



# A vegetation phenology dataset developed by integrating multiple sources using the reliability ensemble averaging method

Yishuo Cui<sup>1</sup>, Shouzhi Chen<sup>1</sup>, Yufeng Gong<sup>1</sup>, Mingwei Li<sup>1</sup>, Zitong Jia<sup>1</sup>, Yuyu Zhou<sup>2,★</sup>, and Yongshuo H. Fu<sup>1,3,‡</sup>

<sup>1</sup>College of Water Sciences, Beijing Normal University, Beijing 100875, China

<sup>2</sup>Department of Geography and Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong SAR, China

<sup>3</sup>Plants and Ecosystems, Department of Biology, University of Antwerp, Antwerp, Belgium

<sup>‡</sup>First correspondence author

<sup>★</sup>Second correspondence author

**Correspondence:** Yuyu Zhou (yuyuzhou@hku.hk) and Yongshuo H. Fu (yfu@bnu.edu.cn)

Received: 7 June 2024 – Discussion started: 15 August 2024

Revised: 5 May 2025 – Accepted: 16 May 2025 – Published: 21 August 2025

**Abstract.** Global change has substantially shifted vegetation phenology, with important implications in the carbon and water cycles of terrestrial ecosystems. Various vegetation phenology datasets have been developed using remote sensing data. However, the significant uncertainties in these datasets limit our understanding of ecosystem dynamics in terms of phenology. It is therefore crucial to generate a reliable large-scale vegetation phenology dataset, by fusing various existing vegetation phenology datasets, to provide a comprehensive and accurate estimation of vegetation phenology with a fine spatiotemporal resolution. In this study, we merged four widely used vegetation phenology datasets to generate a new dataset using the reliability ensemble averaging (REA) fusion method. The new dataset has a spatial resolution of 0.05° and covers the period from 1982 to 2020, with geographic coverage extending above 30° N in the Northern Hemisphere. The evaluation using ground-based phenocam data from 280 sites indicated that the accuracy of the newly merged dataset was substantially improved compared to the four original datasets. The start and end of the growing season (SOS and EOS) in the newly merged dataset showed the highest correlation with ground-based phenocam observations, compared to the original datasets (0.84 and 0.71, respectively) and accuracy in terms of the root mean square error (RMSE) between phenocam data and merged datasets (12 and 17 d, respectively). Using the new dataset, we found that the SOS is occurring approximately 0.19 d earlier per year ( $p < 0.01$ ), while the end of the growing season is occurring 0.18 d later per year ( $p < 0.01$ ) over the period 1982–2020 across regions north of 30° N. This dataset offers a unique and novel source of vegetation phenology data for global ecology studies.

This study uses multiple phenology datasets, including the MCD12Q2 dataset (Friedl et al., 2022, <https://doi.org/10.5067/MODIS/MCD12Q2.061>), the VIP dataset (Didan and Barreto, 2016, [https://doi.org/10.5067/MEaSUREs/VIP/VIPPHEN\\_NDVI.004](https://doi.org/10.5067/MEaSUREs/VIP/VIPPHEN_NDVI.004)), the GIMMS NDVI3g dataset (Wang et al., 2019, [http://data.globalecology.unh.edu/data/GIMMS\\_NDVI3g\\_Phenology/](http://data.globalecology.unh.edu/data/GIMMS_NDVI3g_Phenology/)), the GIMMS NDVI4g dataset (Chen and Fu, 2024, <https://doi.org/10.5281/zenodo.11136967>), the PhenoCam dataset (Richardson et al., 2018, <https://doi.org/10.3334/ORNLDAAAC/1674>), the Internet Nature Information System of Japan (Ministry of the Environment of Japan, 2024, <http://www.sizenken.biodic.go.jp>), the Phenological Eyes Network (PEN, 2024, <http://www.pheno-eye.org>), and the MCD12Q1 land use dataset (Friedl and Sulla-Menashe, 2022, <https://doi.org/10.5067/MODIS/MCD12Q1.061>). The REA phenology dataset developed in this study is available at <https://doi.org/10.5281/zenodo.15165681> (Cui and Fu, 2024).

## 1 Introduction

Global change has notably altered the timing of vegetation phenology (Ettinger et al., 2020; Zhang et al., 2022), leading to important impacts on the carbon and water cycles of terrestrial ecosystems (Peñuelas et al., 2009; Piao et al., 2019; Richardson et al., 2012; Zhou, 2022). Various vegetation phenology datasets using remote sensing data have been produced, but inconsistencies and uncertainties arise when comparing these datasets with ground-based phenological observations and there is large variation in spatiotemporal resolution (Peng et al., 2017). Therefore, there is an urgent need to develop a highly reliable vegetation phenology product to improve our understanding of vegetation phenology dynamics, and to facilitate subsequent research on terrestrial ecosystem responses to climate change.

Ground-based phenological records are commonly used in vegetation phenology studies (Fu et al., 2014; Geng et al., 2020; Sparks and Carey, 1995; Zhou et al., 2020). For example, phenocams, a near-surface remote sensing tool, have been operational for more than 20 years (Richardson et al., 2018a). Although ground-based observations provide high accuracy in terms of phenology dynamics, they are limited to certain locations, resulting in sparse spatial coverage. In contrast, phenology datasets based on remote sensing data can cover large areas, providing comprehensive and continuous monitoring of vegetation phenology across landscapes, regions, or even continents. Additionally, remote sensing datasets and phenocam data are processed using standardized methods that ensure consistency and comparability across different locations and periods. However, phenology datasets based on remote sensing data do have certain limitations. Owing to differences in revisit cycles among satellites, together with sensor characteristics, sun-sensor geometry, and atmospheric conditions during imaging, substantial bias exists among the derived phenology datasets. For example, differences of  $> 50$  d in the start of the growing season (SOS) have been reported among different phenology datasets based on remote sensing data (Peng et al., 2017; Zhou et al., 2020). Additionally, substantial variations in the trends of vegetation phenology exist. For example, a recent study reported that the SOS was delayed by  $0.17 \text{ d yr}^{-1}$  when based on the Global Inventory Modeling and Mapping Studies-3rd Generation (GIMMS 3g) dataset, whereas the SOS was advanced by  $0.58 \text{ d yr}^{-1}$  when based on the Moderate Resolution Imaging Spectroradiometer (MODIS) dataset in the Northern Hemisphere during 2000–2015 (Zhang et al., 2020). Previous studies found that different vegetation phenology datasets have advantages and disadvantages in different regions and over different periods (Fensholt and Proud, 2012; Zhang et al., 2020). The MODIS phenology product for the United States shows a stronger correlation with ground observations compared to the AVHRR (Advanced

Very High-Resolution Radiometer) phenology product (Peng et al., 2017), while the VIPPHEN (Making Earth System Data Records for Use in Research Environments Vegetation Index and Phenology) data have fewer missing values than the MODIS phenology product. For estimates obtained using different extraction methods, such as varying algorithms or approaches for extracting the SOS and EOS (end of the growing season) from the same satellite data, the discrepancies can exceed one month (White et al., 2009). Additionally, the NDVI (normalized difference vegetation index) is a commonly used remote sensing indicator for assessing vegetation cover and health. It is derived from reflectance data in the red and near-infrared bands and is closely related to aboveground net primary productivity (Myneni et al., 1995; Pettorelli et al., 2005). The NDVI threshold required for phenology extraction varies across different biomes due to differences in vegetation types, growth patterns, and environmental conditions, which affect how NDVI values correspond to phenological events such as the SOS and EOS (Reed et al., 1994). Because it is difficult to determine the optimal dataset from the various phenology datasets, producing a merged dataset using a method that selects the most suitable dataset for different times and locations from all input datasets is essential for providing a comprehensive and accurate estimation of vegetation phenology with high spatiotemporal resolution.

Remote sensing vegetation phenology typically reflects key transition dates in the vegetation growth cycle, such as the SOS and EOS, using various vegetation indices such as the NDVI and EVI (enhanced vegetation index) to assess vegetation conditions (Cong et al., 2012; Piao et al., 2019). Ideally, phenological dates extracted using different methods (thresholds, derivatives, smoothing functions, and fitted models) should accurately capture changes in actual physiological conditions (De Beurs and Henebry, 2010). However, in existing phenology datasets, there is often no “best” definition of these transition dates (White et al., 2009). The effectiveness of different extraction methods can vary across regions and periods, and may not always perfectly reflect true vegetation conditions (Cong et al., 2013; De Beurs and Henebry, 2010). For instance, in high-latitude areas, meaningful observations are relatively sparse. If the smoothing method removes too much information, it may reduce the ability to extract phenological signals that accurately reflect surface dynamics (Wang et al., 2015). Therefore, integrating multiple datasets based on their reliability, rather than relying on a single dataset or using a simple averaging method, is a more robust approach. Our study stresses the value of integrating different datasets, and that phenology results can be improved by applying the REA method, which assigns weights based on dataset reliability. This method can reduce uncertainties and provide a more accurate representation of phenological dynamics across different spatial and temporal scales.

Data fusion methods generally include unmixing-based, weight-function-based, and Bayesian-based approaches (Gevaert and García-Haro, 2015; Piao et al., 2019). In vegetation phenology studies, fusion methods based on raw remote sensing data, such as the Spatial and Temporal Adaptive Reflectance Fusion Model (Gao et al., 2006) and the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (Zhu et al., 2010), are often influenced by vegetation types, growth conditions, and methodological assumptions (Sisheber et al., 2022). These methods are typically applied to specific regions, and their performance can be affected by nonlinear spectral mixing, where the reflectance of vegetation endmembers (the pure spectral signatures of distinct land cover types) changes nonlinearly, and the spectral response of a single pixel is no longer a simple linear combination of the endmember spectra (Ma et al., 2015). The nonlinear combination of the ground feature can degrade the accuracy of vegetation phenology extraction. Unlike these approaches, the reliability ensemble averaging (REA) method is not based on the assumption of linear reflectance changes. Instead, it directly merges annual phenology products based on their reliability. Compared to traditional data fusion methods, the REA method shows clear advantages in simplicity and computational efficiency (Lu et al., 2021), while explicitly accounting for dataset reliability, in contrast to simple averaging methods that assume equal reliability across datasets. The simple averaging method treats all datasets equally by calculating the mean value across different vegetation phenology datasets, despite their uncertainties varying across time and space (Lu et al., 2021; Wang et al., 2019), which leads to potential inaccuracies in the result. The REA method considers the temporal consistency of vegetation phenology data and uses the voting principle, whereby the final REA result is generated by assigning different weights to the input data sources based on their agreement (Giorgi and Mearns, 2002). This provides convergence while preserving spatial differences, making it suitable for multi-source data fusion.

In this study, we merged four widely used vegetation phenology datasets to generate a new dataset using the REA fusion method. The spatial resolution of the new dataset is  $0.05^\circ$ , with a temporal scale spanning 1982–2020, and it covers regions north of the  $30^\circ\text{N}$  latitude. The new dataset was evaluated using data from the ground-based phenocam dataset from 280 sites over the period 2000–2018, which provided 1410 site–year combinations. We further explored the phenological trends in the spring and autumn vegetation phenology using the merged dataset. The new vegetation phenology dataset could be used in further studies on the impact of energy and carbon–water cycles within terrestrial ecosystems, together with analyses of their responses and feedbacks to global climate change (Piao et al., 2009, 2019; Tang et al., 2016).

## 2 Data and method

### 2.1 Phenology dataset

Four satellite-based vegetation phenology products were used to create a merged dataset, and the ground-based phenocam dataset was used for validation. The four satellite-based vegetation phenology products included (1) the MCD12Q2 (Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Dynamics data product) phenology dataset, which was extracted from the MODIS Land Cover Dynamics Version 6.1 derived by Friedl et al. (2022); (2) the VIP (Making Earth System Data Records for Use in Research Environments Vegetation Index and Phenology) dataset derived by Didan and Barreto (2016); (3) the GIM\_3g (GIMMS (Global Inventory Modeling and Mapping Studies) NDVI 3rd Generation) based phenology dataset derived by Wang et al. (2019); and (4) the GIM\_4g (GIMMS (Global Inventory Modeling and Mapping Studies) NDVI 4th Generation) based phenology dataset derived by Chen and Fu (2024). These phenological data products were obtained from open sources and used to merge a new set of phenological products. The time span and the spatial resolution of each vegetation phenology dataset are listed in Table 1. The merged data used in this paper were clipped into regions above  $30^\circ$  in the Northern Hemisphere to ensure that the region was covered by the four datasets.

#### 2.1.1 MCD12Q2 phenology dataset

The MCD12Q2 product was derived using data from the MODIS sensor on board the Terra and Aqua satellites. The MCD12Q2 land cover dynamic product v6.1 provides a global surface phenology dataset with a 500 m spatial resolution for the period 2001–2020. The vegetation phenology data were extracted from the Nadir Bidirectional Adjusted Reflectance 2-band Enhanced Vegetation Index (EVI2) using the threshold method (Gray et al., 2019). The threshold method defines the growing state of the vegetation as the time when the vegetation index reaches a certain percentage of the annual amplitude and reflects a specific vegetation physiological growth stage (the schematic diagram is shown in Fig. S1 in the Supplement). The amplitude is calculated as the difference between the maximum and minimum values of the EVI2 time series during the entire growing season. The MCD12Q2 phenology dataset includes greenup and dormancy (equivalent to SOS and EOS in this study, respectively). Greenup (dormancy) is defined as the date when the EVI2 time series first (last) crosses 15 % of the segment EVI2 amplitude (Gray et al., 2019). The time series data were fitted by a penalized cubic smoothing spline. This dataset can be found at <https://doi.org/10.5067/MODIS/MCD12Q2.061> (Friedl et al., 2022).

**Table 1.** List of data sources.

Name	Abbreviations	Sensor	Spatial resolution	Time span	Reference
MODIS MCD12Q2	MCD12Q2	MODIS	500 m	2001–2020	Friedl et al. (2022)
MEaSURES VIPPHEN	VIP	AVHRR& MODIS	0.05°	1982–2015	Didan and Barreto (2016)
GIMMS NDVI3g	GIM_3g	AVHRR	1/12°	1999–2014	Wang et al. (2019)
GIMMS NDVI4g	GIM_4g	AVHRR	1/12°	1982–2020	Chen and Fu (2024)

2.1.2 VIP phenology dataset

The VIP phenology dataset was generated using data from the NASA Making Earth System Data Records for Use in Research Environments (MEaSURES) and the AVHRR over the period 1981–1999, together with MODIS/Terra MOD09 surface reflectance data over the period 2000–2014 (Didan et al., 2018). The VIP dataset includes the SOS and EOS, which were also extracted using the threshold method. The filtering method, which uses confidence intervals and an operational continuity algorithm, was applied to reconstruct the time series curves. The start (end) of the season is defined using the modified Half-Max method (White et al., 2009) as the date when the NDVI time series first (last) crosses 35 % of the growing-season NDVI amplitude. This dataset is organized in a geographic gridded format with a spatial resolution of 0.05°. This dataset can be found at [https://doi.org/10.5067/MEaSURES/VIP/VIPPHEN\\_NDVI.004](https://doi.org/10.5067/MEaSURES/VIP/VIPPHEN_NDVI.004) (Didan and Barreto, 2016).

2.1.3 GIM\_3g phenology dataset

The GIMMS NDVI 3g-based phenology dataset (GIM\_3g) has a spatial resolution of 1/12° and covers the period 1999–2014 (Wang et al., 2019). A double logistic function was applied to fit the NDVI curve and the threshold method was used to extract phenological dates, including the SOS and EOS. This product provides phenology data for the Northern Hemisphere, and it uses the date when the NDVI first (last) crosses 20 % of the segment NDVI amplitude as the SOS (EOS). This dataset can be accessed at [http://data.globalecology.unh.edu/data/GIMMS\\_NDVI3g\\_Phenology/](http://data.globalecology.unh.edu/data/GIMMS_NDVI3g_Phenology/) (last access: 13 May 2024) (Wang et al., 2019).

2.1.4 GIM\_4g phenology dataset

The GIM\_4g dataset, based on the GIMMS NDVI 4g dataset acquired by the AVHRR sensors, has a spatial resolution of 1/12° and a temporal scale spanning 1982–2020.

Two steps were adopted in the process to extract phenological dates. First, the NDVI time series data were fitted and smoothed using five fitting methods: the HANTS-Maximum, Spline-Midpoint, Gaussian-Midpoint, Timesat-SG, and Polyfit-Maximum methods. Second, the threshold method was used to extract phenological dates, using the date when the NDVI first (last) crosses 20 % (50 %) of the segment NDVI amplitude as the SOS (EOS) (Chen et al., 2024; Chen and Fu, 2024). The average spring (SOS) and autumn (EOS) phenological dates were produced by averaging the results obtained from the five fitting methods. The GIM\_4g phenology dataset is available at <https://doi.org/10.5281/zenodo.11136967> (Chen and Fu, 2024).

2.1.5 Camera-based phenology dataset

A ground-based phenocam dataset, with phenological dates extracted from camera-derived images with high spatial resolution and reliable accuracy, was used to validate the merged dataset. The phenocam dataset comprised three datasets. The first dataset, the PhenoCam Dataset v2.0 (Richardson et al., 2018b; Seyednasrollah et al., 2019a, b), included data derived from conventional visible-wavelength automated digital camera imagery through the PhenoCam Network (Richardson et al., 2018a) over the period 2000–2018 and across 393 sites in various ecosystems. For detailed information, please refer to <https://doi.org/10.3334/ORNLDAAAC/1674> (Richardson et al., 2018) and <https://phenocam.nau.edu/webcam/> (last access: 21 May 2024). It comprised all typical vegetation types including deciduous broadleaf, deciduous needleleaf, evergreen broadleaf, evergreen needleleaf, grassland, mixed vegetation, shrubland, tundra, and wetland ecosystems, mainly in regions of Europe and North America (Moon et al., 2021; Ruan et al., 2023). A spline interpolation method was applied to the PhenoCam data to extract transition dates for each region of interest (ROI) using the green chromatic coordinate (GCC), a measure of greenness intensity derived from digital imagery (Sonnentag et al.,



2012) in the PhenoCam Dataset v2.0. We used the date when the GCC first (last) crosses 25 % of the GCC amplitude as the SOS and EOS. The second phenocam dataset is from the Japan Internet Nature Information System digital camera data (<http://www.sizenken.biodic.go.jp/>, last access: 12 May 2024) acquired over the period 2002–2009. Ide and Oguma (2010) provided greenup dates for two phenocam sites with the ROI defined at the species-level scale. The vegetation types included in their data comprised wetland and mixed deciduous forest. The date of greenup each year was estimated as the day of year (DOY) corresponding to the maximum rate of increase in the 2G-RBi index. This is calculated as the maximum of the second derivative of the GCC time series (the maximum of the second derivative of GCC). The third dataset consisted of phenology data for deciduous broadleaf forests in Japan (Inoue et al., 2014), supported by the Phenological Eyes Network (<http://www.pheno-eye.org/>, last access: 13 May 2024), which is a network of ground-based observatories for long-term automatic observation of vegetation dynamics established in 2003 (Nasahara and Nagai, 2015). The SOS and EOS was defined as the first day when 20 % of leaves had flushed and the first day when 80 % of leaves had fallen in the given ROI, respectively. For this study, we excluded 26 sites that only provided one type of transition date (either SOS or EOS) and removed 90 sites where none of the 4 remote sensing datasets provided valid phenology estimates. These excluded sites were primarily located in cropland-dominated areas or regions with sparse vegetation, where the low spatial resolution of remote sensing data limited the reliable detection of phenological transitions. We then selected phenocam data from 280 sites over the period 2000–2018, resulting in 1410 site–year combinations.

### 2.1.6 Land cover dataset

To avoid the impact of human activities and non-vegetated areas on data quality, areas of cropland, cropland/natural vegetation mosaics, permanent snow and ice, barren land, and water bodies were removed based on a land cover dataset obtained by a supervised classification of MODIS reflectance data (Sulla-Menashe and Friedl, 2018). The land cover data generated based on the Annual International Geosphere–Biosphere Programme classification schemes are available from <https://doi.org/10.5067/MODIS/MCD12Q1.061> (Friedl and Sulla-Menashe, 2022).

## 2.2 Ensemble method for estimating phenological dates

### 2.2.1 Reliability ensemble averaging method

The weighting method was applied to obtain more accurate SOS and EOS dates from the four vegetation phenology datasets. The weight assigned to each product was based on the interannual variability of each phenology dataset, as well as the degree of consistency and offset among the four

datasets. Importantly, these weights can vary over time to reflect changes in dataset reliability and performance (Giorgi and Mearns, 2002). The consistency is measured as the difference between each input dataset and the mean value of the four datasets, and the offset is measured as the difference between the REA result and each input dataset. These values are iteratively calculated during the process of determining the final weight coefficients. There are discrepancies in the spatial coverage among the four phenology datasets, and missing data in specific regions for some of the datasets. The ensemble method can fill in missing data accurately, thereby producing a phenology dataset with high accuracy and spatially continuous coverage. Furthermore, the process of merging the phenology datasets does not depend on simple averaging; instead, it is based on the uncertainty (calculated using the merged result and the differences between the REA result and the remote sensing phenology datasets) among the products, which produces data that are more reliable than those obtained using the simple averaging method, and can circumvent the effects of outliers (Giorgi and Mearns, 2002).

The REA method, based on the “voting principle” (where the REA result is generated by assigning different weights to the input data sources), produces data that align with the majority of input phenology products at the pixel level. This approach assumes that most data values at each pixel are accurate, while outliers are down-weighted or excluded. It provides a dataset with high reliability by relying on the temporal consistency of each pixel among the input products, and by minimizing the influence of outliers during the merging process (Giorgi and Mearns, 2002). The REA method has been applied to generate datasets for multiple elemental fields, e.g., temperature, evapotranspiration, and precipitation (Giorgi and Mearns, 2002; Lu et al., 2021; Xu et al., 2010). In this study, the REA method was used to integrate both the SOS and the EOS from the four phenology datasets.

The REA method gives different weights to the various datasets involved in the process of data merging and then obtains the desired result using the following function:

$$\Delta \tilde{\text{Phe}} = \tilde{A}(\Delta \text{Phe}) = \frac{\sum_i R_i \Delta \text{Phe}_i}{\sum_i R_i} \quad (1)$$

where  $\Delta \tilde{\text{Phe}}$  represents the phenology result,  $\Delta \text{Phe}_i$  represents the different datasets involved in the process,  $\tilde{A}$  denotes the REA process, and  $R_i$  represents the model reliability factor, which is defined as follows:

$$R_i = \left[ (R_{B,i})^m \times (R_{D,i})^n \right]^{\left[ \frac{1}{m \times n} \right]} \\ = \left\{ \left[ \frac{\epsilon_{\text{Phe}}}{\text{abs}(B_{\text{Phe},i})} \right]^m \left[ \frac{\epsilon_{\text{Phe}}}{\text{abs}(D_{\text{Phe},i})} \right]^n \right\}^{\left[ \frac{1}{m \times n} \right]} \quad (2)$$

where  $R_{B,i}$  measures the bias of the data compared with that of the average data (the higher the bias, the lower the reliability of the dataset), and  $R_{D,i}$  represents the convergence criterion of the data (the larger the distance between the dataset

and the newly generated REA data, the poorer the convergence; several iterations are required to reach convergence). The values of  $R_{B,i}$  and  $R_{D,i}$  will be set to 1 when  $B_{Phe,i}$  and  $D_{Phe,i}$  are less than  $\epsilon_{Phe}$ , which means the deviation of the dataset is within the limit of natural variation.

$$B_{Phe,i} = \Delta Phe_i - \overline{Phe} \quad (3)$$

$$D_{Phe,i} = \Delta Phe_i - \Delta Phe \quad (4)$$

$$\epsilon_{Phe} = \max(\text{MA}(D_{Phe})) - \min(\text{MA}(D_{Phe})). \quad (5)$$

Eq. (3) explains the derivation of  $B_{Phe,i}$ : it is defined by the difference between the input dataset and the mean value of the four datasets. Eq. (4) explains the arithmetic process of  $D_{Phe,i}$ , which is measured by the difference between the REA result and each input dataset. In Eq. (5),  $\epsilon_{Phe}$  is measured by the natural variability in phenology, which is calculated by estimating the difference between the maximum and minimum values of the multiyear moving averages following the linear detrending of the observed long-term series data, and works with  $B_{Phe,i}$  and  $D_{Phe,i}$  jointly to assign weights to each dataset. Natural variability changes from region to region: in Eqs. (1) and (6),  $\epsilon_{Phe}$  cancels out under the condition that both  $B_{Phe,i}$  and  $D_{Phe,i}$  are greater than  $\epsilon_{Phe}$ . This is based on the assumption that more stringent criteria for consistency and deviation are required to increase reliability in regions characterized by lower natural variability.

$$\begin{aligned} \delta_{Phe} &= \left[ \tilde{A} \left( \Delta Phe_i - \Delta \tilde{Phe} \right)^2 \right]^{\frac{1}{2}} \\ &= \left[ \frac{\sum_i R_i \left( \Delta Phe_i - \Delta \tilde{Phe} \right)^2}{\sum_i R_i} \right]^{\frac{1}{2}} \end{aligned} \quad (6)$$

$$\Delta Phe_+ = \Delta \tilde{Phe} + \tilde{\delta}_{\Delta Phe}, \quad (7a)$$

$$\Delta Phe_- = \Delta \tilde{Phe} - \tilde{\delta}_{\Delta Phe}. \quad (7b)$$

In Eq. (6),  $\delta_{Phe}$  is the uncertainty range calculated using  $R_i$  and the difference between the REA result and the datasets (a higher value of  $\delta_{Phe}$  means larger differences between the REA result and the original phenology datasets). The upper and lower uncertainty limits are measured by  $\Delta \tilde{Phe}$  and  $\tilde{\delta}_{\Delta Phe}$ , respectively, in Eqs. (7a) and (7b).

If an individual data point shows significant discrepancies compared to others, potentially caused by improper extraction methods in that region, the  $B_{Phe,i}$  and  $D_{Phe,i}$  will extract this variance and incorporate it, along with the natural variability  $\epsilon_{Phe}$  of the region, into the weight distribution process. If the natural variability of that region is low, then the weight is assigned to a smaller value, and if the natural variability of the region is large, the weight is assigned by both the natural variability and the deviations.

To evaluate the robustness of the REA method, we used a sensitivity analysis to confirm its reliability, considering

the influence of different time spans and dataset combination choices on the fusion results. In the sensitivity analysis of the time span, we performed fusion experiments using two subsets (2001–2005 and 2006–2010) and compared them to the full 2001–2010 period, and the differences in the fusion results were analyzed. In the sensitivity analysis of the selection of different dataset combinations on the fusion results, we selected three dataset combinations for the fusion experiment and analyzed the influence of data selection on the fusion results. Each group included two datasets selected from the four used in the full REA fusion. These combinations were chosen to assess how different data sources influence the final REA results.

## 2.2.2 Evaluation criteria

In this study, the metrics of the root mean square error (RMSE), bias, correlation coefficient ( $r$ ), unbiased RMSE (Ubrmse), and coefficient of variation (CV) were used for data evaluation:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{Phe}_i - \text{Ref}_i)^2}{n}}, \quad (8)$$

$$\text{bias} = \frac{\sum_{i=1}^n (\text{Phe}_i - \text{Ref}_i)}{n}, \quad (9)$$

$$r = \frac{\sum_{i=1}^n (\text{Phe}_i - \overline{\text{Phe}}) (\text{Ref}_i - \overline{\text{Ref}})}{\sqrt{\sum_{i=1}^n (\text{Phe}_i - \overline{\text{Phe}})^2} \sqrt{\sum_{i=1}^n (\text{Ref}_i - \overline{\text{Ref}})^2}}, \quad (10)$$

$$\text{Ubrmse} = \sqrt{\text{RMSE}^2 - \text{bias}^2}, \quad (11)$$

$$\text{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{Phe}_i - \overline{\text{Phe}})^2}, \quad (12)$$

$$\text{CV} = \frac{\sigma_{\text{Phe}}}{\overline{\text{Phe}}} \quad (13)$$

where  $n$  represents the number of site years,  $\text{Phe}_i$  represents the corresponding vegetation phenological indicator (SOS and EOS) at a given point,  $\text{Ref}_i$  represents data from a phenology camera,  $\sigma_{\text{Phe}}$  represents the standard deviation of  $\text{Phe}_i$ , and  $\overline{\text{Phe}}$  and  $\overline{\text{Ref}}$  represent the average of  $\text{Phe}_i$  and  $\text{Ref}_i$ , respectively.

The RMSE is calculated as the square root of the average of the squares of the residuals, which penalizes larger errors than smaller ones and provides an estimate of the magnitude of errors between the remote sensing estimated value and phenocam datasets. Bias is the average difference between the remote sensing estimated value and the phenocam value, which helps in understanding whether the estimated value is higher or lower than the phenocam value. The correlation coefficient measures the linear relationship between two variables. The Ubrmse measures the deviation between two variables without systematic errors. The standard deviation quantifies the variation of the dataset, which measures the deviation between the data and the mean value.

### 2.2.3 Mann–Kendall trend test

The Mann–Kendall trend test is a nonparametric trend test method, which has the characteristics of not being limited by a specific distribution and a small number of outliers, and can be used to detect the trend of time series data (Kendall, 1975). The Mann–Kendall test was applied to detect trends in the SOS and EOS dates during 1982–2020 in the merged dataset, as well as trends in the growing-season greenness across different phenology datasets. The SOS/EOS trends refer to the temporal change in the timing of phenological events and the greenness trends refer to interannual variations in vegetation greenness during the growing seasons. The basic Mann–Kendall test formulas are as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_j - X_i), \quad (14)$$

$$Z_c = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & S < 0 \end{cases} \quad (15)$$

where  $X_i$  and  $X_j$  are the phenological parameter values of the  $i$ th year and the  $j$ th year of the pixel, respectively,  $n$  is the length of the time series,  $\text{sgn}$  is the sign function, and  $S$  is the test statistic. The null hypothesis  $H_0$  is as follows: the time series data are  $n$  independent samples with identically distributed random variables.  $H_1$ : for any  $i, j \leq n$ , and  $i \neq j$ , the distribution of  $X_i, X_k$  is different. If  $|Z| \geq Z_{1-\frac{\alpha}{2}}$ , the time series is considered to have a statistically significant change; otherwise, any change is considered not statistically significant. When  $Z > 0$ , the time series has an upward trend; when  $Z < 0$ , it has a downward trend.

### 2.2.4 Growing-season greenness

Greenness is a widely used indicator of vegetation growth, typically represented by NDVI (Myneni, 1997). The growing-season greenness (GSG) is calculated as the mean NDVI value within the growing season in this study, defined by the period between the SOS and EOS:

$$\text{GSG} = \text{mean}(\text{NDVI}[\text{Date}_{\text{SOS}}, \text{Date}_{\text{EOS}}]) \quad (16)$$

$\text{Date}_{\text{SOS}}$  is the DOY value of the SOS data, and  $\text{Date}_{\text{EOS}}$  is the DOY value of the EOS value. Mean denotes calculating the mean NDVI value in the corresponding date range.

## 3 Results

### 3.1 Differences in the vegetation phenological dates among the four datasets

Fig. 1 illustrates the spatial distribution of the multiyear mean dates for both the SOS and the EOS above 30° N for each of the four datasets. The mean SOS values for the

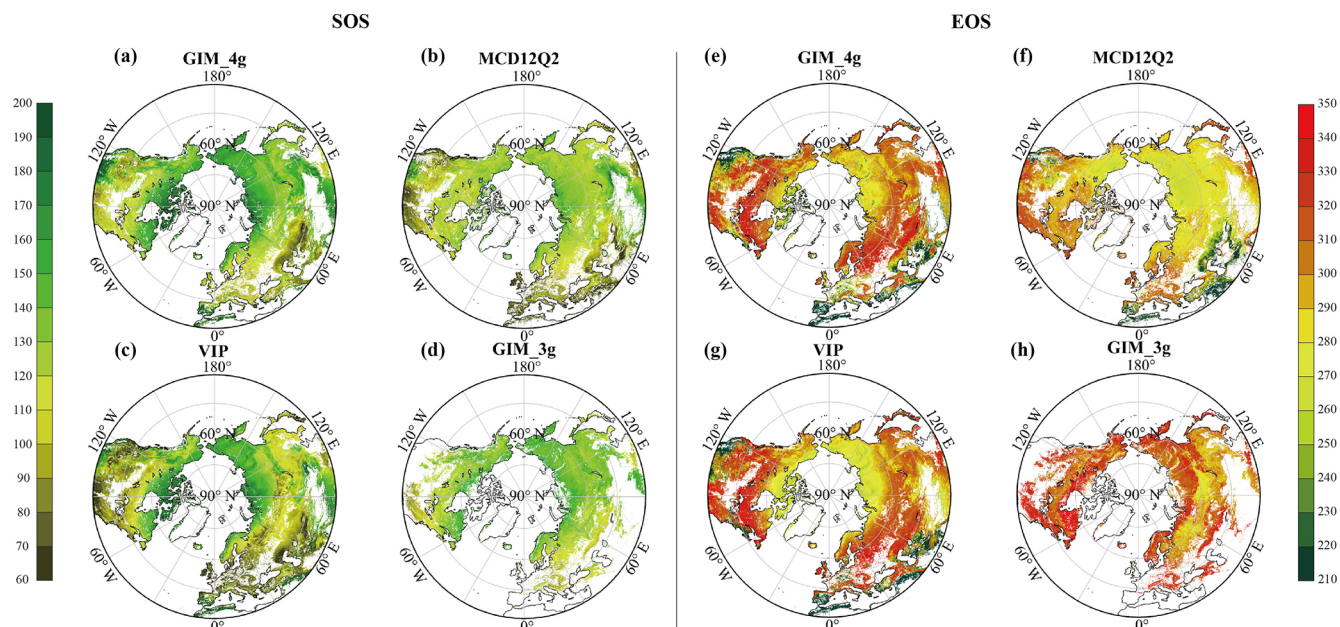
MCD12Q2, VIP, GIM\_3g, and GIM\_4g datasets are DOY 120 (SD = 32 d), 125 (SD = 43 d), 132 (SD = 17 d), and 139 (SD = 32 d), respectively. Discrepancies among the datasets are particularly notable in southwestern North America, North Africa, the Qinghai–Tibet Plateau, and Mongolia. Compared with the SOS, the EOS exhibits greater variability, and the mean EOS values for the MCD12Q2, VIP, GIM\_3g, and GIM\_4g datasets are DOY 281 (SD = 37 d), 290 (SD = 44 d), 315 (SD = 19 d), and 287 (SD = 53 d), respectively. Among the four datasets, the spatial distributions of the GIM\_4g and VIP datasets are the most similar. In comparison with these two datasets, the MCD12Q2 dataset displays earlier EOS values in northern Europe, Central Asia, North America, and the 45–60° N latitudinal belt over Central Asia. Given the substantial differences among these datasets, it is imperative to integrate these datasets into a merged dataset with higher accuracy.

### 3.2 Variation of weights and contributions of the four datasets to the merged phenology dataset

The weight of each dataset, as determined by the REA method, varies largely among years and specific locations. The left panels of Fig. 2 illustrate the mean weight for each dataset in each year over the period 1982–2020, with the upper and lower sections representing the SOS and the EOS, respectively. For the SOS, the overall weight of the VIP dataset during 1982–1998 surpasses that of the GIM\_4g dataset. The GIM\_3g dataset is dominant during 1999–2014, with weights exceeding 65 %. In 2015, the weighting of the MCD12Q2 dataset is highest, at approximately 45 %, with the weights of the other two datasets broadly similar. During 2016–2020, the weights of the MCD12Q2 and GIM\_4g datasets are 61 % and 39 %, respectively. The combinations of data sources for the EOS data are similar to those for the SOS data. Specifically, during 1982–1998, the weight of the VIP dataset is approximately 65 %, with the GIM\_4g dataset accounting for the remaining 35 %. For 1999–2000, the weighting of the GIM\_3g dataset is approximately 10 %, whereas that of the VIP dataset is the highest (approximately 55 %). Throughout the period 2001–2014, the weighting of the VIP dataset is greatest (> 45 %), whereas that of the GIM\_3g dataset is low (< 10 %); the weighting of the GIM\_4g and MCD12Q2 datasets each account for over 20 %. During 2016–2020, the weights of the GIM\_4g and MCD12Q2 datasets are broadly equal, albeit with the weighting of the GIM\_4g dataset slightly exceeding that of the MCD12Q2 dataset.

The latitudinal distribution of the mean weighting of the datasets for the SOS and the EOS is shown in Figs. 2b and 2d, respectively. For the SOS data, the zonal distribution of the GIM\_4g, VIP, and MCD12Q2 datasets is reasonably stable within 30–75° N. The weight of the GIM\_3g dataset is notably higher between 50 and 70° N, primarily because of its spatial distribution, and it shows notable fluctuations in high-





**Figure 1.** Spatial distribution of the multiyear mean SOS and EOS dates from each phenology dataset: (a–d) multiyear mean SOS dates and (e–h) multiyear mean EOS dates derived from the GIM\_4g, MCD12Q2, VIP, and GIM\_3g datasets, respectively.

latitude areas. In contrast, the weighting of the EOS datasets exhibits relatively smooth changes within 30–75° N. There are marked fluctuations in the weighting of the GIM\_4g and VIP datasets in high-latitude areas above 75° N. The weight of the GIM\_4g dataset between 30 and 75° N fluctuates before stabilizing smoothly. Conversely, the weight of the VIP dataset increases with the latitude, displaying a trend opposite to that of the GIM\_4g dataset. Additionally, the weighting of both the MCD12Q2 and the GIM\_3g datasets initially increases and then decreases with the increasing latitude.

Fig. 3 shows the spatial distribution of the mean contribution of the four datasets to the merged SOS and EOS results, calculated as the average weight for each pixel over the time span for the corresponding dataset. For the SOS data, the GIM\_3g dataset exhibits the greatest contribution, followed by similar contributions from the GIM\_4g and VIP datasets; the MCD12Q2 dataset has the smallest contribution. The MCD12Q2 dataset has a greater contribution in high-latitude areas near the Arctic Circle, but makes a smaller contribution in most other regions. The VIP dataset generally has a greater contribution than that of the MCD12Q2 dataset, with values ranging between 0 and 0.5 in 90 % of areas. The overall contribution of the GIM\_3g dataset is reasonably uniform, averaging approximately 0.37. For the EOS data, the VIP dataset has the greatest contribution, followed by the GIM\_4g dataset; the GIM\_3g dataset has the smallest contribution. The contribution of the MCD12Q2 dataset remains relatively small, primarily distributed between 0 and 0.5. The VIP dataset has a positive correlation with the latitude, with approximately 4.7 % of areas of weights exceeding 0.8 in

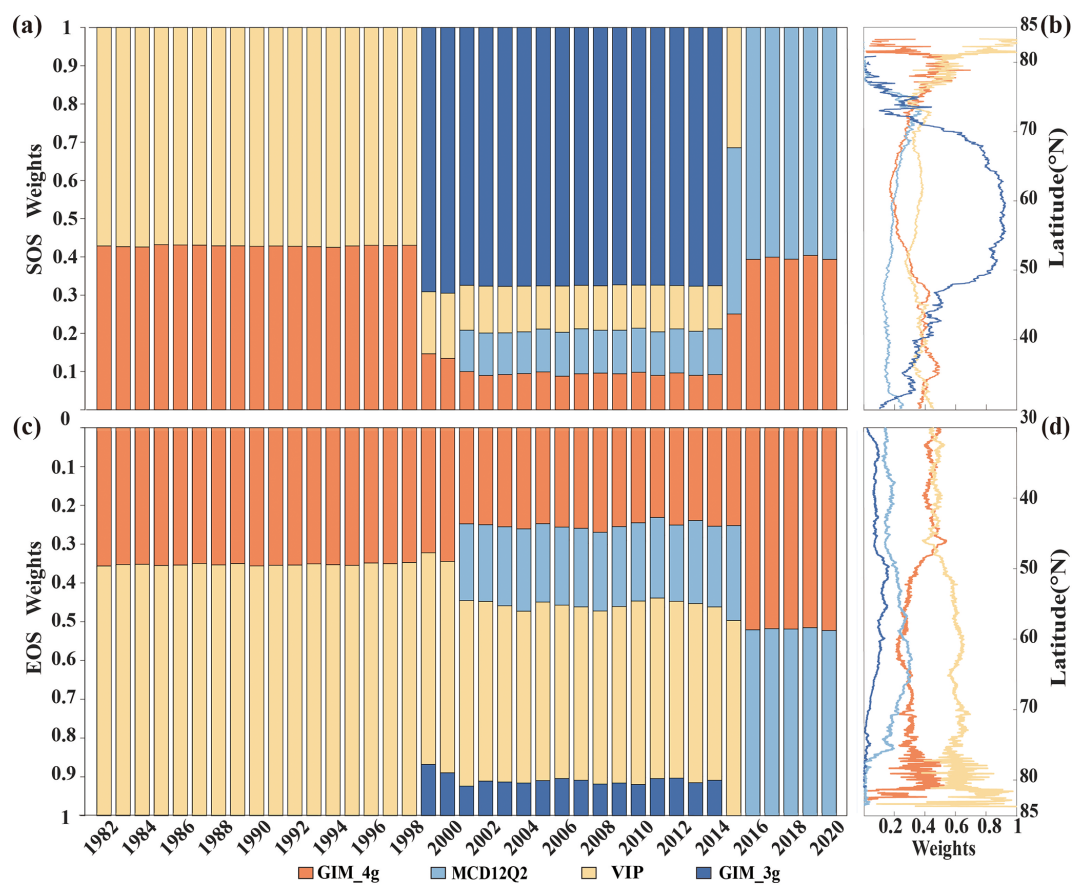
Central Asia and parts of East Asia, whereas the contribution of the GIM\_3g dataset remains lower across the entire region.

### 3.3 Merged phenology dataset using the REA method

Fig. 4 displays the merged mean SOS and EOS dates for the period 1982–2020. For the SOS, a general pattern of increase with the latitude is evident, albeit with a later occurrence of the SOS in southwestern North America and on the Qinghai–Tibet Plateau. The highest proportion of the SOS falls within DOY 120–150 (38.2 %), followed by DOY 90–120 (23.1 %), DOY 150–180 (21.23 %), and DOY 60–90 (11.1 %). Only a small proportion (4.6 %) of areas are experiencing the SOS later than DOY 180. The mean SOS obtained using the REA method is DOY 129 (SD = 28 d). It demonstrates an overall increase in the EOS with the latitude, with fewer trends observed in high-latitude areas above 60° N and the eastern parts of North America. The distribution of the EOS appears more uniform after the merging. Unlike the SOS dates, which exhibit greater variability, the EOS dates are more consistent, predominantly occurring within DOY 270–330 (80.0 %). The mean EOS is DOY 284 (SD = 20 d). The interannual variability in most regions for both the SOS and the EOS data is minimal; however, notable variations are observed in areas such as southwestern North America, Spain, Portugal, North Africa, West Asia, and Mongolia, consistent with the earlier analysis of data sources (Fu et al., 2014; Liu et al., 2016; Piao et al., 2006, 2015).

The mean uncertainty range of the merged SOS and EOS dates, calculated using Eq. (6), is presented in Fig. 4. This



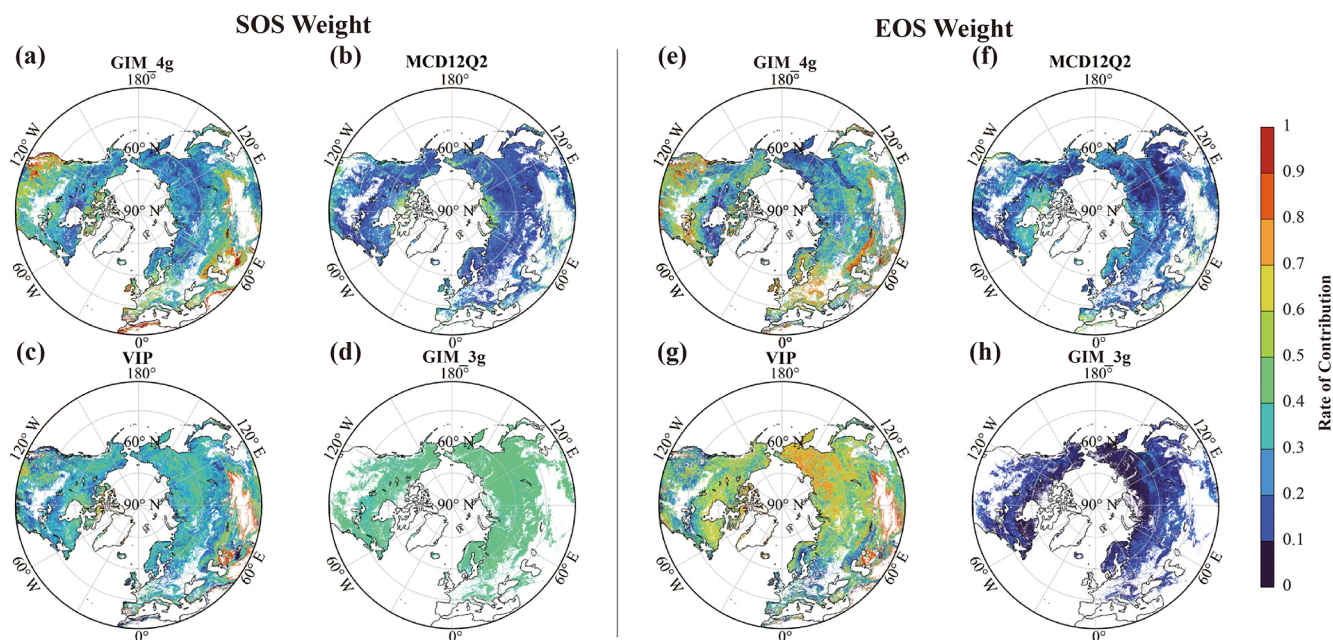


**Figure 2.** (a, c) The weights of the four phenology datasets during 1982–2020 and (b, d) latitudinal differences for (a, b) the SOS and (c, d) the EOS. For panels (b) and (d), the latitudinal weights are the mean values of each dataset over their respective time spans. The four datasets comprise the GIM\_4g, MCD12Q2, VIP, and GIM\_3g datasets (for the full names, see Table 1).

range was determined using the REA method over the period from 1982 to 2020. The mean uncertainty range of the SOS (EOS) dates is below 10 d in more than 96 % (94 %) of regions, with less than 4 % (5 %) of regions exhibiting a mean uncertainty range exceeding 10 or 15 d (Figs. 4b and 4e). The mean uncertainty range of SOS dates shows a negative correlation with the latitude, whereas this trend is not evident in the EOS dates. In Figs. 4c and 4f, the CV of the uncertainty in the SOS and EOS dates from 1982 to 2020 is analyzed. More than 56 % (73 %) of regions exhibit a CV below 1, 31 % (18 %) of regions have a CV between 1 and 1.5, and only 13 % (8 %) of regions show a CV higher than 1.5. No evident correlation is observed between the CV and latitude changes.

The phenocam dataset was used to evaluate each of the four vegetation datasets and the merged dataset. The verification results of the SOS and EOS data indicate that the merged data produced using the REA method has the best performance (Fig. 5). Specifically, the RMSE for the SOS and the EOS is 12 and 17 d, respectively. The correlation between the SOS and phenocam results is notably high, at 0.84; for the EOS, it is 0.71. An evaluation of the four satellite-based

SOS products shows that the GIM\_3g dataset has the highest correlation coefficient and the lowest RMSE among the four datasets. However, it has more missing values spatially and a shorter time span, leading to fewer points for verification. The MCD12Q2 dataset has a correlation coefficient of 0.65 and an RMSE of 20 d, but its wider spatial coverage provides more points for verification. The GIM\_4g dataset has a lower correlation with the phenocam dataset owing to outliers, resulting in an RMSE of 29 d. Compared with the phenocam dataset, the VIP dataset tends to estimate earlier SOS dates than the other datasets, especially in areas where the SOS occurs early in the season (DOY 100–140), leading to a larger RMSE. And comparing with the simple average, the REA-based SOS shows better performance in RMSE (REA and average, 12 and 21 d, respectively (in Fig. S4)), CORR (correlation coefficient) (REA and average, 0.84 and 0.65, respectively), bias (REA and average, −1.5 and −9.7 d, respectively), and UBRMSE (REA and average, 12 and 18 d, respectively). The REA-based SOS dataset outperforms in terms of all indicators, with the lowest RMSE, UBRMSE, and standard deviation, together with the highest correlation



**Figure 3.** Spatial distribution of the mean contribution of the four datasets to the merged SOS and EOS results. The mean SOS (a–d) and EOS (e–h) weights are derived from the GIM\_4g, MCD12Q2, VIP, and GIM\_3g datasets.

and lowest absolute bias, thereby demonstrating high consistency with the phenocam dataset.

In the evaluation of the EOS, the MCD12Q2 dataset has the best results among the four datasets, and except for the REA result, it has the highest correlation coefficient and the lowest RMSE. The GIM\_4g dataset shows good performance but tends to overestimate the EOS, with predicted dates occurring later than observed, resulting in an RMSE of 43 d. Both the VIP and the GIM\_3g datasets also overestimate the EOS due to their spatial and temporal distributions, with RMSEs of 46 and 35 d, respectively. The REA-based EOS also shows better performance compared to the simple average of the original datasets in RMSE (REA and average, 17 and 32 d, respectively), CORR (REA and average, 0.71 and 0.45, respectively), bias (REA and average, 1.0 and 8.0 d, respectively), and UBRMSE (REA and average, 17 and 31 d, respectively). It is evident from Fig. 5 that the REA dataset demonstrates the highest accuracy and best consistency with the phenocam dataset, outperforming the four other datasets in terms of all indicators, with the lowest RMSE, UBRMSE, and standard deviation, together with the highest correlation and lowest absolute bias.

We selected a long-term PhenoCam site (Morganmonroe) from the PhenoCam Network to evaluate the merged dataset across years at a single location. We have chosen a US PhenoCam site characterized by deciduous broadleaved forest, with data from 2010–2020 for both the SOS and EOS. As illustrated in the time series plot in Fig. S5, the consistency between the REA results and PhenoCam data for both the SOS and EOS is the highest when compared to the original

four individual datasets. Additionally, most vegetation phenology products demonstrate higher consistency with phenocam data for the SOS compared to the EOS.

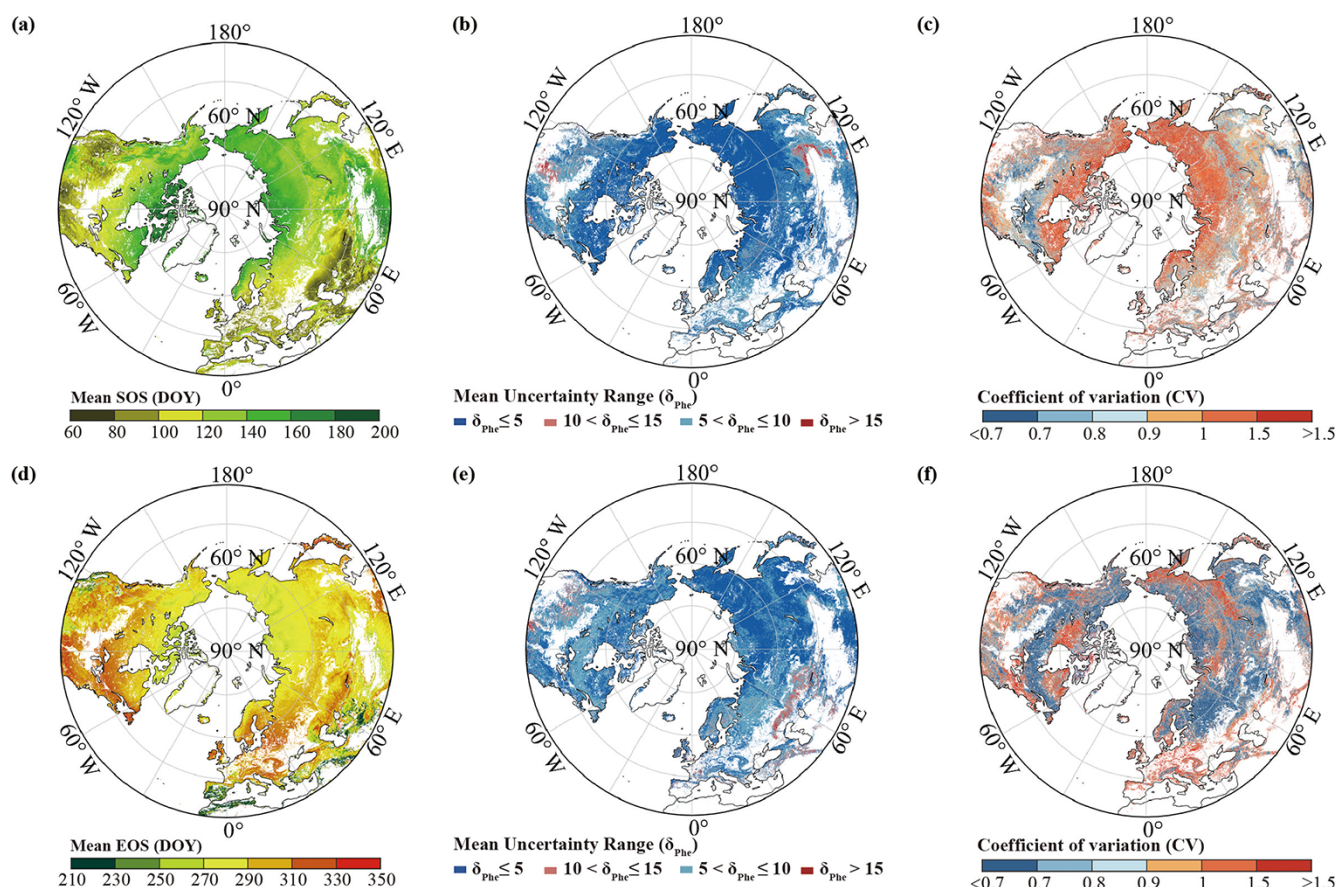
### 3.4 Temporal trends of phenology based on the merged dataset

It is evident from Fig. 6a that the SOS exhibits a significant ( $p < 0.01$ ) trend of advance (earlier dates in the SOS) over the period 1982–2020, with a rate of advance of approximately  $0.19 \text{ d yr}^{-1}$ . Fig. 6b presents the spatial distribution of the SOS trends obtained using the Mann–Kendall test. Approximately 64.37 % of regions exhibit a trend of advance, with 44.88 % exhibiting a significant ( $p < 0.05$ ) trend, while 18.53 % of regions demonstrate a significant trend of delay (later dates in the SOS).

Fig. 6c illustrates that the EOS exhibits a significant trend of delay (later dates in the EOS) with a rate of  $0.18 \text{ d yr}^{-1}$  ( $p < 0.01$ ). It is evident from Fig. 6d that the proportion of areas experiencing a delayed EOS in regions above  $30^\circ \text{N}$  is 68.11 % (comprising 46.60 %; significant at  $p < 0.05$ ), consistent with the corresponding trend depicted in Fig. 6c. Apart from the southwestern to northeastern regions of North America, Europe, the Middle East, and certain high-latitude areas in Asia, EOS delay is predominant.

Over the study area, 46.72 % of regions exhibit SOS advance and EOS delay, 17.73 % show SOS advance and EOS advance, 21.42 % demonstrate SOS delay and EOS delay, and 14.14 % show SOS delay and EOS advance.





**Figure 4.** Merged mean (a) SOS and (d) EOS dates (DOY) obtained using the REA method for the period 1982–2020 and the uncertainty in the REA merged data. Mean uncertainty ( $\delta_{Ph}$ ) of (b) SOS and (e) EOS dates obtained using the REA method for the period 1982–2020, and its CV in (c) the merged SOS and (f) EOS dates.

## 4 Discussion and conclusions

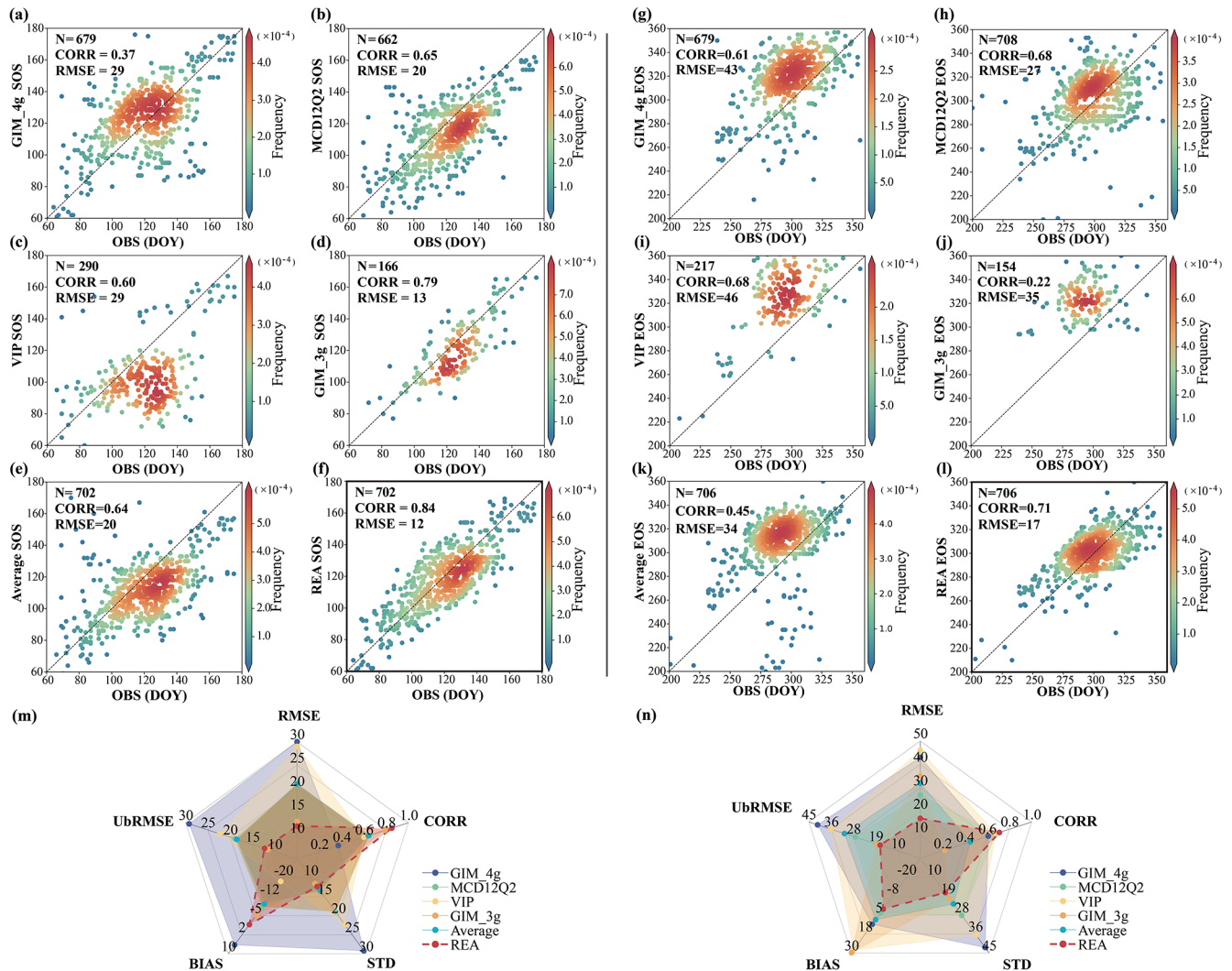
### 4.1 Integrating multi-source phenology data and addressing dataset discrepancies

This study integrates four widely used remote sensing phenology datasets (VIP, GIM\_4g, MCD12Q2, and GIM\_3g) using the reliability ensemble averaging (REA) method to improve the consistency of SOS and EOS estimations. Our analysis shows substantial discrepancies among these datasets, consistent with previous findings that SOS estimates can vary by 60 d (White et al., 2009) or even exceed 100 d (De Beurs and Henebry, 2010), depending on the extraction method. We find the inconsistencies that arise from both differences in data sources and variations in phenology extraction methodologies by evaluating the spatial distribution of the multiyear mean SOS and EOS across datasets.

The REA method produces a more harmonized phenology result by assigning different weights to each dataset based on its reliability to mitigate discrepancies. The effectiveness of this approach is demonstrated by the lower deviation of the SOS and EOS estimates between merged results and pheno-

cam datasets. As shown in Fig. 1, the differences in the SOS and EOS among the four datasets across the Northern Hemisphere reach approximately 15 d on average, with some regions exhibiting even larger discrepancies (more than 120 d). These variations might be caused by factors such as sensor characteristics, data preprocessing techniques, and phenological extraction methods. For instance, similarities in spatial distributions between the GIM\_4g and VIP datasets likely stem from their common use of AVHRR sensor data, whereas their mean difference (14 d) could be attributed to differences in time series reconstruction techniques. Differences between GIM\_4g and GIM\_3g are primarily driven by different data fitting methodologies. The data smoothing techniques aim to preserve vegetation dynamics while reducing noise, but the optimal method is still unclear, as the appropriate method depends on the specific biogeographical characteristics of a given region (Hird and McDermid, 2009).

Additionally, phenology extraction methods exhibit varying effectiveness across different ecological regions (Reed, 2006). Evaluating dataset consistency with PhenoCam observations across different plant functional types (PFTs), we



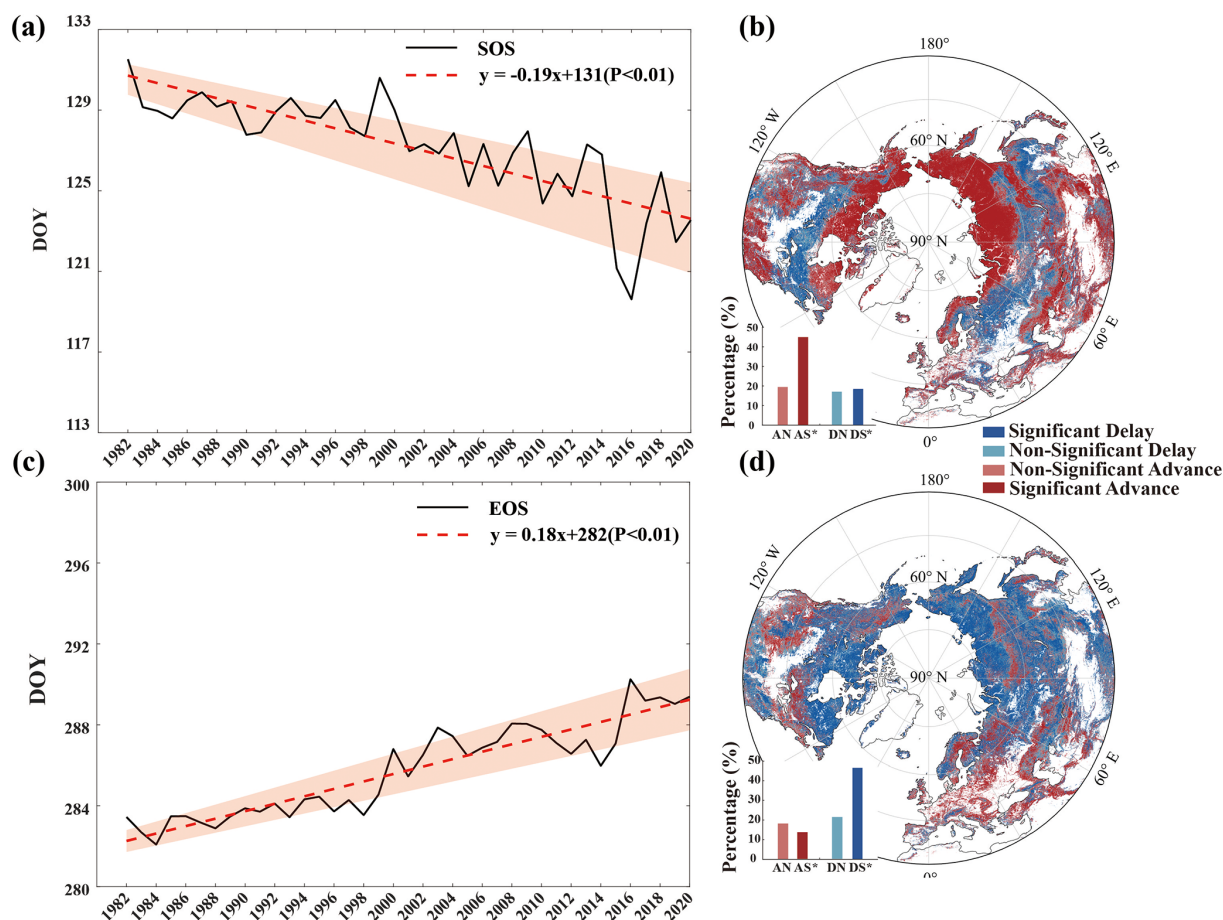
**Figure 5.** Scatterplots and radar charts of performance for each phenology dataset and the merged phenology dataset obtained using the REA method. (a–f) SOS evaluation results of the GIM\_4g, MCD12Q2, VIP, GIM\_3g, average, and REA dataset, (m) radar chart of the SOS evaluation results, (g–l) EOS evaluation results of the GIM\_4g, MCD12Q2, VIP, GIM\_3g, average, and REA dataset, and (n) radar chart of the EOS evaluation results. Each point represents a site year in the figure. OBS indicates the ground-based phenocam phenological dates, RMSE indicates the RMSE, UBRMSE indicates the unbiased RMSE, bias indicates the mean difference between the satellite-based results and the ground-based verification results, SD indicates the standard deviation, and CORR indicates the correlation coefficient. The radar chart is a graphical method used to display multivariate data in the form of a two-dimensional chart with axes starting from the same point.

find that dataset performance varies by PFT and time period. For instance (in Fig. S2), VIP data show consistency with PhenoCam observations in deciduous broadleaf forests in 2001, whereas GIM\_3g outperforms other datasets in 2002. And MCD12Q2 shows better agreement in evergreen needle-leaf forests. These findings stress the challenges of integrating datasets with inherently different methodological foundations.

We further evaluated the greening trend, which reflects the temporal change in the average vegetation greenness during the growing season, as indicated by NDVI trends (Li et al., 2023). The greening trend estimated from

VIP ( $-10.16 \times 10^{-4} \text{ yr}^{-1}$ ) is lower than those from REA ( $-1.14 \times 10^{-4} \text{ yr}^{-1}$ ) and GIM\_4g ( $3.55 \times 10^{-4} \text{ yr}^{-1}$ ). The proportion of areas showing greening trends is 25.22 % (with 41.44 % significantly greening) for VIP, 68.49 % (with 38.32 % significantly greening) for GIM\_4g, and 49.83 % (with 56.83 % significantly greening) for REA. These substantial differences in greenness trends across datasets (Fig. S3) highlight the necessity of an integrated approach for robust phenology estimation.





**Figure 6.** Temporal and spatial trends of the SOS and the EOS over the period 1982–2020 based on the merged dataset obtained using the REA method. (a) Temporal trend of the SOS over the period 1982–2020, (b) spatial trend of the SOS over the period 1982–2020, (c) temporal trend of the EOS over the period 1982–2020, (d) spatial trend of the EOS over the period 1982–2020. The shaded area in (a) and (c) indicates uncertainty at 1 SD, red lines in (a) and (c) are the trend lines of the average SOS/EOS dates for each year, and black lines are the average SOS/EOS date for each year. The abbreviations DS (significant delay), DN (non-significant delay), AS (significant advance), and AN (non-significant advance) are used in the insets of panels (b) and (d).

#### 4.2 Strengths, limitations, and sensitivity analysis of the REA approach

Given the lack of consensus on the “best” approach for extracting phenological dates, our study highlights the value of integrating multiple methodologies rather than relying on a single dataset. The REA method improves phenology estimations by weighting datasets based on reliability rather than assuming equal contributions, as in a simple arithmetic average (Lu et al., 2021). The REA method considers the temporal correlation of vegetation phenology data by employing a voting principle (Giorgi and Mearns, 2002), and this approach facilitates convergence of data while retaining differences of the spatial distribution, thereby offering advantages in multi-source data fusion over simple averaging. Our results indicate that REA merged phenology results are more consistent with PhenoCam observations than individual datasets and simple average results, demonstrating its po-

tential to reduce biases between different datasets caused by different extraction methods. However, while REA enhances consistency, it does not fully eliminate methodological discrepancies.

To evaluate the robustness of this approach, we conducted sensitivity analyses using different dataset and time span combinations. Our results (see Figs. S6 and S7) indicate that the combination of different input datasets has a greater influence on fusion results than the length of the time series. The average deviation of SOS/EOS dates between different dataset combinations and the long-term fusion result using all four datasets is 7.1 d, whereas the SOS and EOS variations across different fusion time spans (e.g., 5-year vs. 10-year fusion) are relatively minor (4.1 and 3.2 d; see Fig. S6). Moreover, fusion weights derived from shorter time series are largely consistent with those obtained using the full dataset (see Fig. S8). In the sensitivity analysis assessing the impact of the fusion time span, the SOS results were more stable

than the EOS, and the fusion results based on longer time series were more similar to the full-period REA fusion result (1982–2020).

The effectiveness of data fusion depends on the complementarity and quality of input datasets. Our analysis suggests that datasets incorporating more reliable sources could produce improved fusion results. For example, GIM\_3g SOS demonstrate higher consistency with phenocam observations (Fig. 5d); groups 2 and 3, which include GIM\_3g SOS, exhibit smaller deviations from the long-term REA fusion results than group 1. In contrast, the EOS estimates are more stable across different dataset combinations. These results show that multi-source data integration could help improve accuracy. The similarity in weight distributions across different dataset combinations (Fig. S9) suggests that dataset selection remains a critical factor influencing final estimates. Incorporating additional high-quality datasets in future studies could further enhance accuracy and robustness.

#### 4.3 Consistency of SOS and EOS trends across studies and key influencing factors

Global climate change has significantly influenced vegetation phenology, with a general advancement in the SOS and a delay in the EOS (Piao et al., 2019). Our REA-based dataset estimates an SOS advancement of 0.19 d per year during 1982–2020 ( $p < 0.01$ ), aligning with previous findings, such as  $1.4 \pm 0.6$  d per decade for 1982–2011 (Wang et al., 2015) and a 5.4 d advance from 1982 to 2008 (Jeong et al., 2011). Similarly, the EOS has been delayed at a rate of 0.18 d per year ( $p < 0.01$ ), consistent with prior estimates of  $0.18 \pm 0.38$  d per year for 1982–2011 (Liu et al., 2016) and a 6.6 d delay from 1982 to 2008 (Jeong et al., 2011). Deviations between REA-based trends and those from individual datasets (Blunden et al., 2023) suggest that variations in data sources and study regions contribute to observed discrepancies.

Differences in the SOS trends across studies arise from multiple factors, including the study period, land cover changes, spatial domain, environmental heterogeneity, and methodological differences. Trend estimates are highly sensitive to the selected time window, particularly at the start and end years (Cong et al., 2013). Interannual and decadal climate variability can further influence these trends, contributing to discrepancies across studies. Land cover changes, such as wildfires and deforestation (Jeong et al., 2011), can influence the extraction result of phenology, then affect SOS trends. Additionally, spatial differences in studies also influence the rate of trends in different studies, as phenology is affected by vegetation type and climate sensitivity. For instance, Jeong et al. (2011) analyzed temperate vegetation between 30–80° N (SOS advance of 0.2 d per year), and Wang et al. (2015) focused on 30–75° N (SOS advance of  $0.14 \pm 0.6$  d per year), excluding evergreen forests and managed landscapes. Since phenology responds differently

across species and locations (Maignan et al., 2008), variations in the study areas contribute to discrepancies. Photo-period constraints introduce additional latitudinal differences in phenological responses to warming (Meng et al., 2021). Moreover, precipitation and other environmental factors vary spatially, further influencing SOS estimates.

Different methods of extracting phenology from remote sensing data could introduce uncertainties between different datasets (Cong et al., 2013; Jeong et al., 2011; White et al., 2009). Differences in filling missing data and filtering methods can affect the continuity of the vegetation index time series, leading to discrepancies in phenology results. Reducing these inconsistencies among different datasets and producing more harmonized data are primary motivations for employing the REA method, which reduces uncertainties by integrating datasets based on their reliability. Recognizing these factors also explains the discrepancies in the SOS/EOS trends and stresses the need for standardized methodologies in phenological trend analyses.

#### 5 Code and data availability

The code used to generate the REA phenology dataset is openly available: <https://doi.org/10.5281/zenodo.16419015> (Cui, 2025). The MCD12Q2 phenology dataset is available at <https://doi.org/10.5067/MODIS/MCD12Q2.061> (Friedl et al., 2022); the VIP phenology dataset is available at [https://doi.org/10.5067/MEaSUREs/VIP/VIPPHEN\\_NDVI.004](https://doi.org/10.5067/MEaSUREs/VIP/VIPPHEN_NDVI.004) (Didan and Barreto, 2016); the GIM\_3g phenology dataset is available at [http://data.globalecology.unh.edu/data/GIMMS\\_NDVI3g\\_Phenology/](http://data.globalecology.unh.edu/data/GIMMS_NDVI3g_Phenology/) (Wang et al., 2019); the GIM\_4g phenology dataset is available at <https://doi.org/10.5281/zenodo.11136967> (Chen and Fu, 2024); and the camera-based phenology datasets are available from the ORNL DAAC (<https://doi.org/10.3334/ORNLDAC/1674>) (Richardson et al., 2018), the Internet Nature Information System of Japan (<http://www.sizenken.biodic.go.jp>, last access: 12 May 2024, Ministry of the Environment of Japan, 2024), and the Phenological Eyes Network (<http://www.pheno-eye.org>, last access: 13 May 2024, PEN, 2024). The land use dataset is available at <https://doi.org/10.5067/MODIS/MCD12Q1.061> (Friedl and Sulla-Menashe, 2022), and the REA phenology dataset is available at <https://doi.org/10.5281/zenodo.15165681> (Cui and Fu, 2024).

#### 6 Conclusion and perspective

Shifts in vegetation phenology affect ecosystem structure (Kharouba et al., 2018; Yang and Rudolf, 2010), consequently affecting biodiversity (Renner and Zohner, 2018), terrestrial carbon and water cycles (Piao et al., 2020), and the climate system (Green et al., 2017; Piao et al., 2020). The

establishment of a comprehensive and reliable vegetation phenology dataset is thus of critical importance. Our study demonstrated that the REA method provides a robust approach for integrating multi-source phenology datasets and produced a vegetation phenology dataset for regions above 30° in the Northern Hemisphere from 1982 to 2020. This dataset could be used for subsequent analyses, such as examining vegetation phenology dynamics and their impacts on the terrestrial carbon cycle and water balance, and providing climatic feedback for global vegetation dynamics modeling. Integrating more vegetation phenology datasets that consider regional characteristics and refining the weight could be done in a future study, which could improve the accuracy and reliability of the merged phenology dataset. Additionally, higher-resolution phenology datasets will also improve the consistency between remote sensing phenology datasets and ground-based results. High-resolution datasets will enhance our ability to assess phenological changes in heterogeneous landscapes and improve local-scale ecological modeling, offering new opportunities to enhance the monitoring and prediction of vegetation dynamics in response to environmental change.

**Supplement.** The supplement related to this article is available online at <https://doi.org/10.5194/essd-17-4005-2025-supplement>.

**Author contributions.** YHF developed the preliminary conceptualization; YC performed the study, YC, YHF, and YZ wrote the manuscript; YHF and YZ supervised the manuscript construction and revision; SC, YG, ML, and ZJ participated in reviewing and editing the paper. All of the authors have read and approved the paper.

**Competing interests.** At least one of the (co-)authors is a member of the editorial board of *Earth System Science Data*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

**Acknowledgements.** We are grateful to the anonymous reviewers and editor for their valuable suggestions that helped us strengthen the quality of our paper. We thank Liwen Bianji for editing the English text of a draft of this paper. Data used in this research were provided by the PhenoCam Network, which has been supported by the National Science Foundation, the Long-Term Agroecosystem Research (LTAR) network supported by the United States

Department of Agriculture (USDA), the US Department of Energy, the US Geological Survey, the Northeastern States Research Cooperative, and the USA National Phenology Network. We thank the PhenoCam Network collaborators, including site principal investigators (PIs) and technicians, for publicly sharing the data that were used in this paper.

**Financial support.** This research has been supported by the National Funds for Distinguished Young Youths (grant no. 42025101), the National Key Research and Development Program of China (grant no. 2023YFF0805604), the Fundamental Research Funds for the Central Universities (grant no. 2243300004), and the 111 Project (grant no. B18006).

**Review statement.** This paper was edited by Andrew Feldman and reviewed by two anonymous referees.

## References

- Blunden, J., Boyer, T., and Bartow-Gillies, E.: State of the Climate in 2022, B. Am. Meteorol. Soc., 104, S107, <https://doi.org/10.1175/2023BAMSStateoftheClimate.1>, 2023.
- Chen, S. and Fu, Y.: Vegetation phenology data based on GIMMS4g NDVI from 1982 to 2020, Zenodo [data set], <https://doi.org/10.5281/zenodo.11136967>, 2024.
- Chen, S., Fu, Y. H., Li, M., Jia, Z., Cui, Y., and Tang, J.: A new temperature–photoperiod coupled phenology module in LPJ-GUESS model v4.1: optimizing estimation of terrestrial carbon and water processes, *Geosci. Model Dev.*, 17, 2509–2523, <https://doi.org/10.5194/gmd-17-2509-2024>, 2024.
- Cong, N., Piao, S., Chen, A., Wang, X., Lin, X., Chen, S., Han, S., Zhou, G., and Zhang, X.: Spring vegetation green-up date in China inferred from SPOT NDVI data: A multiple model analysis, *Agr. Forest Meteorol.*, 165, 104–113, <https://doi.org/10.1016/j.agrformet.2012.06.009>, 2012.
- Cong, N., Wang, T., Nan, H., Ma, Y., Wang, X., Myneni, R. B., and Piao, S.: Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multimethod analysis, *Glob. Change Biol.*, 19, 881–891, <https://doi.org/10.1111/gcb.12077>, 2013.
- Cui, Y.: REA: A codebase for generating REA phenology datasets (Version 1.0), Zenodo [code], <https://doi.org/10.5281/zenodo.16419015>, 2025b.
- Cui, Y. and Fu, Y.: A vegetation phenology dataset by integrating multiple sources using the Reliability Ensemble Averaging method, Zenodo [data set], <https://doi.org/10.5281/zenodo.15165681>, 2024.
- De Beurs, K. M. and Henebry, G. M.: Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology, in: *Phenological Research*, edited by: Hudson, I. L. and Keatley, M. R., Springer Netherlands, Dordrecht, 177–208, [https://doi.org/10.1007/978-90-481-3335-2\\_9](https://doi.org/10.1007/978-90-481-3335-2_9), 2010.
- Didan, K. and Barreto, A.: NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology NDVI Yearly Global 0.05Deg CMG, USGS [data set],

- [https://doi.org/10.5067/MEaSURES/VIP/VIPPHEN\\_NDVI.004](https://doi.org/10.5067/MEaSURES/VIP/VIPPHEN_NDVI.004), 2016.
- Ettinger, A. K., Chamberlain, C. J., Morales-Castilla, I., Buonaiuto, D. M., Flynn, D. F. B., Savas, T., Samaha, J. A., and Wolkovich, E. M.: Winter temperatures predominate in spring phenological responses to warming, *Nat. Clim. Change*, 10, 1137–1142, <https://doi.org/10.1038/s41558-020-00917-3>, 2020.
- Fensholt, R. and Proud, S. R.: Evaluation of Earth Observation based global long term vegetation trends – Comparing GIMMS and MODIS global NDVI time series, *Remote Sens. Environ.*, 119, 131–147, <https://doi.org/10.1016/j.rse.2011.12.015>, 2012.
- Friedl, M. and Sulla-Menashe, D.: MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061, USGS [data set], <https://doi.org/10.5067/MODIS/MCD12Q1.061>, 2022.
- Friedl, M., Gray, J., and Sulla-Menashe, D.: MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V061, USGS [data set], <https://doi.org/10.5067/MODIS/MCD12Q2.061>, 2022.
- Fu, Y. H., Piao, S., Op de Beeck, M., Cong, N., Zhao, H., Zhang, Y., Menzel, A., and Janssens, I. A.: Recent spring phenology shifts in western Europe based on multiscale observations, *Global Ecol. Biogeogr.*, 23, 1255–1263, <https://doi.org/10.1111/geb.12210>, 2014.
- Gao, F., Masek, J., Schwaller, M., and Hall, F.: On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance, *IEEE T. Geosci. Remote*, 44, 2207–2218, <https://doi.org/10.1109/TGRS.2006.872081>, 2006.
- Geng, X., Fu, Y. H., Hao, F., Zhou, X., Zhang, X., Yin, G., Vitasse, Y., Piao, S., Niu, K., and De Boeck, H. J.: Climate warming increases spring phenological differences among temperate trees, *Glob. Change Biol.*, 26, 5979–5987, <https://doi.org/10.1111/gcb.15301>, 2020.
- Gevaert, C. M. and García-Haro, F. J.: A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion, *Remote Sens. Environ.*, 156, 34–44, <https://doi.org/10.1016/j.rse.2014.09.012>, 2015.
- Giorgi, F. and Mearns, L. O.: Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the “Reliability Ensemble Averaging” (REA) Method, *J. Climate*, 15, 1141–1158, [https://doi.org/10.1175/1520-0442\(2002\)015<1141:COAURA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2), 2002.
- Gray, J., Sulla-Menashe, D., and Friedl, M. A.: User guide to Collection 6 MODIS Land Cover Dynamics (MCD12Q2) product, NASA EOSDIS Land Processes DAAC: Missoula, MT, USA, 6, 1–8, 2019, <https://doi.org/10.5067/MODIS/MCD12Q2.061>, 2024.
- Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., Lee, J.-E., and Gentile, P.: Regionally strong feedbacks between the atmosphere and terrestrial biosphere, *Nat. Geosci.*, 10, 410–414, <https://doi.org/10.1038/ngeo2957>, 2017.
- Hird, J. N. and McDermid, G. J.: Noise reduction of NDVI time series: An empirical comparison of selected techniques, *Remote Sens. Environ.*, 113, 248–258, 2009.
- Ide, R. and Oguma, H.: Use of digital cameras for phenological observations, *Ecol. Inform.*, 5, 339–347, <https://doi.org/10.1016/j.ecoinf.2010.07.002>, 2010.
- Inoue, T., Nagai, S., Saitoh, T. M., Muraoka, H., Nasahara, K. N., and Koizumi, H.: Detection of the different characteristics of year-to-year variation in foliage phenology among deciduous broad-leaved tree species by using daily continuous canopy surface images, *Ecol. Inform.*, 22, 58–68, <https://doi.org/10.1016/j.ecoinf.2014.05.009>, 2014.
- Jeong, S.-J., Ho, C.-H., Gim, H.-J., and Brown, M. E.: Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008, *Glob. Change Biol.*, 17, 2385–2399, <https://doi.org/10.1111/j.1365-2486.2011.02397.x>, 2011.
- Kendall, M. G.: Rank correlation methods, 4th edn., London: Griffin, ISBN 9780852641996, 1970.
- Kharouba, H. M., Ehrlén, J., Gelman, A., Bolmgren, K., Allen, J. M., Travers, S. E., and Wolkovich, E. M.: Global shifts in the phenological synchrony of species interactions over recent decades, *P. Natl. Acad. Sci. USA*, 115, 5211–5216, <https://doi.org/10.1073/pnas.1714511115>, 2018.
- Li, M., Cao, S., Zhu, Z., Wang, Z., Myneni, R. B., and Piao, S.: Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022 (V1.2) (V1.2), Zenodo [data set], <https://doi.org/10.5281/zenodo.8253971>, 2023.
- Liu, Q., Fu, Y. H., Zhu, Z., Liu, Y., Liu, Z., Huang, M., Janssens, I. A., and Piao, S.: Delayed autumn phenology in the Northern Hemisphere is related to change in both climate and spring phenology, *Glob. Change Biol.*, 22, 3702–3711, 2016.
- Lu, J., Wang, G., Chen, T., Li, S., Hagan, D. F. T., Kattel, G., Peng, J., Jiang, T., and Su, B.: A harmonized global land evaporation dataset from model-based products covering 1980–2017, *Earth Syst. Sci. Data*, 13, 5879–5898, <https://doi.org/10.5194/essd-13-5879-2021>, 2021.
- Ma, L., Zhou, Y., Chen, J., Cao, X., and Chen, X.: Estimation of Fractional Vegetation Cover in Semiarid Areas by Integrating Endmember Reflectance Purification Into Nonlinear Spectral Mixture Analysis, *IEEE Geosci. Remote S.*, 12, 1175–1179, <https://doi.org/10.1109/LGRS.2014.2385816>, 2015.
- Maignan, F., Bréon, F.-M., Bacour, C., Demarty, J., and Poirson, A.: Interannual vegetation phenology estimates from global AVHRR measurements: Comparison with in situ data and applications, *Remote Sens. Environ.*, 112, 496–505, 2008.
- Meng, L., Zhou, Y., Gu, L., Richardson, A. D., Peñuelas, J., Fu, Y., Wang, Y., Asrar, G. R., De Boeck, H. J., Mao, J., Zhang, Y., and Wang, Z.: Photoperiod decelerates the advance of spring phenology of six deciduous tree species under climate warming, *Glob. Change Biol.*, 27, 2914–2927, <https://doi.org/10.1111/gcb.15575>, 2021.
- Ministry of the Environment of Japan: Internet Nature Information System, Ministry of the Environment of Japan [data set], <http://www.sizenken.biodic.go.jp/>, last access: 12 May 2024.
- Moon, M., Richardson, A. D., and Friedl, M. A.: Multi-scale assessment of land surface phenology from harmonized Landsat 8 and Sentinel-2, PlanetScope, and PhenoCam imagery, *Remote Sens. Environ.*, 266, 112716, <https://doi.org/10.1016/j.rse.2021.112716>, 2021.
- Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L.: The interpretation of spectral vegetation indexes, *IEEE T. Geosci. Remote*, 33, 481–486, 1995.



- Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R.: Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, 386, 698–702, 1997.
- Nasahara, K. N. and Nagai, S.: Review: Development of an in situ observation network for terrestrial ecological remote sensing: the Phenological Eyes Network (PEN), *Ecol. Res.*, 30, 211–223, <https://doi.org/10.1007/s11284-014-1239-x>, 2015.
- PEN (Phenological Eyes Network): Phenological Eyes Network (PEN), PEN [data set], <http://www.pheno-eye.org/>, last access: 13 May 2024.
- Peng, D., Zhang, X., Wu, C., Huang, W., Gonsamo, A., Huete, A. R., Didan, K., Tan, B., Liu, X., and Zhang, B.: Intercomparison and evaluation of spring phenology products using National Phenology Network and AmeriFlux observations in the contiguous United States, *Agr. Forest Meteorol.*, 242, 33–46, <https://doi.org/10.1016/j.agrformet.2017.04.009>, 2017.
- Peñuelas, J., Rutishauser, T., and Filella, I.: Phenology feedbacks on climate change, *Science*, 324, 887–888, <https://doi.org/10.1126/science.1173004>, 2009.
- Pettorelli, N., Vik, J. O., Myrsetrud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C.: Using the satellite-derived NDVI to assess ecological responses to environmental change, *Trends Ecol. Evol.*, 20, 503–510, 2005.
- Piao, S., Fang, J., Zhou, L., Ciais, P., and Zhu, B.: Variations in satellite-derived phenology in China's temperate vegetation, *Glob. Change Biol.*, 12, 672–685, <https://doi.org/10.1111/j.1365-2486.2006.01123.x>, 2006.
- Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S., and Wang, T.: The carbon balance of terrestrial ecosystems in China, *Nature*, 458, 1009–1013, <https://doi.org/10.1038/nature07944>, 2009.
- Piao, S., Tan, J., Chen, A., Fu, Y. H., Ciais, P., Liu, Q., Janssens, I. A., Vicca, S., Zeng, Z., and Jeong, S.-J.: Leaf onset in the northern hemisphere triggered by daytime temperature, *Nat. Commun.*, 6, 6911, <https://doi.org/10.1038/ncomms7911>, 2015.
- Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., and Zhu, X.: Plant phenology and global climate change: Current progresses and challenges, *Glob. Change Biol.*, 25, 1922–1940, <https://doi.org/10.1111/gcb.14619>, 2019.
- Piao, S., Wang, X., Park, T., Chen, C., Lian, X. U., He, Y., Bjerke, J. W., Chen, A., Ciais, P., and Tømmervik, H.: Characteristics, drivers and feedbacks of global greening, *Nature Reviews Earth & Environment*, 1, 14–27, <https://doi.org/10.1038/s43017-019-0001-x>, 2020.
- Reed, B. C.: Trend Analysis of Time-Series Phenology of North America Derived from Satellite Data, *GIScience Remote Sens.*, 43, 24–38, <https://doi.org/10.2747/1548-1603.43.1.24>, 2006.
- Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O.: Measuring phenological variability from satellite imagery, *J. Veg. Sci.*, 5, 703–714, <https://doi.org/10.2307/3235884>, 1994.
- Renner, S. S. and Zohner, C. M.: Climate change and phenological mismatch in trophic interactions among plants, insects, and vertebrates, *Annual Rev. Ecol. Evol. S.*, 49, 165–182, <https://doi.org/10.1146/annurev-ecolsys-110617-062535>, 2018.
- Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., and Desai, A. R.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, *Glob. Change Biol.*, 18, 566–584, <https://doi.org/10.1111/j.1365-2486.2011.02562.x>, 2012.
- Richardson, A. D., Hufkens, K., Milliman, T., and Frolking, S.: Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1.0 and MODIS satellite remote sensing, *Sci. Rep.*, 8, 5679, <https://doi.org/10.1038/s41598-018-23804-6>, 2018a.
- Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A., and Frolking, S.: Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery, *Sci. Data*, 5, 180028, <https://doi.org/10.1038/sdata.2018.28>, 2018b.
- Ruan, Y., Ruan, B., Xin, Q., Liao, X., Jing, F., and Zhang, X.: phenoC++: An open-source tool for retrieving vegetation phenology from satellite remote sensing data, *Frontiers in Environmental Science*, 11, 215, <https://doi.org/10.3389/fenvs.2023.1097249>, 2023.
- Seyednasrollah, B., Young, A. M., Hufkens, K., Milliman, T., Friedl, M. A., Frolking, S., and Richardson, A. D.: Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset, *Scientific Data*, 6, 222, <https://doi.org/10.1038/s41597-019-0229-9>, 2019a.
- Seyednasrollah, B., Young, A. M., Hufkens, K. et al.: Vegetation Collection PhenoCam Dataset v2.0: Vegetation Phenology from Digital Camera Imagery, 2000–2018, ORNL DAAC [data set], <https://doi.org/10.3334/ORNLDAAC/1674>, 2019b.
- Sisheber, B., Marshall, M., Mengistu, D., and Nelson, A.: Tracking crop phenology in a highly dynamic landscape with knowledge-based Landsat – MODIS data fusion, *Int. J. Appl. Earth Obs.*, 106, 102670, <https://doi.org/10.1016/j.jag.2021.102670>, 2022.
- Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B. H., Milliman, T., O'Keefe, J., and Richardson, A. D.: Digital repeat photography for phenological research in forest ecosystems, *Agr. Forest Meteorol.*, 152, 159–177, 2012.
- Sparks, T. H. and Carey, P. D.: The responses of species to climate over two centuries: an analysis of the Marsham phenological record, 1736–1947, *J. Ecol.*, 321–329, <https://doi.org/10.2307/2261570>, 1995.
- Sulla-Menashe, D. and Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product, USGS: Reston, VA, USA, 1–18, [https://lpdaac.usgs.gov/documents/101/MCD12\\_User\\_Guide\\_V6.pdf](https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf) (last access: 23 May 2024), 2018.
- Tang, J., Körner, C., Muraoka, H., Piao, S., Shen, M., Thackeray, S. J., and Yang, X.: Emerging opportunities and challenges in phenology: a review, *Ecosphere*, 7, e01436, <https://doi.org/10.1002/ecs2.1436>, 2016.
- Wang, X., Piao, S., Xu, X., Ciais, P., MacBean, N., Myneni, R. B., and Li, L.: Has the advancing onset of spring vegetation green-up slowed down or changed abruptly over the last three decades?, *Global Ecol. Biogeogr.*, 24, 621–631, <https://doi.org/10.1111/geb.12289>, 2015.
- Wang, X., Xiao, J., Li, X., Cheng, G., Ma, M., Zhu, G., Altaf Arain, M., Andrew Black, T., and Jassal, R. S.: 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 (data available at: [http://data.globalecology.unh.edu/data/GIMMS\\_NDVI3g\\_Phenology/](http://data.globalecology.unh.edu/data/GIMMS_NDVI3g_Phenology/), last access: 13 May 2024).

- White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., O'keefe, J., Zhang, G., Nemani, R. R., and van Leeuwen, W. J.: Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006, *Glob. Change Biol.*, 15, 2335–2359, <https://doi.org/10.1111/j.1365-2486.2009.01910.x>, 2009.
- Xu, Y., Gao, X., and Giorgi, F.: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections, *Clim. Res.*, 41, 61–81, <https://doi.org/10.3354/cr00835>, 2010.
- Yang, L. H. and Rudolf, V. H. W.: Phenology, ontogeny and the effects of climate change on the timing of species interactions, *Ecol. Lett.*, 13, 1–10, <https://doi.org/10.1111/j.1461-0248.2009.01402.x>, 2010.
- Zhang, H., Chuine, I., Regnier, P., Ciais, P., and Yuan, W.: Deciphering the multiple effects of climate warming on the temporal shift of leaf unfolding, *Nat. Clim. Change*, 12, 193–199, <https://doi.org/10.1038/s41558-021-01261-w>, 2022.
- Zhang, J., Zhao, J., Wang, Y., Zhang, H., Zhang, Z., and Guo, X.: Comparison of land surface phenology in the Northern Hemisphere based on AVHRR GIMMS3g and MODIS datasets, *ISPRS J. Photogramm.*, 169, 1–16, 2020.
- Zhou, X., Geng, X., Yin, G., Hänninen, H., Hao, F., Zhang, X., and Fu, Y. H.: Legacy effect of spring phenology on vegetation growth in temperate China, *Agr. Forest Meteorol.*, 281, 107845, <https://doi.org/10.1016/j.agrformet.2019.107845>, 2020.
- Zhou, Y.: Understanding urban plant phenology for sustainable cities and planet, *Nat. Clim. Change*, 12, 302–304, <https://doi.org/10.1038/s41558-022-01331-7>, 2022.
- Zhu, X., Chen, J., Gao, F., Chen, X., and Masek, J. G.: An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions, *Remote Sens. Environ.*, 114, 2610–2623, <https://doi.org/10.1016/j.rse.2010.05.032>, 2010.