



OPEN

DATA DESCRIPTOR

Vegetation photosynthetic phenology dataset in northern terrestrial ecosystems

Jing Fang^{1,2}, Xing Li³, Jingfeng Xiao⁴, Xiaodong Yan⁵, Bolun Li³ & Feng Liu^{1,2} ✉

Vegetation phenology can profoundly modulate the climate-biosphere interactions and thus plays a crucial role in regulating the terrestrial carbon cycle and the climate. However, most previous phenology studies rely on traditional vegetation indices, which are inadequate to characterize the seasonal activity of photosynthesis. Here, we generated an annual vegetation photosynthetic phenology dataset with a spatial resolution of 0.05 degrees from 2001 to 2020, using the latest gross primary productivity product based on solar-induced chlorophyll fluorescence (GOSIF-GPP). We combined smoothing splines with multiple change-point detection to retrieve the phenology metrics: start of the growing season (SOS), end of the growing season (EOS), and length of growing season (LOS) for terrestrial ecosystems above 30° N latitude (Northern Biomes). Our phenology product can be used to validate and develop phenology or carbon cycle models and monitor the climate change impacts on terrestrial ecosystems.

Background & Summary

Vegetation phenology, the cycle sequence of the vital activities, is a highly sensitive indicator of the climate impacts on terrestrial ecosystems^{1–4}. Most phenology studies focus on the structural changes of plants, such as using the growth process of leaves represented by the greenness indicators^{5,6}. However, these indices work well for capturing the variations in chlorophyll content or structural changes but are less sensitive to physiological changes in vegetation photosynthesis, especially for evergreen vegetation^{7,8}. Besides, the vegetation indices such as the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) have been widely used to estimate GPP^{9–11}. Recent studies found that the methods based on vegetation indices cannot capture the photosynthesis changes in some vegetation types (e.g., evergreen forests) since the greenness and photosynthesis are sometimes decoupled^{12,13}. The inaccurate estimation of vegetation greenness-based phenology can lead to substantial uncertainties in estimating plant productivity and carbon sequestration^{3,11,14}.

The plant photosynthetic cycle on the seasonal time scale is termed as ‘vegetation photosynthetic phenology’. Unlike the structure change by the traditional phenology, such as bud break and leaf coloring, vegetation photosynthetic phenology represents the functional aspects of plant activities¹⁵. Plants regulate the carbon cycle process through photosynthesis, and the changes in photosynthetic phenology have a feedback effect on climate¹⁵. Therefore, tracking photosynthetic phenology on large scales may provide essential clues about the carbon cycle and help understand the drivers of carbon dynamics¹⁶. The photosynthetic phenology definition is based on the photosynthesis transition dates extracted from the gross primary productivity (GPP) time series. Thus, the accuracy of extracted phenology metrics largely depends on the data source and resolution of GPP. For example, the coarse spatial resolution of satellite data may include the mixed cells where the mixture of species in different phenological states appear simultaneously². Currently, the GPP can either be derived from Eddy Covariance (EC) flux towers at the ecosystem scale or from satellite remote sensing or modeling at the regional or global scale¹⁷. The EC technique, which is considered as the most accurate observation method¹⁸, has provided long-term GPP estimates for more than 20 years. However, the spatial distribution limits these observations, and some key areas are still underrepresented¹⁷. For example, recent studies report that the

¹CAS Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, 430074, China. ²Center of Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, 430074, China. ³Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea. ⁴Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA. ⁵State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China. ✉e-mail: liufeng@wbqcas.cn

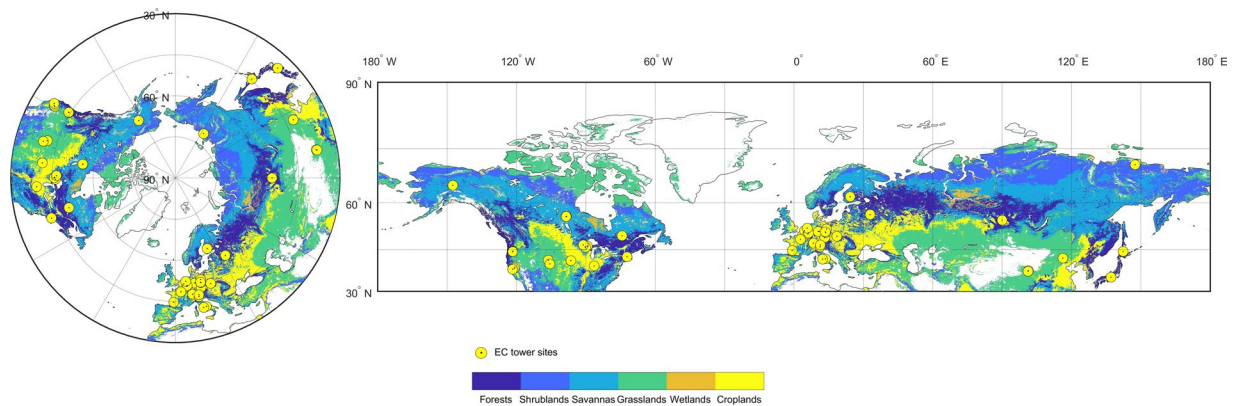


Fig. 1 The spatial distribution of vegetation types and EC tower sites under the different projections in the Northern Biomes (0.05° spatial resolution). The left part uses the projection of Equidistant Azimuthal, and the right part uses the projection of Bolshoi Sovietskii Atlas Mira.

warming-related greening trend in Arctic and boreal regions is identified as one of the clearest examples of climate change impacts on carbon cycles in terrestrial ecosystems¹⁹. Arctic amplification enhances the effects of climate change on vegetation in the regions, but only a few EC sites provide public datasets in these regions^{20,21}. GPP derived from satellite remote sensing can investigate large-scale phenology across the globe²². Recently, the emergence of satellite-based solar-induced chlorophyll fluorescence (SIF) has offered unprecedented opportunities for developing more accurate photosynthetic phenology data products on large scales^{7,23–25}. SIF, a signal emitted by plant chlorophyll molecules after absorbing photosynthetically active radiation (APAR), is considered an effective tool for diagnosing terrestrial photosynthesis and a better proxy of GPP than traditional vegetation indices^{8,12,26–30}. Based on the SIF product, recent studies used the relationship between the GPP and SIF to estimate the regional or global GPP (SIF-GPP)^{31,32}. Previous studies reported that SIF-GPP could better capture the GPP dynamics in evergreen vegetation and dryland ecosystems than traditional vegetation indices^{12,33}.

In addition, the retrieval of phenology in previous studies mainly used a logistic regression model to fit the time series of smoothed vegetation indices or GPP. The predetermined thresholds or inflection points (e.g., using the peaks in the second derivative as the points) are identified as the transition dates of phenology in the fitted curve^{6,34,35}. However, this method needs to reconstruct the original data sequence using a double-sigmoidal logistic model and thus results in uncertainty from the model parameterization³⁶. Furthermore, this method usually captures a single growing season instead of multiple growing seasons in a given year⁶. Correspondingly, Richardson, *et al.*³⁷ proposed a method that combined smoothing spline and multiple change-point detection to retrieve the phenology transition dates from the camera data. They constructed the PhenoCam network to automate the monitoring of canopy phenology in North America using near-surface remote sensing (i.e., cameras). Their method has excellent strength in two aspects: (1) it is not limited by the uncertainty of additional model parameters; (2) it can also be applied in ecosystems having multiple growing seasons. The method has been successfully used at multiple sites in North America³⁷ and can be further extended to large scales.

Here, we aim to generate photosynthetic phenology metrics dataset based on the GPP product derived from satellite SIF data. Our data can detect multiple growing seasons, which can be used to evaluate the photosynthesis activity of vegetation on large scales. The metrics include the start state-transition dates of photosynthesis (SOS), the end state-transition dates of photosynthesis (EOS), and the duration length of photosynthesis (LOS). With this goal, we developed a method combining a smoothing filter and change-point detection to retrieve photosynthetic phenology from a recently developed SIF-based GPP product (GOSIF-GPP: 2001–2020) with a 0.05° spatial resolution. This method enables us to acquire multiple photosynthesis activity periods of vegetation within one year. The remainder of this paper describes the GOSIF-GPP and land cover data, the adopted method for retrieving photosynthetic phenology metrics, the results and discussion of the metrics and their uncertainties, and the conclusions.

Methods

SIF-GPP Data. We used the GOSIF-GPP dataset from 2001–2020 to derive the large-scale phenology metrics in this study (<http://data.globalecology.unh.edu/>)³¹. GOSIF-GPP was estimated from the GOSIF dataset based on eight linear SIF-GPP relationships with 0.05° spatial and 8-day temporal resolutions (i.e., 46 GPP estimates per year for each 0.05° grid cell). The GOSIF dataset used discrete SIF soundings from the Orbiting Carbon Observatory-2 (OCO-2), remote sensing data from MODIS, and reanalysis data from MERRA-2 based on the machine learning method³⁸. The GOSIF-GPP showed good seasonal and spatial patterns and was highly correlated with GPP from FLUXNET³¹. Here, we identified the vegetation type of each grid cell according to the International Geosphere-Biosphere Programme (IGBP) classification from the MODIS Land Cover Type Product Version 6 (Fig. 1, 0.05° spatial resolution). The current study used six broad vegetation types (i.e., **forests**: evergreen needleleaf forests, evergreen broadleaf forests, deciduous needleleaf forests, deciduous broadleaf forests, and mixed forests; **shrublands**: closed canopy shrublands and open shrublands; **savannas**: savannas and woody savannas; **grasslands**; **wetlands**; **croplands**) in the Northern Biomes. To reduce noise generated by non-vegetation signals, we excluded the area covered with bare soil and sparse vegetation (i.e., 8-day maximum GPP over

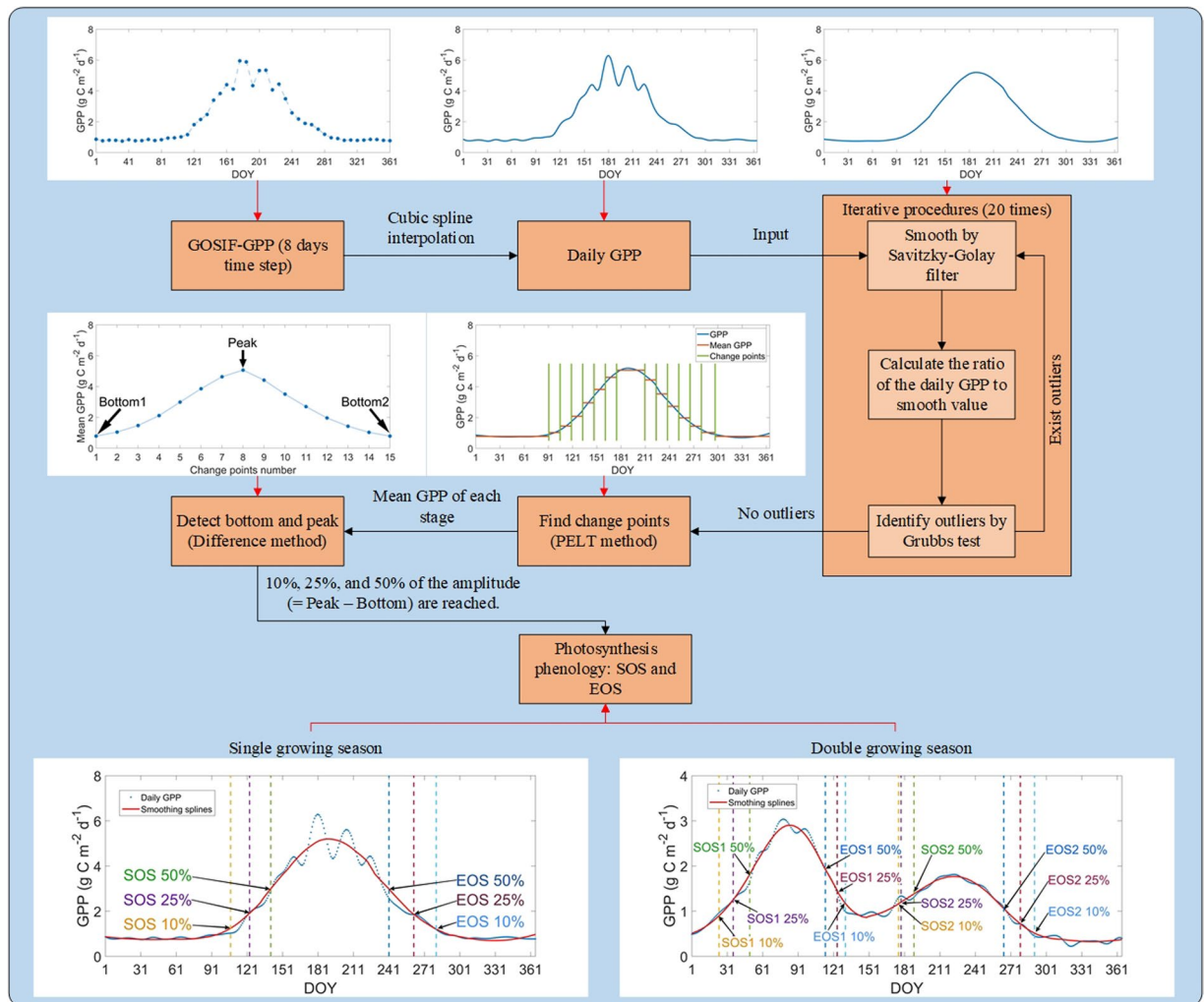


Fig. 2 Illustration of the method for identifying the transition dates of photosynthetic phenology. The method is based on three thresholds, 10%, 25%, and 50%. Bottom1: a baseline for dormancy season before the growing season; Peak: the peak value in one single cycle; Bottom2: a baseline for dormancy season after the growing season; GPP: gross primary productivity; SOS: start time of the growing season; EOS: end time of the growing season; LOS: length of the growing season; PELT: Pruned Exact Linear Time. The example of the single growing season is from one forest site (latitude: 60.0° N, longitude: 15.5° E); the example of the double growing season is from one cropland site (latitude: 36.5° N, longitude: 36.0° E).

2001–2020 was lower than 2.0 g C m⁻² day⁻¹³⁹. Since the seasonal variation of vegetation photosynthesis in the tropical region is relatively small², we focused on the area above 30° N latitude. The final dataset provided the 0.05° grid for 20 years in the six terrestrial ecosystems of the Northern Hemisphere (>30° N).

To evaluate phenology estimates based on GOSIF-GPP, we used the daily GPP data from EC flux towers across the Northern Biomes based on the Variable Ustar Threshold (VUT) mean values from the FLUXNET2015 dataset (<https://fluxnet.org/data/fluxnet2015-dataset/>)⁴⁰. The FLUXNET2015 dataset provided GPP observations based on both nighttime and daytime approaches. We used the GPP average from the two methods. We retained the EC flux sites that were relatively homogeneous because the footprint of 0.05° GOSIF product and EC tower may not exactly match³¹. In this study, one given site was considered homogeneous when the dominant land cover type in 1 × 1 0.05° pixel was similar to that of the site. We selected the flux sites having available GPP data for more than one year. The selected flux tower GPP dataset includes 49 sites with 389 site-year data (detailed information on these flux sites can be found in Table S1 and Fig. 1). Here, we chose GOSIF-GPP for consistent comparison with other GPP products because the GOSIF and GOSIF-GPP data were linearly correlated in each pixel and the extracted phenological metrics based on GOSIF-GPP were the same as those based on GOSIF. We also compared the performance of GOSIF-GPP-based phenology metrics with those based on the vegetation indices and GPP products from the MODIS datasets. For each site, we extracted and calculated three vegetation indices from the Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance dataset MCD43A4 (produced 1-day temporal and 500 m spatial resolution), including the NDVI, the EVI, and the near-infrared reflectance of vegetation (NIR_v)⁴¹; the 8-day, 500 m MODIS-GPP data⁴²; the 8-day, 0.05° GLASS-GPP data⁴³; and the 8-day, 1 km BESS-GPP data⁴⁴. The MODIS-GPP data was extracted

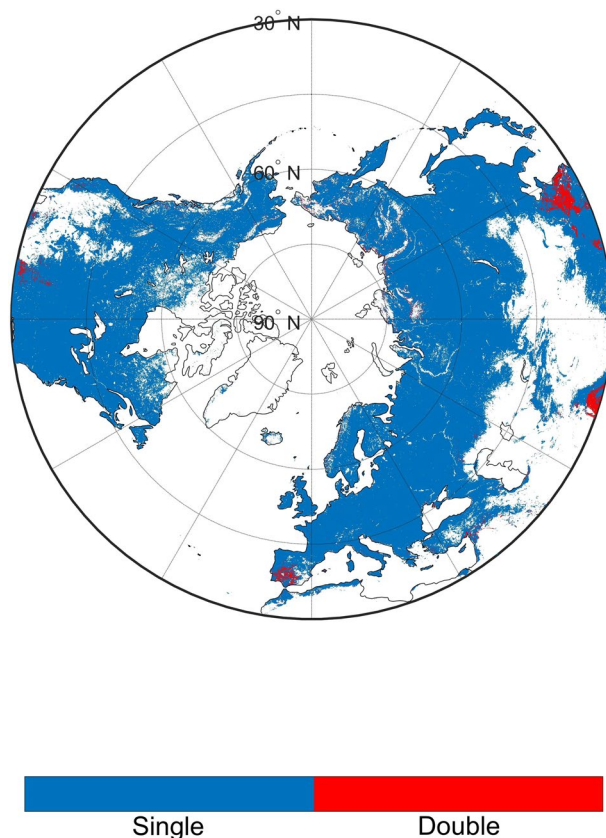


Fig. 3 The spatial distribution of the number of growing seasons in the Northern Biomes (0.05° spatial resolution). The double seasons mean there are two photosynthesis cycles in one year. We used the Pruned Exact Linear Time (PELT) method to detect the change points of each photosynthesis cycle.

from the MOD17A2H dataset, and it was generated by Moderate Resolution Imaging Spectroradiometer (MODIS) Leaf Area Index (LAI)/Fraction of Photosynthetically Active Radiation (FPAR). The GLASS-GPP data was generated by a light use efficiency (EC-LUE) model and the environmental variables (i.e., atmospheric CO₂ concentration, radiation components, and atmospheric vapor pressure deficit)⁴³. The BESS-GPP data was generated by a simplified process-based model, the Breathing Earth System Simulator (BESS), and MODIS Atmosphere and Land products⁴⁴. The period of all data was consistent with the observations of EC towers.

Photosynthetic phenology metrics. The phenology metrics in this study include SOS, EOS, and LOS. Unlike the traditional phenological events from the structural changes of leaf or flower, photosynthetic phenology is defined as the start (i.e., SOS) and end (i.e., EOS) state-transition dates of the photosynthesis cycles. These transition dates are used as the phenology metrics. One full cycle generally has five distinctive stages, including (1) photosynthesis dormancy period, a season before the growing season; (2) photosynthesis development period, a GPP rising stage; (3) photosynthesis peak period, a peak stage of GPP; (4) photosynthesis recession period, a GPP falling stage; and (5) photosynthesis dormancy period, the photosynthetically inactive stage after the growing season. Most previous studies used sigmoid-based methods (e.g., the double-logistic model) to extract the phenology, but these methods are limited to a single cycle⁶. Because some regions or ecosystems had multiple cycles in one year, we used (1) the smoothing splines to minimize the influence of outliers; (2) the change points to identify the transition dates of photosynthesis. In this study, all transition dates were extracted from the daily GPP sequence of each grid cell. Thus, we interpolated the 8-day GOSIF-GPP data to the daily scale using cubic spline interpolation before the extraction. For the cubic spline interpolation, a tridiagonal linear system (possibly with several right-hand sides) was solved for the information needed to describe the coefficients of the various cubic polynomials that made up the interpolating spline (the detailed information could be seen in the ‘spline’ method in Matlab).

We constructed an automatic method to retrieve transition dates (i.e., SOS and EOS) of photosynthetic phenology using GPP data. The algorithm of this method is outlined in the flowchart in Fig. 2. The critical basis for acquiring phenological events was the data reconstruction using smoothing methods to minimize the impact of abnormal values⁴⁵. We applied the iterative procedure to conduct the smoothing process (Fig. 2): (1) Smoothing the GPP time series by the Savitzky-Golay filter, which can reduce the influence of outliers and retain the major change characteristics of the original data sequence⁴⁶; (2) Calculating the ratio of the daily GPP value to the smooth value; (3) Identifying outliers in these ratios by using the Grubbs test and using the smooth value instead of the daily GPP value when the ratios were larger than one standard deviation below or above the mean ratio;

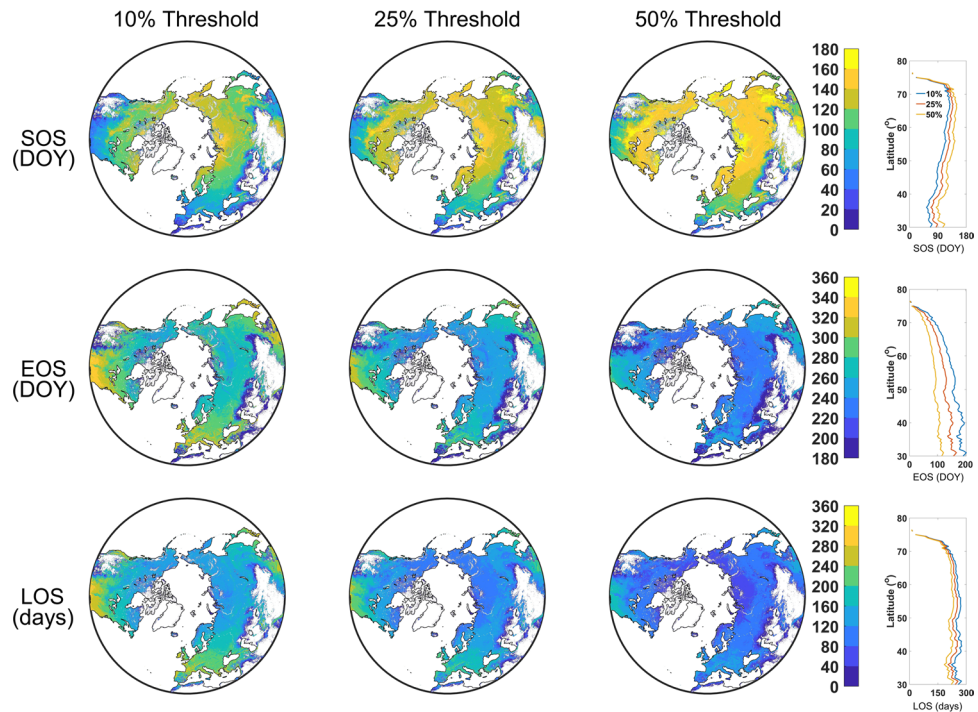


Fig. 4 The spatial distribution of the mean photosynthetic phenology metrics (first growing season) in the Northern Biomes of 2001–2020 (0.05° spatial resolution). SOS: start time of the growing season; EOS: end time of the growing season; LOS: length of the growing season; DOY: day of the year. The right parts are the latitudinal pattern.

(4) Applying the iterative procedure up to 20 times or until no outliers were detected from one iteration to the next. This procedure can largely keep the raw seasonal pattern of photosynthesis.

The potential change points in the final smoothing splines were identified with the Pruned Exact Linear Time (PELT) method. This method can accurately detect the significant change points in the data time series and does not need to preset the number of change points³⁷. For each photosynthesis cycle, we followed Richardson, *et al.*³⁷ to set the penalty factor and the minimum segment length of PELT as 0.5 and 14 days, respectively. The penalty factor was used to limit the number of returned significant changes by applying the additional penalty to each prospective changepoint. The minimum segment length regulated the minimum number of days between the changepoints. Killick, *et al.*⁴⁷ first used the PELT method, describing the calculation processes and how to find the change points in time series in detail. We calculated the mean GPP value of the adjacent change points as the potential peak and bottom baseline in one full cycle. According to the time series of mean GPP value, we used the difference method to detect the bottoms and peaks (i.e., the minimum and maximum value in each cycle). The adjacent bottoms and one peak were formed as a full cycle, and the value of these points was considered as the baselines. The difference between the baselines was used as the amplitude. Some GOSIF-GPP data affected by the weak vegetation SIF signals could have unreliable cycles. These cycles with peaks less than 0.25 of the maximum peak were excluded from the current study.

Here, the SOS and EOS dates of each cycle were determined by amplitude thresholds (i.e., the value reached 10%, 25%, and 50% of the amplitude). The amplitude was equal to the peak baseline minus the bottom baseline. Although the “true” onset of photosynthesis may correspond most closely to the 10% amplitude threshold¹¹, the most tightly-constrained transition dates (i.e., the accurate transition dates) tended to occur in the later dates of the GPP rising stage and the earlier dates of the GPP falling stage³⁷. Thus, we followed Richardson, *et al.*³⁷ to provide the SOS and EOS dates using three amplitude thresholds: 10%, 25%, and 50%. The SOS and EOS were determined when the GPP smoothing splines reached the value of amplitude thresholds, and the LOS was defined as EOS minus SOS:

$$SOS_i = t, \text{ if } GPP_S(t) = (Peak - Bottom_1) \times threshold_i \quad (1)$$

$$EOS_i = t, \text{ if } GPP_S(t) = (Peak - Bottom_2) \times threshold_i \quad (2)$$

$$LOS_i = EOS_i - SOS_i \quad (3)$$

where i is the threshold (10%, 25%, and 50%); t is the day of the year (DOY); GPP_S is the daily value of the smoothed splines; $Bottom_1$ is the baseline for dormancy season before the growing season; $Bottom_2$ is the baseline for dormancy season after growing season. Note that we retrieved the phenology of vegetation indices (i.e.,

Terrestrial ecosystems	Threshold	Mean SOS (uncertainty)	Mean EOS (uncertainty)	Mean LOS (uncertainty)
Forests	10%	108.26 (7.34)	271.29 (18.39)	163.03 (25.73)
	25%	122.12 (8.28)	255.40 (17.31)	133.28 (25.59)
	50%	138.44 (9.38)	236.24 (16.01)	97.80 (25.40)
Shrublands	10%	75.10 (5.09)	144.33 (9.78)	69.22 (14.88)
	25%	80.97 (5.49)	137.14 (9.30)	56.17 (14.79)
	50%	88.14 (5.97)	129.02 (8.75)	40.89 (14.72)
Savannas	10%	106.37 (7.21)	244.25 (16.56)	137.88 (23.77)
	25%	117.25 (7.95)	230.10 (15.60)	112.85 (23.55)
	50%	130.85 (8.87)	213.10 (14.45)	82.24 (23.32)
Grasslands	10%	56.72 (3.84)	153.48 (10.40)	96.76 (14.25)
	25%	65.21 (4.42)	140.37 (9.52)	75.16 (13.94)
	50%	74.71 (5.06)	127.55 (8.65)	52.84 (13.71)
Wetlands	10%	106.58 (7.23)	213.13 (14.45)	106.55 (21.67)
	25%	115.24 (7.81)	202.70 (13.70)	86.83 (21.51)
	50%	126.24 (8.56)	189.61 (12.85)	63.37 (21.41)
Croplands	10%	86.60 (5.87)	272.47 (18.47)	185.87 (24.34)
	25%	102.20 (6.93)	250.78 (17.00)	148.58 (23.93)
	50%	120.56 (8.17)	226.01 (15.32)	105.45 (23.49)

Table 1. The mean value and uncertainty of photosynthetic phenology metrics in the different terrestrial ecosystems.

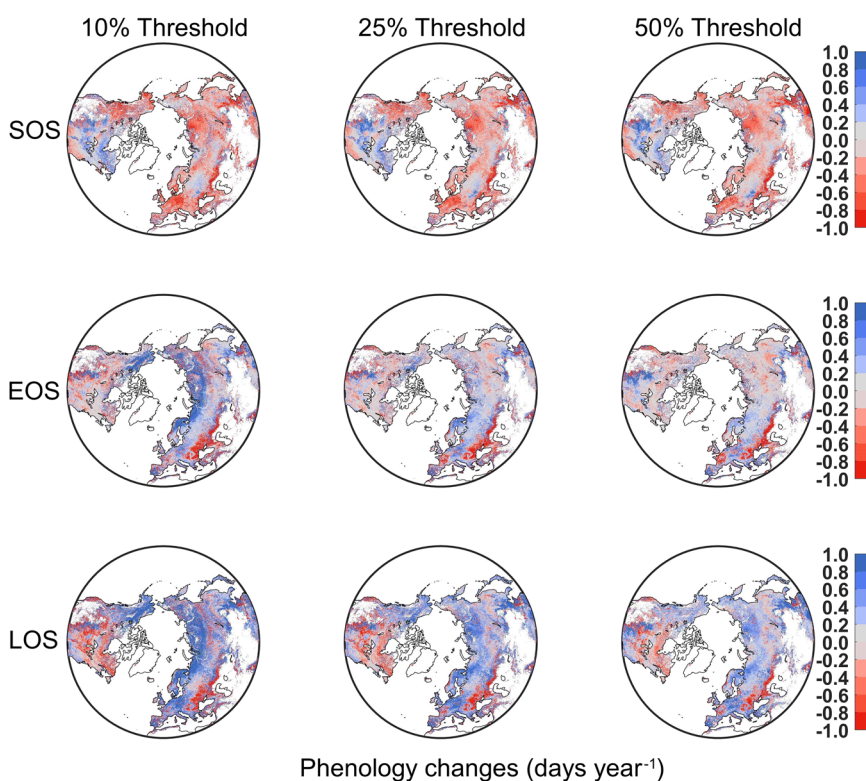


Fig. 5 Changes in photosynthetic phenology metrics in the Northern Biomes from 2001 to 2020. SOS: start time of the growing season; EOS: end time of the growing season; LOS: length of the growing season.

daily data), MODIS-GPP (i.e., interpolating the 8-day data to the daily scale), GLASS-GPP (i.e., interpolating the 8-day data to the daily scale), BESS-GPP (i.e., interpolating the 8-day data to the daily scale), and observed GPP from EC tower (i.e., EC-GPP, daily data) by using the same method.

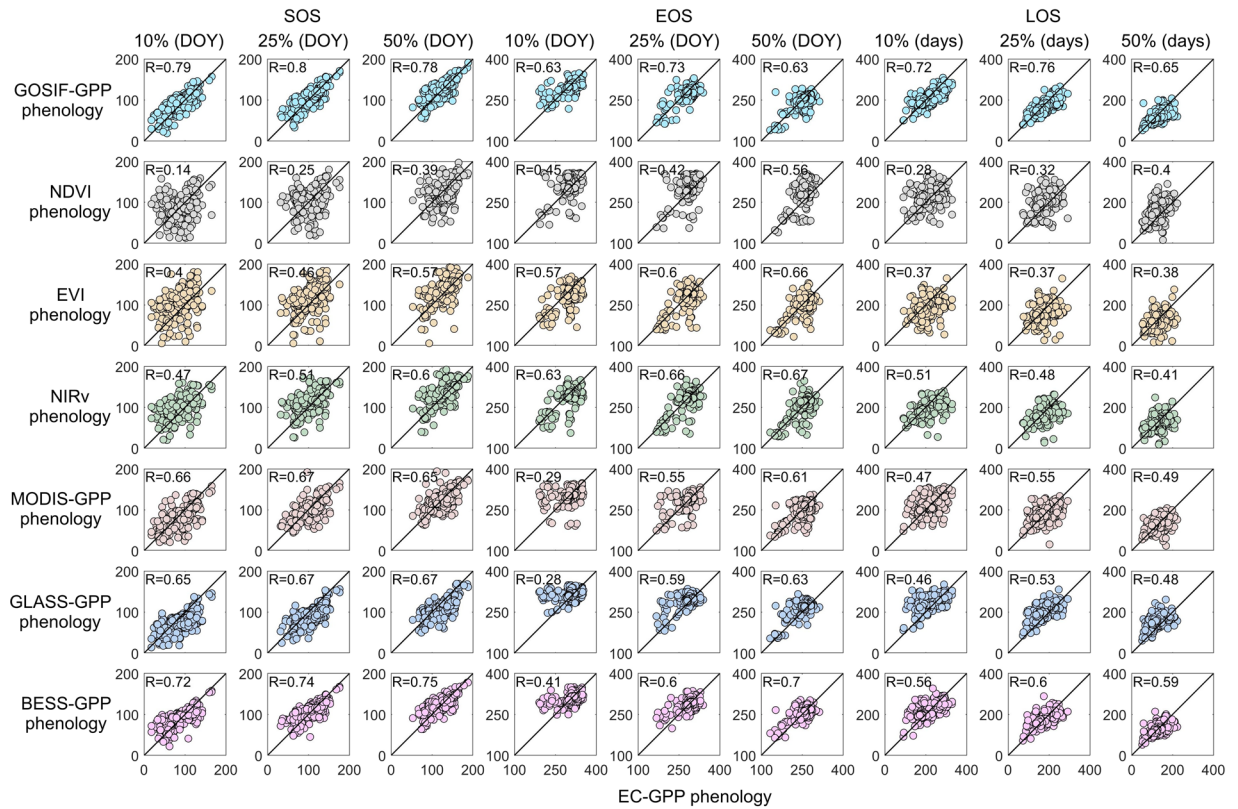


Fig. 6 The comparison of the phenology metrics retrieves from EC tower GPP (EC-GPP) and GOSIF-GPP, NDVI, EVI, NIR_v, MODIS-GPP, GLASS-GPP, and BESS-GPP. Each subplot has 389 site-year data. The significant correlations of all results are less than 0.05 ($p < 0.05$). The solid line represents a 1:1 line. SOS: start time of the growing season; EOS: end time of the growing season; LOS: length of the growing season; DOY: day of the year; R: correlation coefficient.

Data Records

During 2001–2020, our product provides annual 0.05 degrees vegetation photosynthetic phenology (i.e., start of season-SOS; end of season-EOS; length of season-LOS) in terrestrial ecosystems of the Northern Hemisphere (latitude $> 30^{\circ}\text{N}$). The entire dataset⁴⁸ is deposited at the open-access repository Figshare (<https://doi.org/10.6084/m9.figshare.17195009.v3>). This dataset is divided into single and double growing seasons. The files include two types: the first growing season in ecosystems of the single and the double growing season; the second growing season in ecosystems of the double growing season. Figure 3 showed the spatial distribution of the number of growing seasons. Most regions in the Northern Biomes had a single growing season, while some croplands had a double growing season in a given year. We showed the spatial distribution of the first growing season in Fig. 4. For different ecosystems (Table 1), grasslands showed the earliest SOS and EOS among all biomes; forests and savannas had the latest SOS; croplands and forests exhibited the latest EOS and the longest LOS, while shrublands and grasslands had the shortest LOS. Fig. S1 showed the spatial distribution of the second growing season. Figure 5 presented the linear regression analysis using the transition dates of phenology and the time series in each grid cell, and the regression coefficient was considered as the changing trend of the grid cell.

Each RAR file includes the SOS, EOS, or LOS of every year (format: *.nc*, NetCDF). The file names are structured according to the file naming scheme “<year number> <name of phenology metrics>.nc”. In each *.nc* file, including: <lat>, latitudes; <lon>, longitudes; <name of phenology metrics> <10%>, Phenology metrics with 10% threshold; <name of phenology metrics> <10%> <low>, 5th percentile phenology metrics with 10% threshold; <name of phenology metrics> <10%> <up>, 95th percentile phenology metrics with 10% threshold; <name of phenology metrics> <25%>, Phenology metrics with 25% threshold; <name of phenology metrics> <25%> <low>, 5th percentile phenology metrics with 25% threshold; <name of phenology metrics> <25%> <up>, 95th percentile phenology metrics with 25% threshold; <name of phenology metrics> <50%>, Phenology metrics with 50% threshold; <name of phenology metrics> <50%> <low>, 5th percentile phenology metrics with 50% threshold; <name of phenology metrics> <50%> <up>, 95th percentile phenology metrics with 50% threshold.

The presented dataset is Version 3. You can extract the areas from the *.nc* files if you are only interested in a specific area (area of interest).

Data source	SOS 10%	SOS 25%	SOS 50%	EOS 10%	EOS 25%	EOS 50%	LOS 10%	LOS 25%	LOS 50%
	R								
GOSIF-GPP	0.79	0.80	0.78	0.63	0.73	0.63	0.72	0.76	0.65
NDVI	0.14	0.25	0.39	0.45	0.42	0.56	0.28	0.32	0.40
EVI	0.40	0.46	0.57	0.57	0.60	0.66	0.37	0.37	0.38
NIR _v	0.47	0.51	0.60	0.63	0.66	0.67	0.51	0.48	0.41
MODIS-GPP	0.66	0.67	0.65	0.29	0.55	0.61	0.47	0.55	0.49
GLASS-GPP	0.65	0.67	0.67	0.28	0.59	0.63	0.46	0.53	0.48
BESS-GPP	0.72	0.74	0.75	0.41	0.60	0.70	0.56	0.60	0.59

Table 2. Correlation of the phenology metrics retrieved from EC tower GPP and GOSIF-GPP, NDVI, EVI, NIR_v, MODIS-GPP, GLASS-GPP, and BESS-GPP. 10%, 25%, and 50% mean the amplitude thresholds. The bold means the highest R.

Data source	SOS 10%	SOS 25%	SOS 50%	EOS 10%	EOS 25%	EOS 50%	LOS 10%	LOS 25%	LOS 50%
	RMSE (days)								
GOSIF-GPP	18.03	15.83	14.99	23.55	21.89	24.38	33.93	29.14	27.89
NDVI	36.91	32.18	26.20	34.13	39.68	41.86	53.87	53.92	52.39
EVI	36.97	31.34	24.97	31.75	29.41	26.09	58.53	49.04	39.28
NIR _v	29.86	26.00	21.74	27.26	25.67	24.95	46.11	40.03	35.12
MODIS-GPP	22.98	20.56	18.83	30.22	24.76	23.88	43.79	36.33	32.25
GLASS-GPP	25.15	23.47	22.17	29.09	23.8	24.07	46.88	41.11	39.41
BESS-GPP	20.04	17.43	15.24	25.7	21.59	19.16	38.23	32.21	26.61
Bias (days)									
GOSIF-GPP	-3.73	-2.72	-1.31	9.05	10.55	10.05	12.78	13.26	11.36
NDVI	-0.98	4.08	4.27	-10.94	-19.96	-26.9	-9.9	-24.0	-31.2
EVI	-17.84	-12.04	-8.33	11.89	5.27	0.77	29.74	17.31	9.10
NIR _v	-13.68	-9.53	-6.98	9.17	3.67	0.20	22.85	13.19	7.18
MODIS-GPP	6.37	5.47	4.0	1.60	2.30	3.27	-4.77	-3.18	-0.73
GLASS-GPP	12.67	13.47	13.53	-7.29	-7.51	-9.54	-19.9	-21.0	-23.0
BESS-GPP	-4.02	-3.39	-2.43	0.88	1.03	0.94	4.90	4.42	3.37

Table 3. Root mean square error (RMSE) and mean bias (Bias) of the phenology metrics retrieved from EC tower GPP and GOSIF-GPP, NDVI, EVI, NIR_v, MODIS-GPP, GLASS-GPP, and BESS-GPP. 10%, 25%, and 50% mean the amplitude thresholds. The bold means the lowest value.

Technical Validation

Comparison with phenology derived from vegetation indices, MODIS-GPP, GLASS-GPP, BESS-GPP, and EC tower data. We used the photosynthetic phenology metrics extracted from the daily GPP of the flux towers to examine the corresponding metrics extracted from the GOSIF-GPP product. We also use the same method to retrieve phenology from the NDVI, EVI, NIR_v, MODIS-GPP, GLASS-GPP, and BESS-GPP for the EC tower sites. The period of all comparison data was consistent with the EC data (these periods could be found in Table S1). According to the different thresholds, the metrics were divided into nine groups (SOS_{10%}, SOS_{25%}, SOS_{50%}: SOS with 10%, 25%, and 50% amplitude threshold; EOS_{10%}, EOS_{25%}, EOS_{50%}: EOS with 10%, 25%, and 50% amplitude threshold; LOS_{10%}, LOS_{25%}, LOS_{50%}: LOS with 10%, 25%, and 50% amplitude threshold) (Fig. 6, Tables 2, 3).

Overall, the phenology metrics of GOSIF-GPP showed the highest correlations with the phenology metrics of EC tower GPP, while the phenology of NDVI showed the lowest correlations. For each metric, (1) SOS, the GOSIF-GPP had the best correlation coefficient ($R = 0.78$ – 0.80 , calculated across all the site-years), the lowest root mean square error ($RMSE = 14.99$ – 18.03 days), and the lowest mean bias ($Bias$, from -3.73 to -1.31 days) in the 10%, 25%, and 50% thresholds. (2) EOS, the highest R between 10%, 25%, and 50% EOS of EC tower GPP and other data was GOSIF-GPP (0.63 – 0.73), and the lowest R was NDVI (0.42 – 0.56). (3) LOS, the highest R between 10%, 25%, and 50% LOS of EC tower GPP and other data was GOSIF-GPP (0.65 – 0.76), and the lowest R was NDVI (0.28 – 0.40). The comparisons indicated that GOSIF-GPP consistently performed better than the vegetation indices (i.e., NDVI, EVI, and NIR_v) for different photosynthesis phenology metrics and thresholds. MODIS-GPP and GLASS-GPP had larger deviations than GOSIF-GPP, highlighting the need for improvement on light use efficiency models. The BESS-GPP performs slightly worse than the GOSIF-GPP. NIR_v, the product of near-infrared reflectance and NDVI, is slightly better at capturing the phenology metrics of tower GPP than EVI and NDVI.

The derived phenology of GOSIF-GPP and EC tower GPP (i.e., EC-GPP) showed a close correspondence across the 389 site-years. The best performance of the different thresholds in SOS, EOS, and LOS was 25%

($R = 0.80, 0.73$, and 0.76 ; $RMSE = 15.83, 21.89$, and 29.14 days; $Bias = -2.72, 10.55$, and 13.26 days, respectively), and the threshold of 10% had relatively low performance in SOS and EOS ($R = 0.79$ and 0.63 ; $RMSE = 18.03$ and 23.55 days; $Bias = -3.73$ and 9.05 days, respectively) and 50% had relatively low performance in LOS ($R = 0.65$; $RMSE = 27.89$ days; $Bias = 11.36$ days). Our results showed that our method better captured the SOS than the EOS. For example, Table S3 presented that the SOS extracted from evergreen forests performed better than the EOS. Table S2 showed the performance of the GOSIF-GPP phenology in the different terrestrial ecosystems. The GOSIF-GPP phenology had a high performance for grasslands, wetlands, and croplands ($R > 0.84$) and a moderate performance for forests and shrublands ($R = 0.77$ and 0.51 , respectively).

Uncertainties of photosynthetic phenology metrics. The uncertainties in the estimates of phenology metrics mainly arise from the gridded SIF-based GPP estimates, such as using the limited explanatory variables to acquire the gridded SIF estimates (i.e., GOSIF) and the relationship between the SIF and GPP. Previously, Li and Xiao³¹ had assessed the quality of the underlying SIF and GPP data. In this study, we used the Monte Carlo Bootstrapping method⁴⁹ to estimate the related uncertainties of the GOSIF-GPP phenology. Bootstrapping provides valuable information about uncertainties without making assumptions about the underlying data distributions^{48,50}. We used bootstrapping for each year of the individual grid cell to replace the transition dates with 100 times random uniform sampling⁶. Bootstrapping was a statistical procedure that resampled a single dataset to create many simulated samples. In this study, each grid had 100 transition dates created by the bootstrapping method. The 5th and 95th percentiles of the 100 bootstrapped data were considered as the confidence interval of the mean estimated from the original transition dates.

The uncertainty used in this study was defined as the 5th and 95th percentiles of the 100 Monte Carlo bootstrapping samples ranging from a few days to several weeks (Table 1). The uncertainty was the lowest for SOS and the highest for LOS; EOS had intermediate uncertainty. The highest uncertainty in LOS may be because of the compounding effect of SOS and EOS⁶. Generally, metrics of grasslands had the lowest uncertainty: SOS uncertainty ranged from 3.8 to 5.1 days, EOS uncertainty ranged from 8.6 to 10.4 days, and LOS uncertainty ranged from 13.7 to 14.2 days; forests have the largest uncertainty, with SOS uncertainty ranging from 7.3 to 9.4 days, EOS uncertainty from 16.0 to 18.4 days, and LOS uncertainty from 25.4 to 25.7 days.

Code availability

The code⁵¹ of processes is available at Zenodo dataset (<https://zenodo.org/record/7266229#.Y19tWHZBwuU>). The code needs to be run in the version of Matlab after 2018.

Received: 4 November 2022; Accepted: 10 May 2023;

Published online: 19 May 2023

References

- Keenan, T. F., Richardson, A. D. & Hufkens, K. On quantifying the apparent temperature sensitivity of plant phenology. *New Phytol.* **225**, 1033–1040, <https://doi.org/10.1111/nph.16114> (2020).
- Piao, S. *et al.* Plant phenology and global climate change: Current progresses and challenges. *Glob Chang Biol* **25**, 1922–1940, <https://doi.org/10.1111/gcb.14619> (2019).
- Richardson, A. D. *et al.* Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric For Meteorol* **169**, 156–173, <https://doi.org/10.1016/j.agrformet.2012.09.012> (2013).
- Wang, X. *et al.* No trends in spring and autumn phenology during the global warming hiatus. *Nat. Commun.* **10**, 2389, <https://doi.org/10.1038/s41467-019-10235-8> (2019).
- Seyednasrollah, B. *et al.* Seasonal variation in the canopy color of temperate evergreen conifer forests. *New Phytol.* **229**, 2586–2600, <https://doi.org/10.1111/nph.17046> (2021).
- Yang, L. & Noormets, A. Standardized flux seasonality metrics: a companion dataset for FLUXNET annual product. *Earth Syst. Sci. Data* **13**, 1461–1475 (2021).
- Joiner, J. *et al.* The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange. *Remote Sens Environ* **152**, 375–391, <https://doi.org/10.1016/j.rse.2014.06.022> (2014).
- Li, X. & Xiao, J. Global climatic controls on interannual variability of ecosystem productivity: Similarities and differences inferred from solar-induced chlorophyll fluorescence and enhanced vegetation index. *Agric For Meteorol* **288–289**, 108018, <https://doi.org/10.1016/j.agrformet.2020.108018> (2020).
- Dai, J. *et al.* Divergent changes of the elevational synchronicity in vegetation spring phenology in North China from 2001 to 2017 in connection with variations in chilling. *Int J Climatol* **41**, 6109–6121, <https://doi.org/10.1002/joc.7170> (2021).
- Huang, X., Xiao, J. & Ma, M. Evaluating the Performance of Satellite-Derived Vegetation Indices for Estimating Gross Primary Productivity Using FLUXNET Observations across the Globe. *Remote Sens.* **11**, 1823 (2019).
- Wu, C. *et al.* Land surface phenology derived from normalized difference vegetation index (NDVI) at global FLUXNET sites. *Agric For Meteorol* **233**, 171–182, <https://doi.org/10.1016/j.agrformet.2016.11.193> (2017).
- Smith, W. K. *et al.* Chlorophyll Fluorescence Better Captures Seasonal and Interannual Gross Primary Productivity Dynamics Across Dryland Ecosystems of Southwestern North America. *Geophys. Res. Lett.* **45**, 748–757, <https://doi.org/10.1002/2017GL075922> (2018).
- Walther, S. *et al.* Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests. *Glob Chang Biol* **22**, 2979–2996, <https://doi.org/10.1111/gcb.13200> (2016).
- Fang, J., Lutz, J. A., Wang, L., Shugart, H. H. & Yan, X. Using climate-driven leaf phenology and growth to improve predictions of gross primary productivity in North American forests. *Glob Chang Biol* **26**, 6974–6988, <https://doi.org/10.1111/gcb.15349> (2020).
- Gu, L. *et al.* in *Phenology of Ecosystem Processes: Applications in Global Change Research* (ed Noormets, A.) 35–58 (Springer New York, 2009).
- Dronova, I. & Taddeo, S. Remote sensing of phenology: Towards the comprehensive indicators of plant community dynamics from species to regional scales. *J. Ecol.* **110**, 1460–1484, <https://doi.org/10.1111/1365-2745.13897> (2022).
- Xiao, J. *et al.* Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sens Environ* **233**, 111383, <https://doi.org/10.1016/j.rse.2019.111383> (2019).
- Baldocchi, D. *et al.* FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bull Am Meteorol Soc* **82**, 2415–2434, [10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2) (2001).

19. Myers-Smith, I. H. *et al.* Complexity revealed in the greening of the Arctic. *Nature Climate Change* **10**, 106–117, <https://doi.org/10.1038/s41558-019-0688-1> (2020).
20. Previdi, M., Smith, K. L. & Polvani, L. M. Arctic amplification of climate change: a review of underlying mechanisms. *Environ. Res. Lett.* **16**, 093003, <https://doi.org/10.1088/1748-9326/ac1c29> (2021).
21. Ueyama, M. *et al.* Growing season and spatial variations of carbon fluxes of Arctic and boreal ecosystems in Alaska (USA). *Ecol Appl* **23**, 1798–1816, <https://doi.org/10.1890/11-0875.1> (2013).
22. Sjöström, M. *et al.* Evaluation of MODIS gross primary productivity for Africa using eddy covariance data. *Remote Sens Environ* **131**, 275–286, <https://doi.org/10.1016/j.rse.2012.12.023> (2013).
23. Frankenberg, C. *et al.* Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sens Environ* **147**, 1–12, <https://doi.org/10.1016/j.rse.2014.02.007> (2014).
24. Köhler, P. *et al.* Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophys. Res. Lett.* **45**, 10,456–410,463, <https://doi.org/10.1029/2018GL079031> (2018).
25. Li, X. *et al.* Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations. *Glob Chang Biol* **24**, 3990–4008, <https://doi.org/10.1111/gcb.14297> (2018).
26. Kim, J. *et al.* Solar-induced chlorophyll fluorescence is non-linearly related to canopy photosynthesis in a temperate evergreen needleleaf forest during the fall transition. *Remote Sens Environ* **258**, 112362, <https://doi.org/10.1016/j.rse.2021.112362> (2021).
27. Meroni, M. *et al.* Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote Sens Environ* **113**, 2037–2051, <https://doi.org/10.1016/j.rse.2009.05.003> (2009).
28. Verma, M. *et al.* Effect of environmental conditions on the relationship between solar-induced fluorescence and gross primary productivity at an OzFlux grassland site. *J. Geophys. Res. Biogeosci.* **122**, 716–733, <https://doi.org/10.1002/2016JG003580> (2017).
29. Wood, J. D. *et al.* Multiscale analyses of solar-induced fluorescence and gross primary production. *Geophys. Res. Lett.* **44**, 533–541, <https://doi.org/10.1002/2016GL070775> (2017).
30. Zhang, Z., Chen, J. M., Guanter, L., He, L. & Zhang, Y. From Canopy-Leaving to Total Canopy Far-Red Fluorescence Emission for Remote Sensing of Photosynthesis: First Results From TROPOMI. *Geophys. Res. Lett.* **46**, 12030–12040, <https://doi.org/10.1029/2019GL084832> (2019).
31. Li, X. & Xiao, J. A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data. *Remote Sens.* **11**, 517 (2019).
32. Zhang, Z. *et al.* Reduction of structural impacts and distinction of photosynthetic pathways in a global estimation of GPP from space-borne solar-induced chlorophyll fluorescence. *Remote Sens Environ* **240**, 111722, <https://doi.org/10.1016/j.rse.2020.111722> (2020).
33. Bertani, G., Wagner, F. H., Anderson, L. O. & Aragão, L. E. O. C. Chlorophyll Fluorescence Data Reveals Climate-Related Photosynthesis Seasonality in Amazonian Forests. *Remote Sens.* **9**, 1275 (2017).
34. Garrity, S. R. *et al.* A comparison of multiple phenology data sources for estimating seasonal transitions in deciduous forest carbon exchange. *Agric For Meteorol* **151**, 1741–1752, <https://doi.org/10.1016/j.agrformet.2011.07.008> (2011).
35. Wang, X. *et al.* No Consistent Evidence for Advancing or Delaying Trends in Spring Phenology on the Tibetan Plateau. *J. Geophys. Res. Biogeosci.* **122**, 3288–3305, <https://doi.org/10.1002/2017JG003949> (2017).
36. Klosterman, S. T. *et al.* 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> (2014).
37. Richardson, A. D. *et al.* Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Sci. Data* **5**, 1–24 (2018).
38. Li, X. & Xiao, J. Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: A global, fine-resolution dataset of gross primary production derived from OCO-2. *Remote Sens.* **11**, 2563 (2019).
39. Liu, Q. *et al.* Delayed autumn phenology in the Northern Hemisphere is related to change in both climate and spring phenology. *Glob Chang Biol* **22**, 3702–3711, <https://doi.org/10.1111/gcb.13311> (2016).
40. Pastorello, G. *et al.* The FLUXNET2015 dataset and the ONEFLUX processing pipeline for eddy covariance data. *Sci. Data* **7**, 225, <https://doi.org/10.1038/s41597-020-0534-3> (2020).
41. Badgley, G., Field Christopher, B. & Berry Joseph, A. Canopy near-infrared reflectance and terrestrial photosynthesis. *Sci. Adv.* **3**, e1602244, <https://doi.org/10.1126/sciadv.1602244>.
42. Zhao, M., Heinsch, F. A., Nemani, R. R. & Running, S. W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens Environ* **95**, 164–176, <https://doi.org/10.1016/j.rse.2004.12.011> (2005).
43. Zheng, Y. *et al.* Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017. *Earth Syst. Sci. Data* **12**, 2725–2746, <https://doi.org/10.5194/essd-12-2725-2020> (2020).
44. Jiang, C. & Ryu, Y. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). *Remote Sens Environ* **186**, 528–547, <https://doi.org/10.1016/j.rse.2016.08.030> (2016).
45. Li, X. *et al.* A dataset of 30 m annual vegetation phenology indicators (1985–2015) in urban areas of the conterminous United States. *Earth Syst. Sci. Data* **11**, 881–894 (2019).
46. Chen, J. *et al.* A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens Environ* **91**, 332–344, <https://doi.org/10.1016/j.rse.2004.03.014> (2004).
47. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *J Am Stat Assoc* **107**, 1590–1598 (2012).
48. Fang, J. *et al.* Vegetation photosynthetic phenology metrics in northern terrestrial ecosystems: a dataset derived from a gross primary productivity product based on solar-induced chlorophyll fluorescence. *figshare* <https://doi.org/10.6084/m9.figshare.17195009.v3> (2021).
49. Efron, B. in *Breakthroughs in Statistics: Methodology and Distribution* (eds Kotz, S. & Johnson, N. L.) 569–593 (Springer New York, 1992).
50. Elmore, A. J., Guinn, S. M., Minsley, B. J. & Richardson, A. D. Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. *Glob Chang Biol* **18**, 656–674, <https://doi.org/10.1111/j.1365-2486.2011.02521.x> (2012).
51. Fang, J. Find phenology dates of GPP (GPP_phenology). *Zenodo* <https://doi.org/10.5281/zenodo.7266229> (2022).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (32101349, 32171599). This study also was supported by the National Key R&D Program of China (2019YFA0606904) and the Key Program of the National Natural Science Foundation of China (32130069). J.X. was supported by the University of New Hampshire.

Author contributions

J.F. planned the project. F.L. and J.F. conducted the processing. J.F. and X.L. contributed to data collection. X.L., J.X., B.L. and J.F. contributed to data analysis and interpretation of the results. X.L. and J.F. took the lead in writing the manuscript. J.F. feedback and approval from co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02224-w>.

Correspondence and requests for materials should be addressed to F.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023