C., Pacholczyk, P., ... & Calvet, J. C. (2011). Geoland2—towards an operational GMES land monitoring core service: the biogeophysical parameter core mapping service. In *Proceedings of the 34th International Symposium on Remote Sensing of Environment*.
- Liu, Y., Wu, C., Peng, D., Xu, S., Gonsamo, A., Jassal, R. S., ... & Chen, J. M. (2016). Improved modeling of land surface phenology using MODIS land surface reflectance and temperature at evergreen needleleaf forests of central North America. *Remote Sensing of Environment*, 176, 152–162.
- Ma, X., Huete, A., Yu, Q., Restrepo Coupe, N., Davies, K.P., Broich, M., ... & Eamus, D. (2013). Spatial patterns and temporal dynamics in savanna vegetation phenology across the North Australian Tropical Transect, *Remote Sensing of Environment*, 139(1), 97–115.
- Mazer, S. J., Gerst, K. L., Matthews, E. R., & Evenden, A. (2015). Species-specific phenological responses to winter temperature and precipitation in a water – limited ecosystem. *Ecosphere*, 6(6), 1–27.
- McCabe, G.J., Ault, T. R., Cook, B.I., Betancourt, J. L., & Schwartz, M. D. (2012). Influences of the El Niño Southern oscillation and the pacific decadal oscillation on the timing of the North American spring. *International Journal of Climatology*, 32, 2301–2310.
- Melaas, E. K., Richardson, A. D., Friedl, M. A., Dragoni, D., Gough, C. M., Herbst, M., ... & Moors, E. (2013). Using FLUXNET data to improve models of

- springtime vegetation activity onset in forest ecosystems. *Agricultural and Forest Meteorology*, 171, 46–56.
- Melaas, E. K., Friedl, M. A., & Richardson, A. D. (2015). Multiscale modeling of spring phenology across Deciduous Forests in the Eastern United States. *Global change biology*, 22(6), 792–805.
- Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., ... & Zust, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology*, 12(10):1969–1976.
- Morisette, J. T., Richardson, A. D., Knapp, A. K., Fisher, J. I., Graham, E. A., Abatzoglou, J., ... & Liang, L. (2009). Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Frontiers in Ecology and the Environment*, 7(5), 253–260.
- Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, 37, 637–669.
- Piao, S., Friedlingstein, P., Ciais, P., Viovy, N., & Demarty, J. (2007). Growing season extension and its impact on terrestrial carbon cycle in the Northern Hemisphere over the past 2 decades. *Global Biogeochemical Cycles*, 21(3), 1148–1154.
- Piao, S., Tan, J., Chen, A., Fu, Y. H., Ciais, P., Liu, Q., ... & Li, Y. (2015). Leaf onset in the northern hemisphere triggered by daytime temperature. *Nature communications*, 6, 6911.

- Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of vegetation science*, 5(5), 703–714.
- Reed, B.C., (2006). Trend analysis of time–series phenology of North America derived from satellite data. *GIScience and Remote Sensing*, 43, 1–15.
- Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., & Ollinger, S.V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, 19, 1417–1428.
- Rodriguez-Galiano, V.F., Dash, J., & Atkinson, P.M. (2015). Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. *Geophysical Research Letters*, 42, 2253-2260
- Rosenzweig, C., Casassa, G., Karoly, D.J. Imeson, A., Liu, C., Menzel, A., ... Tryjanowski, P. (2007). Assessment of observed changes and responses in natural and managed systems. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 79–131.
- Rosemartin, A.H., Denny, E.G., Weltzin, J.F., Marsh, R.L., Wilson, B.E., Mehdipoor, H., ... Schwartz, M.D. (2015). Lilac and honeysuckle phenology data 1956–2014. Scientific Data 2: 150038. <http://dx.doi.org/10.1038/sdata.2015.38>

- Ryu, Y., Lee, G., Jeon, S., Song, Y., & Kimm, H. (2014). Monitoring multi-layer canopy spring phenology of temperate deciduous and evergreen forests using low-cost spectral sensors. *Remote Sensing of Environment*, 149, 227–238.
- Schwartz, M. D., Ahas, R., & Aasa, A. (2006). Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biology*, 12, 343–351.
- Schwartz, M. D., & Hanes, J. M. (2009). Intercomparing multiple measures of the onset of spring in eastern North America. *International Journal of Climatology*, 30(11), 1614–1626.
- Schwartz, M. D., Betancourt, J. L., & Weltzin, J. F. (2012). From Caprio's lilacs to the USA National Phenology Network. *Frontiers in Ecology and the Environment*, 10 (6), 324–327.
- Shen, M., Tang, Y., Desai, A. R., Gough, C., & Chen, J. (2014). Can EVI-derived land surface phenology be used as a surrogate for phenology of canopy photosynthesis? *International Journal of Remote Sensing*, 35, 1162–1174.
- Tan, B., Morisette, J. T., Wolfe, R. E., Gao, F., Ederer, G. A., Nightingale, J., & Pedelty, J. A. (2008). Vegetation phenology metrics derived from temporally smoothed and gap-filled MODIS data. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International* (Vol. 3, pp. III–593).
- Tan, B., Morisette, J. T., Wolfe, R. E., Gao, F., Ederer, G. A., Nightingale, J., & Pedelty, J. A. (2011). An enhanced TIMESAT algorithm for estimating

- vegetation phenology metrics from MODIS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 361–371.
- USA National Phenology Network. (2015). Data from: Lilac and honeysuckle phenology data 1956–2014. Dryad Digital Repository.
<http://dx.doi.org/10.5061/dryad.0262m>.
- Verstraete, M. M., Gobron, N., Aussedat, O., Robustelli, M., Pinty, B., Widlowski, J. L., & Taberner, M. (2008). An automatic procedure to identify key vegetation phenology events using the JRC–FAPAR products. *Advances in Space Research*, 41(11), 1773–1783.
- Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western US forest wildfire activity. *Science*, 313(5789), 940–943.
- White, M.A., Thornton, P.E., & Running, S.W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global biogeochemical cycles*, 11(2): 217–34.
- White, M. A., Beurs, D., Kirsten, M., Didan, K., Inouye, D.W., Ruchardson, A. D., & Leeuwen, V. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, 15 (10), 2335–2359.

- White, M. A., & Nemani, R. R. (2003). Canopy duration has little influence on annual carbon storage in the deciduous broad leaf forest. *Global change biology*, 9(7), 967–972.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2, 184–194.
- Wilson, K., & Baldocchi, D. (2001). Comparing independent estimates of carbon dioxide exchange over 5 years at a deciduous forest in the southeastern United States. *Journal of Geophysical Research. D. Atmospheres*, 106, 34.
- Wu, C., Chen, J. M., Black, T. A., Price, D. T., Kurz, W. A., Desai, A. R., ... & Dragoni, D. (2013). Interannual variability of net ecosystem productivity in forests is explained by carbon flux phenology in autumn. *Global Ecology and Biogeography*, 22(8), 994–1006.
- Wu, C., Gonsamo, A., Gough, C. M., Chen, J. M., & Xu, S. (2014). Modeling growing season phenology in North American forests using seasonal mean vegetation indices from MODIS. *Remote Sensing of Environment*, 147, 79–88.
- Wu, C., Hou, X., Peng, D., Gonsamo, A., & Xu, S. (2016). Land surface phenology of China's temperate ecosystems over 1999–2013: Spatial–temporal patterns, interaction effects, covariation with climate and implications for productivity. *Agricultural and Forest Meteorology*, 216, 177–187.
- Zhang, Q.Y., Xiao, X.M., Braswell, B., Linder, E., Ollinger, S., Smith, M.L., ... & Minocha, R. (2006). Characterization of seasonal variation of forest canopy in a

- temperate deciduous broadleaf forest, using daily MODIS data. *Remote Sensing of Environment*, 105(3), 189–203
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote sensing of environment*, 84(3), 471–475.
- Zhang, X., Friedl, M. A., & Schaaf, C. B. (2006). Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *Journal of Geophysical Research: Biogeosciences*, 111(G4).
- Zhang, X., & Goldberg, M. (2011). Monitoring fall foliage coloration dynamics using time-series satellite data. *Remote Sensing of Environment*, 115, 382–391.
- Zhang, X., Tan, B., & Yu, Y. (2014). Interannual variations and trends in global land surface phenology derived from enhanced vegetation index during 1982–2010. *International journal of biometeorology*, 58(4), 547–564.
- Zhang, X. (2015). Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data. *Remote Sensing of Environment*, 156, 457–472.

<InlineImage1>

Fig. 1. Spatial distribution of *in-situ* observations superimposed over a land cover map. The *in-situ* observations include the first leaf dates of lilacs from the USA National Phenology Network (USA-NPN) sites and the green-up onset dates from daily gross primary production at AmeriFlux sites. Land cover types include evergreen needle-leaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needle-leaf Forest (DNF), deciduous broadleaf forest (DBF), mixed forest (MF), shrublands (SHR), savannas (SAV), grasslands (GRA), permanent wetlands (WET), croplands (CRO), urban and built-up (URB), cropland/natural vegetation mosaic (CNV), snow and ice (SAI), and barren or sparsely vegetated (BSV).

<InlineImage2>

Fig. 2. Spatial pattern of mean green-up onset dates in day of year (DOY) from MCD12Q2 (a), MOD09Q1PEVI (b), MOD15PHN (c), CMGLSP (d), VIPPHENEVI2 (e), and AVHRRP (f). The acronym for land surface phenology products is the same as defined in Table 1.

<InlineImage3>

Fig. 3. Relationships of green-up onset dates, reported as day of year (DOY), among MCD12Q2, MOD09Q1PEVI, MOD15PHN, CMGLSP, VIPPHENEVI2, and AVHRRP. The acronym for land surface phenology products is the same as defined in Table 1. Colour tone level shows the density of scatter points (see panel (a) for scale). *** indicates the significant correlation at the level of $p < 0.001$.

<InlineImage4>

Fig. 4. Short-term trends (days per year) of green-up onset dates (GUD) during 2001–2007 from MCD12Q2 (a), MOD09Q1PEVI (b), MOD15PHN (c), CMGLSP (d), VIPPHENEVI2 (e), and AVHRRP (f). Positive values indicate delays in GUD while negative values indicate advances in GUD. The acronym for land surface phenology products is the same as defined in Table 1.

<InlineImage5>

Fig. 5. Long-term trends (days per year) of green-up onset dates (GUD) during 1982–2013 from CMGLSP (a) and VIPPHENEVI2 (b). Positive values indicate delays in GUD while negative values indicate advances in GUD. The acronym for land surface phenology products is the same as defined in Table 1.

Fig. 6. The correlation significance (p -value) between ground observations and green-up onset dates (GUD) from land surface phenology products. NS indicates insignificant. The ground observations in (a) and (c) were the lilac first leaf dates (USA-NPN) for the periods of 2001–2007 and 1982–2013, respectively. The ground observations in (b) were GPP GUD at AmeriFlux sites during the period of 2001–2007. The number at the right end of the rows in (a), (b) and (c) indicates the site-years of field observations. The acronym for land surface phenology products and land cover types is the same as defined in Table 1 and Fig. 1. "Total" includes all observations.

<InlineImage7>

Fig. 7. Comparison of the green-up onset dates (GUD) estimates in land surface phenology (LSP) products with the ground observations using root mean square error (RMSE, days), systematic RMSE ($RMSE_s$, days) and unsystematic RMSE ($RMSE_u$, days). The four short-term (2001–2007) LSP products were compared with the USA National Phenology Network (USA-NPN) first leaf dates (a–c) and the GPP GUD at AmeriFlux sites (d–f). The two long-term (1982–2013) LSP products were compared with the USA-NPN first leaf dates (g–i). The acronym for LSP products and land cover types is the same as defined in Table 1 and Fig. 1, respectively. "Total" includes all observations.

<InlineImage8>

Fig. 8. The index of agreement (IOA) between the land surface phenology (LSP) green-up onset dates (GUD) estimates and the ground-based observations. The GUD estimates in four short-term (2001–2007) LSP products were compared with the USA National Phenology Network (USA-NPN) first leaf dates (a) and the AmeriFlux GPP GUD dates (b). The GUD values in two long-term (1982–2013) LSP products were compared with the USA-NPN first leaf dates (c). The acronym for LSP products and land cover types is the same as defined in Table 1 and Fig. 1, respectively. "Total" includes all observations.

<InlineImage9>

Fig. 9. Comparisons of the green-up onset dates (GUD) estimates from CMGLSP and VIPPHENEVI2 using the USA National Phenology Network first leaf dates at sites where CMGLSP and VIPPHENEVI2 showed opposite temporal GUD trends in 1982–2013 (Fig. 5). See the explanations of the acronyms for CMGLSP and VIPPHENEVI2 in Table 1. *** indicates the significant correlation at the level of $p < 0.001$.