Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe

Feng Tian, Zhanzhang Cai, Hongxiao Jin, Koen Hufkens, Helfried Scheifinger, Torbern Tagesson, Bruno Smets, Roel van Hoolst, Kasper Bonte, Eva Ivits, et al.

HAL Id: hal-03273281
https://hal.inrae.fr/hal-03273281

Submitted on 29 Jun 2021

Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe

Feng Tian [a,b,*], Zhanzhang Cai [b], Hongxiao Jin [b,c], Koen Hufkens [d,e], Helfried Scheifinger [f], Torbern Tagesson [b,g], Bruno Smets [h], Roel Van Hoolst [h], Kasper Bonte [h], Eva Ivits [i], Xiaoye Tong [g], Jonas Ardö [b], Lars Eklundh [b,*]

[a] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China*
[b] *Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden*
[c] *Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark*
[d] *Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium*
[e] *INRA, UMR ISPA, Villenave d'Ornon, France*
[f] *Klima, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria*
[g] *Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark*
[h] *Remote Sensing Unit, Flemish Institute for Technological Research (VITO), B-2400 Mol, Belgium*
[i] *European Environment Agency (EEA), Copenhagen, Denmark*

ARTICLE INFO

ABSTRACT

Vegetation phenology obtained from time series of remote sensing data is relevant for a range of ecological applications. The freely available Sentinel-2 imagery at a 10 m spatial resolution with a ~ 5-day repeat cycle provides an opportunity to map vegetation phenology at an unprecedented fine spatial scale. To facilitate the production of a Europe-wide Copernicus Land Monitoring Sentinel-2 based phenology dataset, we design and evaluate a framework based on a comprehensive set of ground observations, including eddy covariance gross primary production (GPP), PhenoCam green chromatic coordinate (GCC), and phenology phases from the Pan-European Phenological database (PEP725). We test three vegetation indices (VI) — the normalized difference vegetation index (NDVI), the two-band enhanced vegetation index (EVI2), and the plant phenology index (PPI) — regarding their capability to track the seasonal trajectories of GPP and GCC and their performance in reflecting spatial variabilities of the corresponding GPP and GCC phenometrics, i.e., start of season (SOS) and end of season (EOS). We find that for GPP phenology, PPI performs the best, in particular for evergreen coniferous forest areas where the seasonal variations in leaf area are small and snow is prevalent during wintertime. Results are inconclusive for GCC phenology, for which no index is consistently better than the others. When comparing to PEP725 phenology phases, PPI and EVI2 perform better than NDVI regarding the spatial correlation and consistency (i.e., lower standard deviation). We also link VI phenometrics at various amplitude thresholds to the PEP725 phenophases and find that PPI SOS at 25% and PPI EOS at 15% provide the best matches with the ground-observed phenological stages. Finally, we demonstrate that applying bidirectional reflectance distribution function correction to Sentinel-2 reflectance is a step that can be excluded for phenology mapping in Europe.

1. Introduction

Vegetation phenology is the timing of repeated biological events, e.g., leaf emergence, flowering, leaf coloration, and leaf fall (Lieth, 1974; Richardson et al., 2013). It is controlled by molecular mechanisms, related to the plant functioning and driven by cues like photoperiod and temperature (Singh et al., 2017), and recognized as a key indicator of terrestrial ecosystem functioning (Peñuelas and Filella, 2001). Time series of satellite imagery are capable of tracking seasonal foliage variations and phenological events across space and time (Buitenwerf et al., 2015; Cong et al., 2013; Malingreau, 1986; Melaas et al., 2018; Piao et al., 2019).

The mapping of vegetation phenology using satellite data started with the separation of near-infrared (*NIR*) and *red* reflectance bands in the Advanced Very High Resolution Radiometer (AVHRR) sensor in the early 1980s (Justice et al., 1985). The AVHRR sensor has since then enabled daily monitoring of the abundance of photosynthetically active vegetation (Tucker, 1979). The later MODerate resolution Imaging Spectroradiometer (MODIS) sensor improved vegetation monitoring capabilities with a spatial resolution of 250–500 m and a daily revisit period since the year 2000 (Justice et al., 2002). These two sensors have been widely used for regional to global scale phenology studies due to their long-term records, high temporal resolution, and ready-to-use vegetation index (VI) products, and they have greatly advanced our understanding of vegetation phenology and its responses to global change (Brown et al., 2010; de Beurs and Henebry, 2004; Jeong et al., 2011; Malingreau, 1986; Myneni et al., 1997; Piao et al., 2019; Tong et al., 2019; White et al., 1997; Wu et al., 2018; Zhang et al., 2004). Also other coarse-resolution sensors are being used for phenology mapping, e.g. the Visible Infrared Imaging Radiometer Suite (VIIRS), VEGETA-TION, and PROBA-V (Bórnez et al., 2020; Zhang et al., 2018b). To distinguish remotely sensed phenology from in-situ monitoring of specific plant phenophases (e.g., bud-burst, flowering, and leaf-fall), the term land surface phenology is often used (de Beurs and Henebry, 2004; Ganguly et al., 2010; Schwartz and Reed, 1999; White and Nemani, 2006).

With the coarse spatial resolution of these sensors (>250 m), pixels often cover plant communities with a mixture of species and individuals characterized by distinct seasonal patterns, leading to inconclusive biological interpretations for certain conditions (Archibald and Scholes, 2007; Helman, 2018; Zhang et al., 2017). Satellite data with high spatial and temporal resolution are therefore needed. With a 30 m resolution, Landsat has been used for mapping vegetation phenology in the US (Fisher et al., 2006; Melaas et al., 2018, 2016). However, its 16-day revisit period along with the data gaps induced by cloud contamination results in often too sparse time series to obtain robust phenology metrics (Wulder et al., 2016). The Sentinel-2 sensors with a 10 m spatial and 5-day temporal resolution provide an opportunity for continental-scale high-resolution phenology mapping (Drusch et al., 2012), and have been successfully applied at local scales (Cai, 2019; d'Andrimont et al., 2020; Vrieling et al., 2018). A combined Landsat-8 and Sentinel-2 dataset (Claverie et al., 2018) was utilized in the development of the Multi-Source Land Imaging (MuSLI) 30-m land surface phenology dataset covering North America (Bolton et al., 2020). Recently, the European Environmental Agency (EEA) commissioned the development of an operational Sentinel-2 based High-Resolution Vegetation Phenology and Productivity (HR-VPP) dataset at 10 m resolution for Europe, as part of the Copernicus Land Monitoring Service (Copernicus, 2020). This study was conducted as part of that development.

Satellite observations of vegetation phenology need calibration and validation against ground observations. Plant growth stages collected by professional or volunteer observers in national and international phenological networks, such as the Pan European Phenology (PEP725) project (Menzel, 2000; Templ et al., 2018) are valuable for studying plant responses to climate variability and trends (Scheifinger et al., 2003). However, these data suffer from some limitations for remote sensing calibration/validation purposes, including uneven spatial distribution and coverage, often poor geolocation accuracy, and variability of prior experience of citizen scientists. In recent years, digital camera networks for phenology (PhenoCam) have developed, generating near-continuous data that are useful for satellite validation purposes (Richardson et al., 2018, 2007). The Red-Green-Blue (RGB) imagery from these cameras is normally used for generating time-series of greenness indices, e.g. the green chromatic coordinate (GCC, Richardson et al., 2018). A third commonly used datatype is gross primary productivity (GPP) from eddy covariance measurements at flux tower sites. These measure the photosynthetic carbon uptake of the vegetation canopy, and have been used in a number of studies as proxy of

vegetation phenology (Jin et al., 2017; Melaas et al., 2013; Peng et al., 2017). Whereas manual phenology observations and PhenoCam data represent phenology as seen by the human eye, GPP relates more to the photosynthetic phenology that is controlled by both the plant canopy development and the light use efficiency (Medlyn, 1998). Thus, all these data reflect different aspects of vegetation phenological variations, and are useful and complementary for the calibration and validation of satellite-based phenology.

The first step in producing a Europe-wide phenology dataset is to select an optimal VI. The normalized difference vegetation index (NDVI, Tucker, 1979) has commonly been used as a proxy for vegetation productivity and for studying land surface phenology in the past decades. However, NDVI-derived phenology is uncertain for evergreen coniferous forest areas where the seasonal amplitudes are small, and for areas with snow cover where a strong NDVI increase is seen at the time of snowmelt (Delbart et al., 2005; Jin et al., 2017; Jönsson et al., 2010; Norris and Walker, 2020). The enhanced vegetation index (EVI) was found to track vegetation phenology better than NDVI in areas with both dense and sparse vegetation (Klosterman et al., 2014; Zhang et al., 2018a), but may still be problematic in snow covered areas (Huete et al., 2002; Jin and Eklundh, 2014). Bolton et al. (2020) showed the high capability of the two-band EVI (EVI2) in reflecting phenology of ecosystems with strong seasonality, but less so in evergreen ecosystems. As a solution to the problem of mapping phenology of evergreen coniferous forests and of snow-covered regions, Jin and Eklundh (2014) developed the plant phenology index (PPI). While PPI has potential for high-resolution phenology mapping across different ecosystems, it has primarily been tested in boreal regions based on MODIS data (Jin et al., 2019, 2017; Karkauskaite et al., 2017). Thus, it is necessary to test the performance of PPI in more biomes and also with higher spatial resolution data. Besides the choice of VI, there are many other considerations in the derivation of land surface phenology from remotely sensed data, including methods for data smoothing, gap-filling, and derivation of the phenological metrics (Cai et al., 2017; de Beurs and Henebry, 2010; Liu et al., 2017; White et al., 2009).

The aim of this study is to assess the performance of NDVI, EVI2, and PPI, derived from Sentinel-2 data for Europe-wide phenology mapping across different land cover types, and link the satellite-derived phenometrics of start of season (SOS) and end of season (EOS) to ground-observed plant development stages (i.e., assign phenological meanings to the SOS and EOS metrics). Given the sensitivity of VIs to sun-sensor geometry (Jin and Eklundh, 2014; Tagesson et al., 2015) we also analyze the influence of bidirectional reflectance distribution function (BRDF) on the phenometrics. The overarching goal is to support the operational production of the Europe-wide Sentinel-2 phenology dataset through calibration with ground observation networks of eddy covariance, PhenoCam, and PEP725.

2. Methods

The overall framework of the calibration process includes seven steps (Fig. 1):

1) Data collection and creation of time series of NDVI, EVI2, and PPI from April 2017 to March 2020.
2) Comparison of VI time series to GPP and GCC seasonality for 2018.
3) Fitting of regular functions to the noisy and irregular VI time series and extraction of SOS and EOS at various amplitude thresholds.
4) Evaluation of the spatial consistency between VI and GPP/GCC phenometrics.
5) Linking VI phenometrics to PEP725 phenophases to assign certain phenological meanings to the obtained SOS and EOS.
6) Analyzing variabilities in VI phenometrics along latitudinal and elevational transects.
7) Analyzing BRDF effects on VI phenometrics.
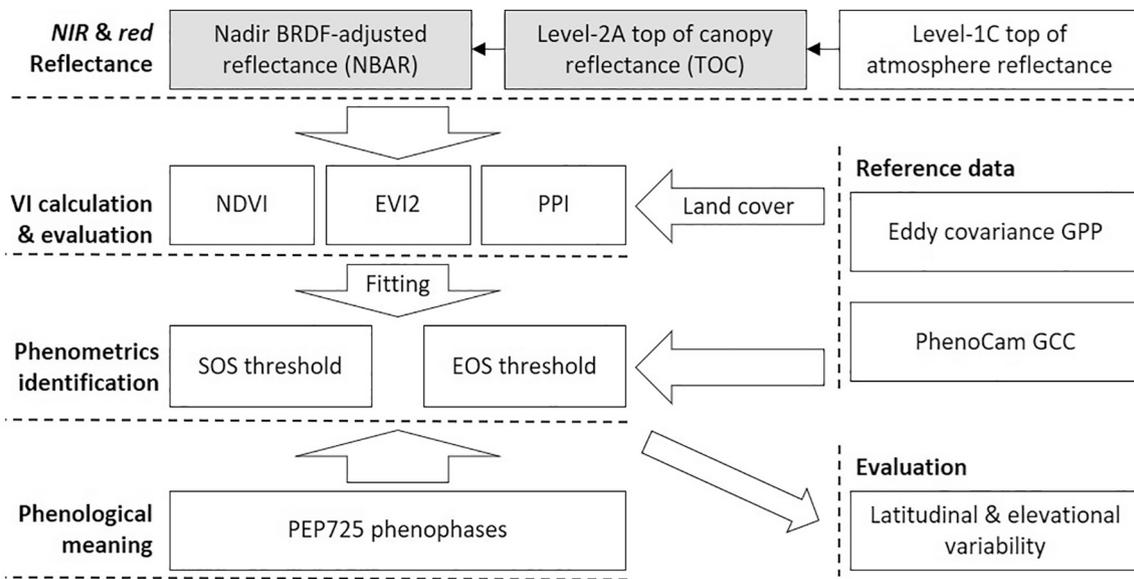
Fig. 1. The overall framework of the calibration process. The impact of using Nadir BRDF-adjusted reflectance (NBAR) or top of canopy reflectance (TOC) as inputs on the phenometrics are assessed.

2.1. Sentinel-2 reflectance data processing

Sentinel-2 reflectance data were extracted from Google Earth Engine (GEE), which were Level-2A top of canopy reflectance (TOC) processed using the Sen2Cor tool by the Copernicus Scientific Data Hub (Doxani et al., 2018). These data include a scene classification (SCL) layer at 20 m resolution indicating the pixel status, based on which we filtered the Sentinel-2 data by removing pixels characterized by clouds, cloud shadows, cirrus, snow/ice, and water, that is, only retaining SCL values of 4 (soil) and 5 (vegetation). Topographic effects were corrected for in the Level-2A surface reflectance products (Supplementary Fig. S1), whereas BRDF correction was not implemented. Despite the narrow field of view of Sentinel-2 (within $\pm 10.3°$ view zenith angle), the BRDF may need to be adjusted for to make the data consistent across space and time, especially for continental-scale studies covering large latitudinal gradients. To generate nadir BRDF-adjusted reflectance (NBAR) for Sentinel-2 reflectance, we used the *c-factor* BRDF normalization approach utilizing the MODIS MCD43A1 BRDF/Albedo products (Roy et al., 2008). The MODIS reflectance at a certain wavelength, location, and time $\rho^{modis}(\theta_v, \theta_s, \varphi)$ for viewing zenith angle θ_v, solar zenith angle θ_s, and relative azimuth angle φ was modeled as:

$$\rho^{modis}(\theta_v, \theta_s, \varphi) = f_{iso} + f_{vol} \cdot k_{vol}(\theta_v, \theta_s, \varphi) + f_{geo} \cdot k_{geo}(\theta_v, \theta_s, \varphi) \quad (1)$$

where $k_{vol}(\theta_v, \theta_s, \varphi)$ and $k_{geo}(\theta_v, \theta_s, \varphi)$ are the volumetric scattering and geometric-optical model kernels, defined by the Ross-Thick (Ross, 1981) and Li-Sparse-reciprocal (Li and Strahler, 1992) equations, respectively; while f_{iso}, f_{vol}, and f_{geo} are the isotropic, volumetric, and geometric model weighting parameters, respectively, provided in the MCD43A1 product. The band dependent *c-factor* for calculating Sentinel NBAR at the local noon solar zenith angle (θ_s^{noon}) and the sensor nadir viewing direction ($\theta_v = 0$) were determined as:

$$c = \frac{\rho^{modis}(0, \theta_s^{noon}, 0)}{\rho^{modis}(\theta_v^{sentinel2}, \theta_s^{sentinel2}, \varphi^{sentinel2})} \quad (2)$$

Then the Sentinel-2 NBAR for *NIR* and *red* bands were obtained by multiplying the Level-2A surface reflectance with their respective *c-factor*.

Based on the MODIS BRDF/Albedo parameters product, Roy et al. (2016) generalized a set of constant f_{iso}, f_{vol}, and f_{geo} values to normalize Landsat reflectance using the *c-factor* approach and applied them to Sentinel-2A reflectance (Roy et al., 2017). These constant values ($f_{iso} = 0.1690$, $f_{vol} = 0.0574$, and $f_{geo} = 0.0227$ for the *red* band, and $f_{iso} = 0.3093$, $f_{vol} = 0.1535$, and $f_{geo} = 0.0330$ for the *NIR* band) were used to fill the MCD43A1 data gaps to generate spatiotemporally consistent Sentinel-2 NBAR data in this study.

2.2. Ground data processing and Sentinel-2 data extraction

Ground measurements of eddy covariance GPP, PhenoCam GCC, and PEP725 phenophases were distributed over Europe (Fig. 2) covering the period 2018–2019, representing the major land cover types of deciduous broad-leaved forest, evergreen coniferous forest, mixed forest, cropland, grassland, and wetland.

The GPP data were obtained from the Drought2018 project (https://www.icos-cp.eu/data-products/YVR0-4898), which includes eddy covariance sites across Europe, providing half-hourly net ecosystem exchange measurements and GPP estimations for 1989–2018. The GPP database contained 52 sites (accessed in September 2020), of which 49 were located in the study area (Fig. 2 and Supplementary Table S1) and only used data from 2018 for comparison with Sentinel-2 data. The land cover at each flux tower was determined based on the CORINE land cover map for 2018 (https://land.copernicus.eu/pan-european/corine-land-cover/clc2018) and was grouped into six categories: broad-leaved forest (code ID 311), coniferous forest (code ID 312), mixed forest (code ID 313 and 324), agriculture (code ID 211, 212, 221, 243, and 244), grassland (code ID 231 and 321), and wetland (code ID 411 and 412). We aggregated the half-hourly GPP values to daily values by averaging, as has been done for the FLUXNET2015 dataset. To extract the Sentinel-2 pixels matching the eddy covariance data, we used a buffer of 100 m radius around the flux tower and computed the mean values of *NIR and red* reflectance. This is a reasonable size matching the eddy covariance surface source areas; in a recent study it was shown that due to the homogeneity of most flux tower sites, the effect of varying the size of the sampling area is limited when staying below 500 m (Cai et al., 2021).

The GCC data were calculated from the RGB camera images using the ratio of the green channel digital numbers to the total brightness of the image as described in Hufkens et al. (2018) and Richardson et al. (2018). The original half-hour GCC time series were aggregated to 3 days by calculating the 90th percentiles to reduce adverse illumination effects caused by atmospheric influences such as rain, snow, and fog. Due to
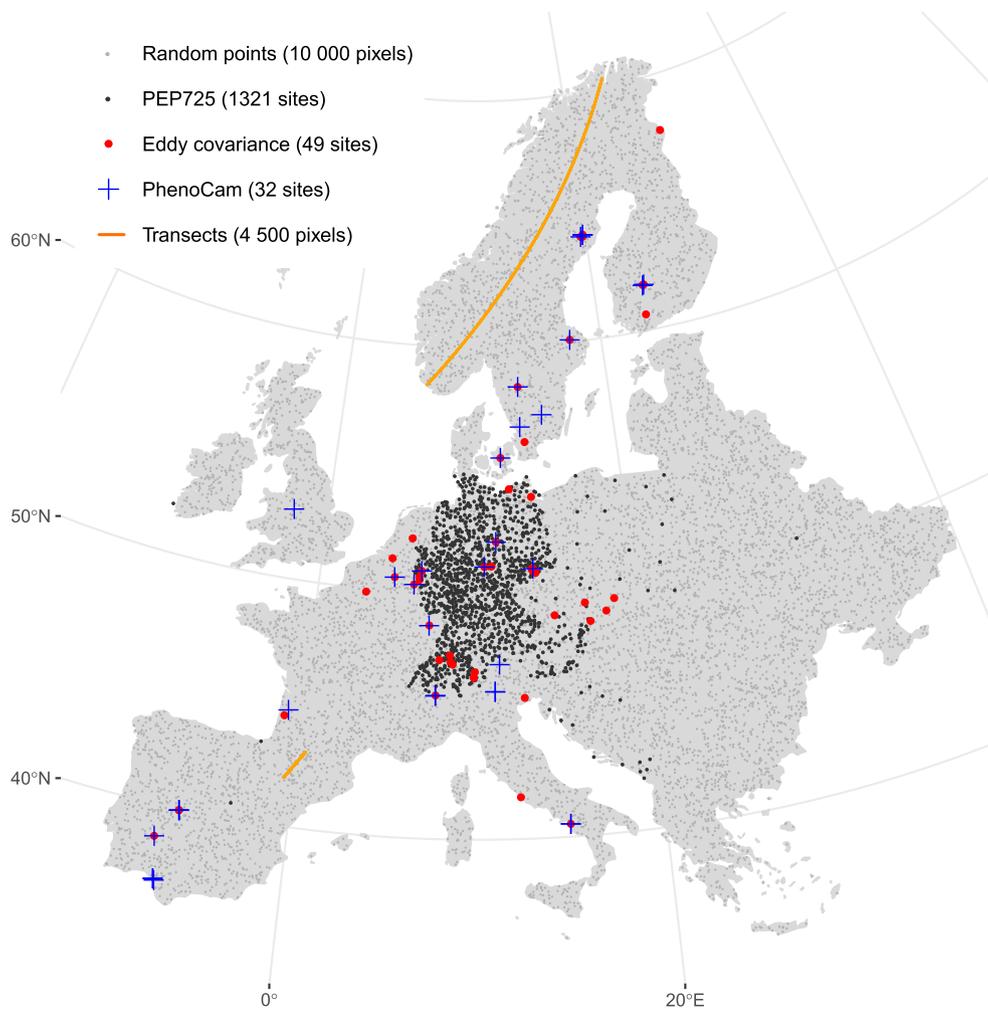
Fig. 2. Locations of the ground data, including the eddy covariance GPP, PhenoCam GCC, and PEP725 phenology phases in the calibration process, the randomly selected points for evaluating the impacts of the BRDF, and the two transects in northern and southern Europe used for evaluation purpose. The northern transect contains 3000 individual Sentinel-2 pixels at every 500 m step and the southern transect contains 1500 individual pixels at every 100 m step.

varying camera settings, the quality of GCC time series at times varied widely and thus required a post-hoc assessment, through which we excluded 19 sites (7 with large gaps and 12 with no seasonal patterns) for the comparison analysis. Given the high resolution of Sentinel-2 data, we did not use the camera location but visually georeferenced the location of each region of interest (ROI, i.e., the parts in the image with studied vegetation canopies; there can be multiple ROIs in one image for heterogeneous regions). The spatial coverage of each ROI was only about one to a few tree crowns, so we used one Sentinel-2 pixel covering the ROI location for the comparison analysis. For sites with contrasting vegetation types (e.g., a sparse evergreen canopy with grass understory), uncertainties in the ROI georeferencing process would introduce a mismatch between the GCC and VI time series, and therefore we excluded 6 sites after visual checking. As a result, among all the 57 PhenoCam sites in Europe, 32 sites with 49 site-year observations during 2018–2019 were used in the analysis (Fig. 2 and Supplementary Table S2).

The PEP725 database contains millions of plant phenological records across Europe (www.pep725.eu), including several plant species and growth stages, collected by both European professionals and volunteers (Templ et al., 2018). Currently, 1321 sites are available in the database for 2018–2019 (Fig. 2). Each site contains phenological records of tens to hundreds of individual plants, but without precise geographical locations for individual records. The coordinates provided for each site roughly mark the locations of phenological records for that site at a

precision of 2–4 km. To link the Sentinel-2 SOS/EOS and the PEP725 records, we created a 4 km buffer around each site, and classified the corresponding Sentinel-2 pixels into three vegetation groups, i.e., deciduous broad-leaved tree, evergreen coniferous tree, and meadow, according to the plant genus information in the PEP725 database and the CORINE land cover map (Table 1). The average time series of Sentinel-2 reflectance bands were then extracted from GEE for each vegetation group. Pixels that did not belong to any of the selected CORINE land

Table 1

Grouping strategy for linking Sentinel-2 SOS/EOS with PEP725 phenology observations for 2018–2019.

Vegetation group	PEP725 genus	CORINE land cover map
Deciduous broad-leaved trees	*Acer (705), Aesculus (2488), Alnus (859), Betula (2787), Fagus (2521), Fraxinus (806), Populus (17), Quercus (2800), Robinia (302), Salix (128), Sambucus (30), Sorbus (1728), Syringa (70), Tilia (954)*	Broad-leaved forest, Transitional woodland-shrub, Mixed forest, Green urban areas
Evergreen coniferous trees	*Picea (1181), Pinus (1)*	Coniferous forest
Meadow	*Meadow (794)*	Pastures, Natural grasslands

The number of phenology records for each genus is shown in brackets.

Table 2

Selected phases in the PEP725 database for linking with Sentinel-2 phenometrics.

Phase_id	Description in the PEP725 database	Short name	Metrics
10	Gramineae: First true leaf emerged from coleoptile; Dicotyledons and Monocotyledons: Cotyledons completely unfolded; Perennial plants: First leaves separated	Leaf emerged	SOS
11	First true leaf, leaf pair or whorl unfolded; Perennial plants: First leaves unfolded	Leaf unfolded	SOS
223	Leaf unfolding ($\geq$ 50%)	Leaf unfolded 50%	SOS
182	25% of the permanent grassland shows fresh green	Fresh green 25%	SOS
205	Autumnal coloration $\geq$ 50%	Coloration 50%	EOS
95	50% of leaves fallen	Leaf fallen 50%	EOS

cover types were excluded from the analysis. An example of buffer areas and the seasonal curves of different vegetation groups is illustrated in Fig. 3. Information on crops in PEP725 was either summer crop, winter crop, or different species during summer/winter, characterized by distinct seasonal patterns that cannot be separated from the CORINE land cover map, and crops were thus excluded from further analysis.

2.3. VI calculation

Based on the Sentinel-2 *NIR* and *red* reflectance data (both NBAR and TOC), we calculated the NDVI, EVI2, and PPI values using the following equations:

$$NDVI = \frac{DVI}{NIR + red} \tag{3}$$

$$EVI2 = \frac{2.5 \cdot DVI}{NIR + 2.4 \cdot red + 1} \tag{4}$$

$$PPI = -K \cdot ln\left(\frac{DVI_{max} - DVI}{DVI - DVI_{soil}}\right) \tag{5}$$

where DVI is the difference between *NIR* and *red* reflectance:

$$DVI = NIR - red. \tag{6}$$

DVI_{max} is the pixel-specific maximum DVI, estimated as the 95% upper quantile from the multi-year time series of DVI values, and DVI_{soil} is an estimate of the DVI of soil background (an empirical value of 0.09 is used in this study and further discussed in Section 4.2). K is the canopy light extinction efficiency, including both light absorption and scattering, computed as:

$$K = \frac{1}{4 \cdot (d_c + 0.5 \cdot (1 - d_c)/cos(\theta_s))} \cdot \frac{1 + DVI_{max}}{1 - DVI_{max}}, \tag{7}$$

where d_c is an instantaneous diffuse fraction of solar radiation at sun zenith angle θ_s (obtained from the scene metadata), computed as:

$$d_c = 0.0336 + 0.0477/cos(\theta_s) \tag{8}$$

2.4. Comparing VI and GPP/GCC seasonal trajectories

We assessed the capability of each VI for tracking vegetation seasonal dynamics in a range of land cover types by comparing the VI values to the GPP and GCC time series. We first used the locally estimated scatterplot smoothing (LOESS, span = 0.2) method to identify and remove residual outliers in each of the VI time series with >1.5 *S.D.* from the smoothed curve. This process was mainly for removing single unrealistic high PPI values in the growing season (see discussion Section 4.2) and was also useful for removing parts of the residual outliers missed by the SCL filtering. To eliminate the high-frequency day-to-day variations in GPP and GCC data, we used TIMESAT (Section 2.5) to obtain smoothed GPP and GCC seasonality. Then, we compared the irregular raw VI values with the smoothed GPP/GCC curves from site level to land cover level, and further to continental level.

2.5. Curve fitting and SOS/EOS extraction

To reconstruct continuous seasonal trajectories from the irregular and noisy Sentinel-2 VI time series, we used a newly developed method with box constrained fits to a superposition of double-logistic model functions (Jönsson et al., 2018). The double-logistic model function was chosen because of its proven robustness and ability to handle large data gaps, e.g., during winter season in the boreal regions (Cai et al., 2017; Fisher et al., 2006; Zhang et al., 2003). The fits are stabilized by a base value, and pixels identified as cloudy are given a weight of zero. Before the double-logistic fitting, the algorithm identifies single or double growing seasons following an initial spline fitting to the original data to enable modeling of irregular seasonality, e.g. shifts from single to
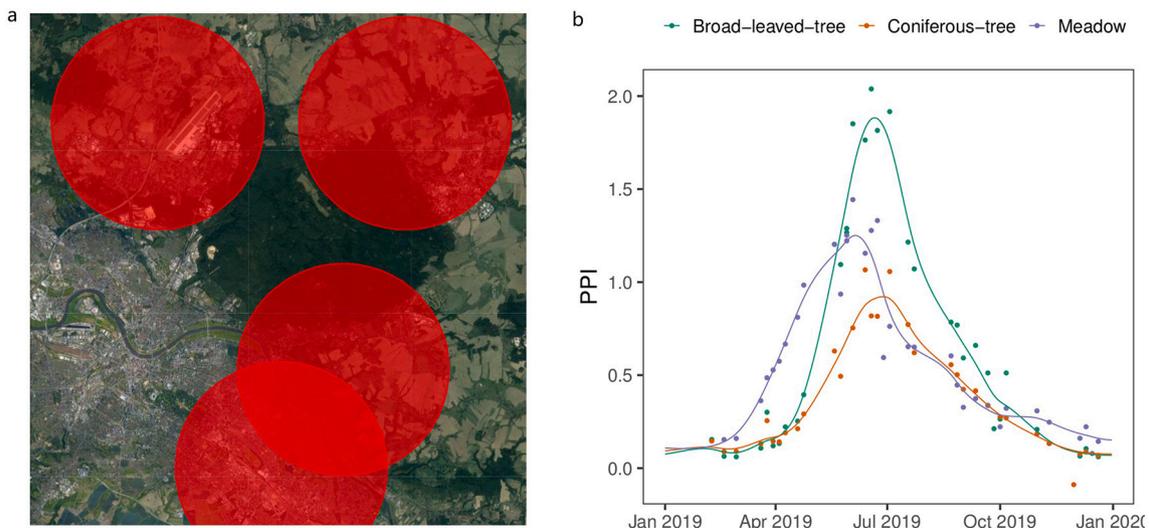
Fig. 3. (a) Example of PEP725 sites with buffer areas of 4 km radius. Each buffer area contains multiple land cover types, which can be characterized by distinct seasonal patterns, as shown in (b) for the upper-right site (site_id 8212 in the PEP725 database).

double-cropping patterns. This method has been implemented in the new version 4.1 of the TIMESAT software package (Eklundh and Jönsson, 2015; Jönsson and Eklundh, 2004).

There are many methods to identify SOS and EOS metrics, which could be broadly grouped into the threshold category and the inflection point category (de Beurs and Henebry, 2004; Jönsson and Eklundh, 2002; White et al., 2009; Zhang et al., 2003). Whereas the inflection point category uses a fixed definition for determining SOS/EOS dates (the mid-point of the fitted function, though other mathematical parameters could also be chosen), the threshold category allows choosing any fraction of the seasonal amplitude from the base level for defining the SOS/EOS dates, and is a method that has been widely used in vegetation phenology studies. The threshold values can be optimized to certain locations and/or land cover types to define SOS and EOS that can be linked with specific phenophases of interest. However, in a continental application, it is advantageous to use a common threshold for all areas. Here we tested a range of threshold values from 5% to 50% by a step of 5% with the primary aim to find single optimal values for SOS and EOS to use at the continental scale.

2.6. Assessing spatial consistency between VI and GPP/GCC phenometrics

We also calculated GPP and GCC phenometrics at varying thresholds from 5% to 50% by a step of 5%, which were used as references for analyzing how well the VI-derived SOS/EOS reflect the spatial variability in the GPP and GCC-derived SOS/EOS across all the sites. The performance of each VI and threshold was quantified by the mean absolute bias and Pearson correlation coefficient (R). More than one SOS/EOS value were obtained for some site-years due to double growing seasons, drought impacts, and VI sensitivities to background disturbances not successfully removed by SCL filtering (Supplementary Figs. S2 and S3). For these site-years, we only used one pair of phenometrics with a minimum bias between VI and GPP/GCC for each site each year to facilitate the comparison analysis.

2.7. Assigning phenological meanings to the VI phenometrics

To explore the phenological meaning of the VI-derived SOS and EOS metrics, i.e. their translation into events that indicate certain development stages in the phenological cycle, we calculated the mean bias (and standard deviation) between the PEP725 phenophases and the VI phenometrics at varying thresholds. Many different phases were recorded in the PEP725 database (www.pep725.eu/pep725_phase.php), from which we selected the six most relevant ones for comparison with the Sentinel-2 SOS and EOS metrics (Table 2). For each site and each phase in the PEP725 database, we calculated the mean value of the different records that belong to the same vegetation group (Table) to obtain a unique value for comparison with the Sentinel-2 SOS/EOS. Given the various limitations of using PEP725 database for calibrating/validating satellite data, we also checked the spatial correlation between VI phenometrics and PEP725 phenophases as a way to verify the robustness of our method.

3. Results

3.1. Time series of VI, GPP, and GCC

Seasonal patterns of the VIs show distinctly different temporal patterns, similarly to the differences between GPP and GCC (see example in Fig. 4 from the DE-HoH site located in central Germany; note that GPP and GCC observations are derived from different pixels), which can be attributed to their different functional meanings (GPP expressing productivity vs. GCC expressing canopy colors). The GPP curves show distinct seasonal patterns with a peak in the beginning of the growing season, and a continuous decrease thereafter, whereas the GCC curves are characterized by large changes in the beginning and the end of the
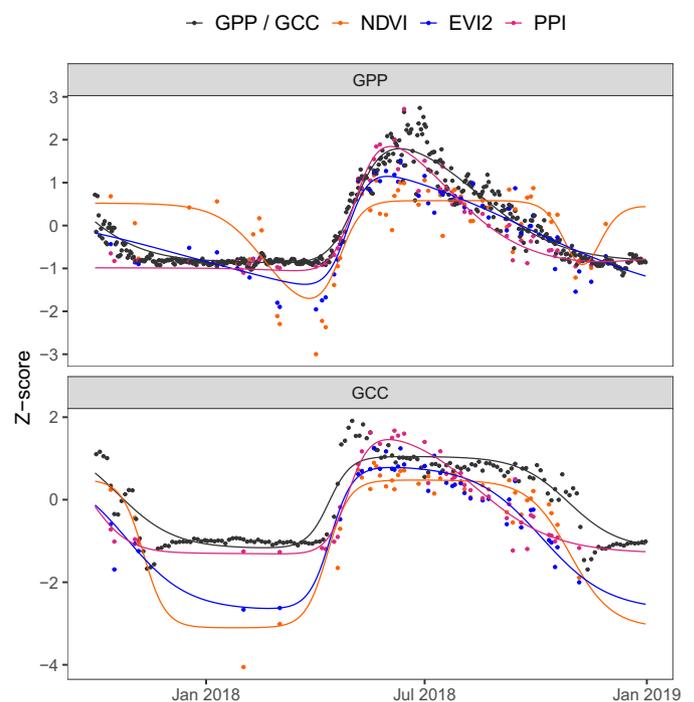
Fig. 4. Time series at the DE-HoH deciduous broad-leaved forest site, comparing the VI observations (colored dots), TIMESAT fitted curves (colored lines), and GPP or GCC (dark grey lines) during 2018. Note that the pixels corresponding to GPP and GCC observations were extracted from one forest but ~200 m apart from each other. All variables were normalized to *Z*-scores using their respective mean and standard deviation values for better visualizations.

growing season, but a plateau in-between (Fig. 4). Visually, PPI is closer to the GPP curve, NDVI is closer to the GCC curve, whereas EVI2 is in-between (Fig. 4). It is noticeable that NDVI and EVI2 show some high values in the winter season for the GPP buffer area, whereas PPI is as flat as the GPP observations. Time series of all the flux and PhenoCam sites are plotted in Supplementary Figs. S2 and S3 respectively, and these data provide the basis for all the following comparisons and analyses.

3.2. Performance of VI in tracking GPP/GCC seasonal variations

PPI performs the best in tracking the GPP seasonal dynamics, with less variation across sites within each land cover type (Fig. 5a), and higher correlation coefficients (Fig. 5b–c), particularly for the evergreen coniferous forest sites. The performance of EVI2 is for some comparisons as high or higher than PPI, but the range between low and high correlations is generally larger and lacks consistency across land cover types. NDVI shows the lowest correlation coefficients for all the land cover types, particularly for the evergreen coniferous forests, where the NDVI and GPP correlation is negative. For the comparison with the PhenoCam GCC observations, none of the VIs shows a distinctly better performance than the others for various land cover types (Fig. 5d and f), but EVI2 shows the highest correlation with GCC for all the sites together ($R = 0.65$), closely followed by NDVI ($R = 0.63$), and PPI performing the worst ($R = 0.56$, Fig. 5e).

3.3. Spatial consistency between VI and GPP/GCC phenometrics

Comparisons between VI and GPP/GCC phenometrics were carried out for all the nested pairs of thresholds, and is shown as a set of correlations and biases (supplementary Fig. S5). The GPP/GCC threshold with the minimum bias as compared to each VI threshold was selected to harmonize the different levels of shifts between seasonal curves of VI and GPP/GCC across land cover types (Fig. 6). As seen, PPI consistently
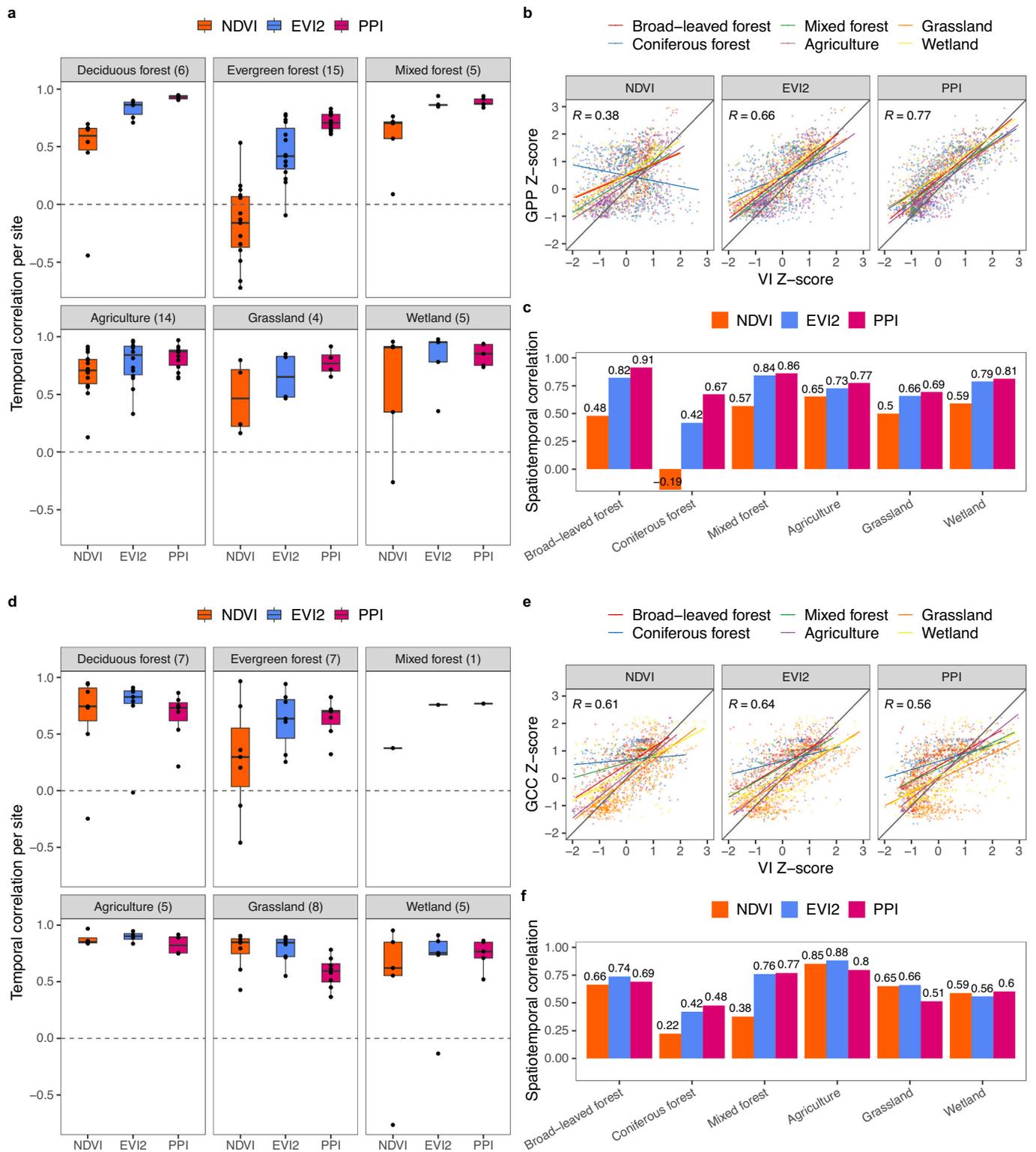
Fig. 5. Performances of raw NDVI, EVI2, and PPI when compared to (a-c) GPP and (d-f) GCC seasonality during 2018. (a, d) Site-level comparison grouped by land cover types, with each black point in the boxplots represents one site, and the total number of sites is noted in each strip caption. (b, e) Scatterplots of all the sites, colored by land cover types with separated regression lines, the Pearson correlation coefficient (R) in each subplot is for all points (i.e., continent-level comparison). (c, f) Land cover-level comparison indicated by R values, corresponding to the regression lines in (b, e).

a   Mean absolute bias

b   Spatial correlation
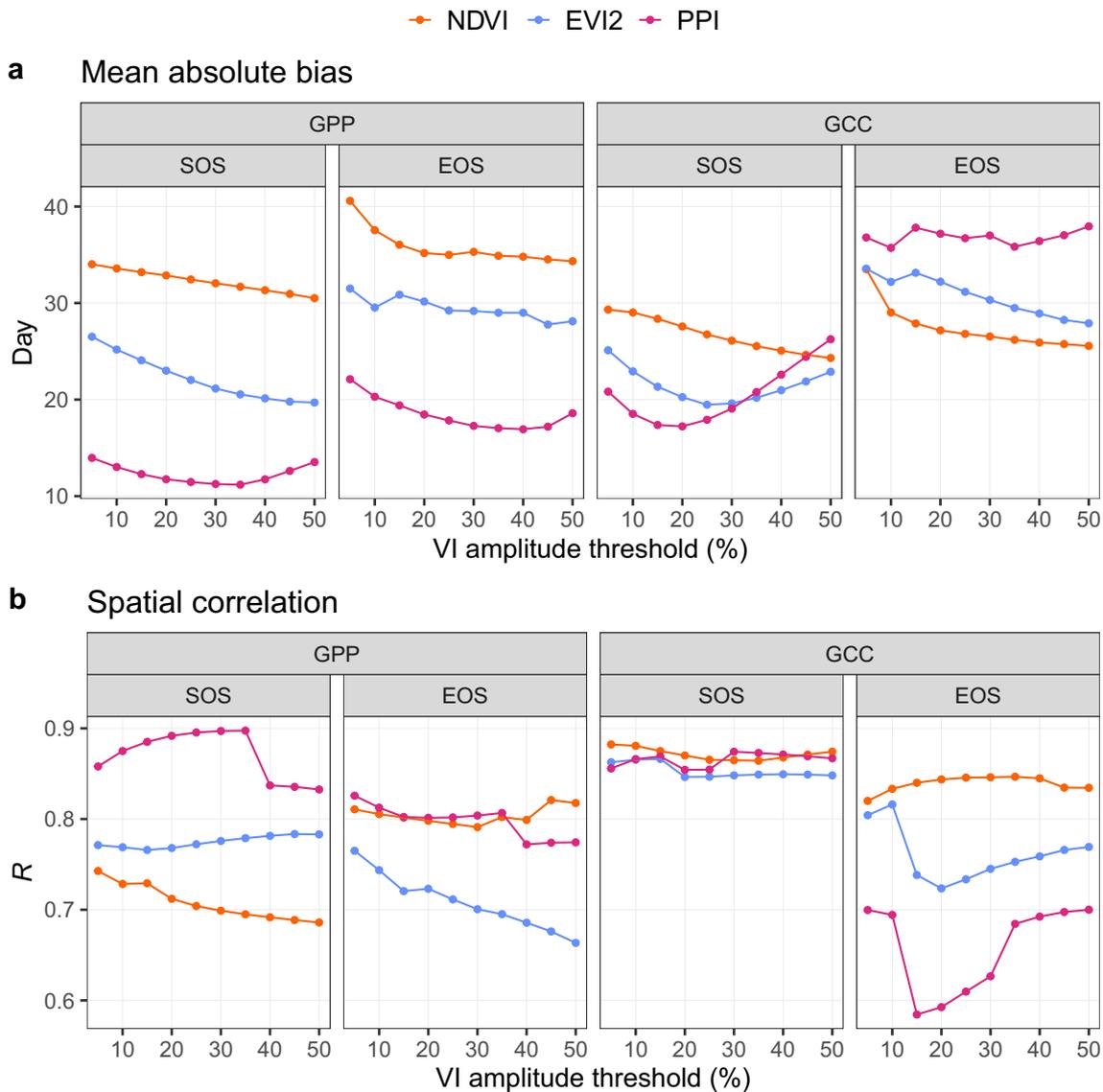
Fig. 6. Comparison of SOS and EOS derived from VI and GPP/GCC ($n = 49$) at various seasonal-amplitude thresholds as quantified by the (a) mean absolute bias and (b) spatial correlation coefficients across all the calibration data sites.

shows the lowest bias and highest R value when comparing to both GPP SOS and GPP EOS (Fig. 6a). As for the comparisons with GCC SOS, all the VIs have comparable performances regarding the spatial correlation coefficients at all tested amplitude thresholds (Fig. 6b), while PPI has lower bias at thresholds below 30% and EVI2 has lower bias above the 30% threshold. For GCC EOS, NDVI consistently shows the best performance with the highest R values and lowest bias (Fig. 6b). The plots are useful for selecting the optimal VI thresholds and for providing information about the sensitivity of the relationships to the choice of threshold values. In most cases the variation in bias and R is low when changing the VI threshold value, but in a few cases the choice of a wrong threshold value can have a negative impact on the accuracy, e.g., R values for GCC EOS, which drop clearly from a threshold of 10% to 15% for PPI and EVI2.

For further illustrating the relationships, we selected the VI thresholds with the highest and lowest accuracy for each comparison to GPP and GCC phenometrics (Fig. 7). The highest and lowest accuracy thresholds were defined based on the ratio between the bias and R values so that the bias value was minimized and the R value was maximized. The similarities between these pairwise plots show that relationships are quite robust to the selection of threshold values.

3.4. Comparison between VI phenometrics and PEP725 phenophases

Most of the paired phases show R values close to zero (Fig. 8a), which could be attributed to 1) the lack of accurate locations of the PEP725 observations, 2) the small spatial dynamic ranges of the PEP725 observations used in this study (mainly in Germany), and 3) uncertainties originated from the subjective interpretation of various observers. However, there are clear positive correlations for the leaf unfolded 50% phase, particularly for the EVI2 and PPI phenometrics, which have R values of 0.61 and 0.41–0.46 at VI threshold 50% for deciduous broad-leaved trees and evergreen coniferous trees, respectively (Fig. 8b). The leaf unfolded phase for evergreen coniferous trees also have relatively high R values (0.52–0.63), although the number of observations is low (only 12, Fig. 8a). These results support the method we employed to compare PEP725 phenophases and Sentinel-2 VI phenometrics.

The optimal VI thresholds matching the best with PEP725 phenophases (both the mean bias and standard deviation should be as low as possible) are different among paired phases and VIs (Fig. 9). To find an overall spring phenology threshold value that is usable across all three vegetation types, we calculated the mean bias and standard deviation values weighted by the number of observations for each paired phase
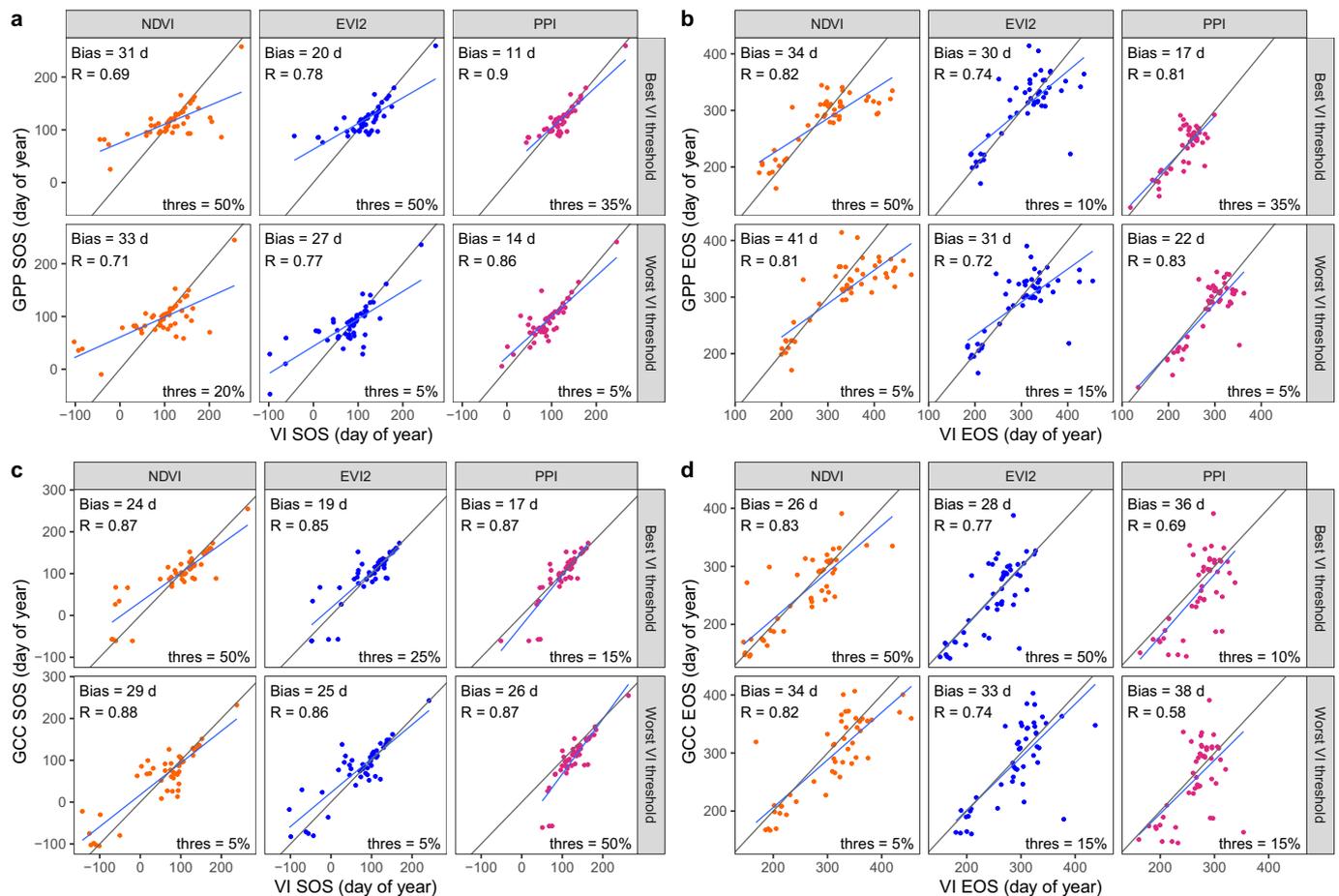
Fig. 7. Scatterplots of (a, c) SOS and (b, d) EOS derived from VI and (a, b) GPP ($n = 49$)/(c, d) GCC ($n = 49$) at the best (highest accuracy) and worst (lowest accuracy) VI thresholds (indicated at the lower right corner of each subplot). The criterion used is the ratio between the mean absolute bias and Pearson R values (all are significant at $p < 0.05$) to minimize the bias value and maximize the R value.

(column-wise in Fig. 9a). The weighted spring mean biases are smallest at threshold 40% for NDVI (5 ± 25 days) and EVI2 (6 ± 14 days) and at threshold 25% for PPI (6 ± 12 days). For these best overall thresholds we checked which of the spring phenophases had the smallest bias for each vegetation type: Deciduous broad-leaved trees matches the best with the leaf unfolded phase for all the three VIs (NDVI: −1 ± 12 days, EVI2: −1 ± 10 days, and PPI: −1 ± 8 days); evergreen coniferous trees matches the best with the leaf unfolded 50% phase for NDVI (6 ± 29 days), and with the leaf emerged phase for EVI2 (8 ± 12 days) and PPI (11 ± 9 days); and the meadow fresh green 25% phase have a match with NDVI of −5 ± 41 days, with EVI2 of −10 ± 22 days, and with PPI of −5 ± 20 days (Fig. 9a). The autumn phases only contain deciduous broad-leaved trees, with the best match at 50% threshold for NDVI for the leaf fallen 50% phase (−12 ± 28 days), at 30% threshold for EVI2 comparing to the leaf fallen 50% phase (−1 ± 20), and at 15% threshold for PPI for the leaf coloration 50% phase (−4 ± 19 days) (Fig. 9b).

3.5. Latitudinal and elevational variability in SOS and EOS

For studying how the estimated SOS and EOS parameters vary across latitudinal and elevational gradients we used the VI SOS/EOS extracted at amplitude threshold that best matched with the PEP725 phases (Section 3.4). Two transects were selected, one located in northern Europe ranging from N58.4° to N70.4°, and one located in southern Europe depicting steeper elevation variations across a mountain region (Fig. 2). The latitudinal and elevational variabilities in SOS and EOS along these two transects are shown in Fig. 10. Overall, EVI2 and PPI are highly consistent with each other and follow the latitudinal and

elevational fluctuations well, particularly for SOS. NDVI aligns with EVI2 and PPI for parts of the transects, e.g., SOS from N42.6° to N43.1° for the southern transect, but shows a different pattern as compared to EVI2 and PPI for other parts of the transects, e.g., SOS and EOS from N42.1° to N42.6° for the southern transect and EOS for the northern transect.

3.6. Impact of BRDF

In addition to using VIs calculated from Sentinel-2 NBAR *NIR* and *red* reflectance for the analyses (Sections 3.2–3.5), we also performed all the analyses based on VIs calculated from the non-BRDF corrected TOC data (Supplementary Figs. S4 and S6-S11). The TOC-based results are almost the same as the NBAR-based results for all the comparisons with GPP, GCC, and PEP725 metrics, except for GPP EOS where NDVI shows a worse performance and EVI2 shows a better performance for TOC-based results as compared to the NBAR-based results (Fig. 6 and Supplementary Fig. S7). To assess the BRDF impacts on the phenological parameters, we plotted SOS and EOS for the 10,000 pixels across Europe calculated from NBAR and TOC reflectance (Fig. 2). The SOS/EOS derived from NBAR and TOC inputs for the three VIs show high consistency, with R and slope values both around 0.94 (Fig. 11). Overall, we found that using TOC rather than NBAR-corrected reflectance would lead to the same conclusions regarding choice of VI and threshold values.
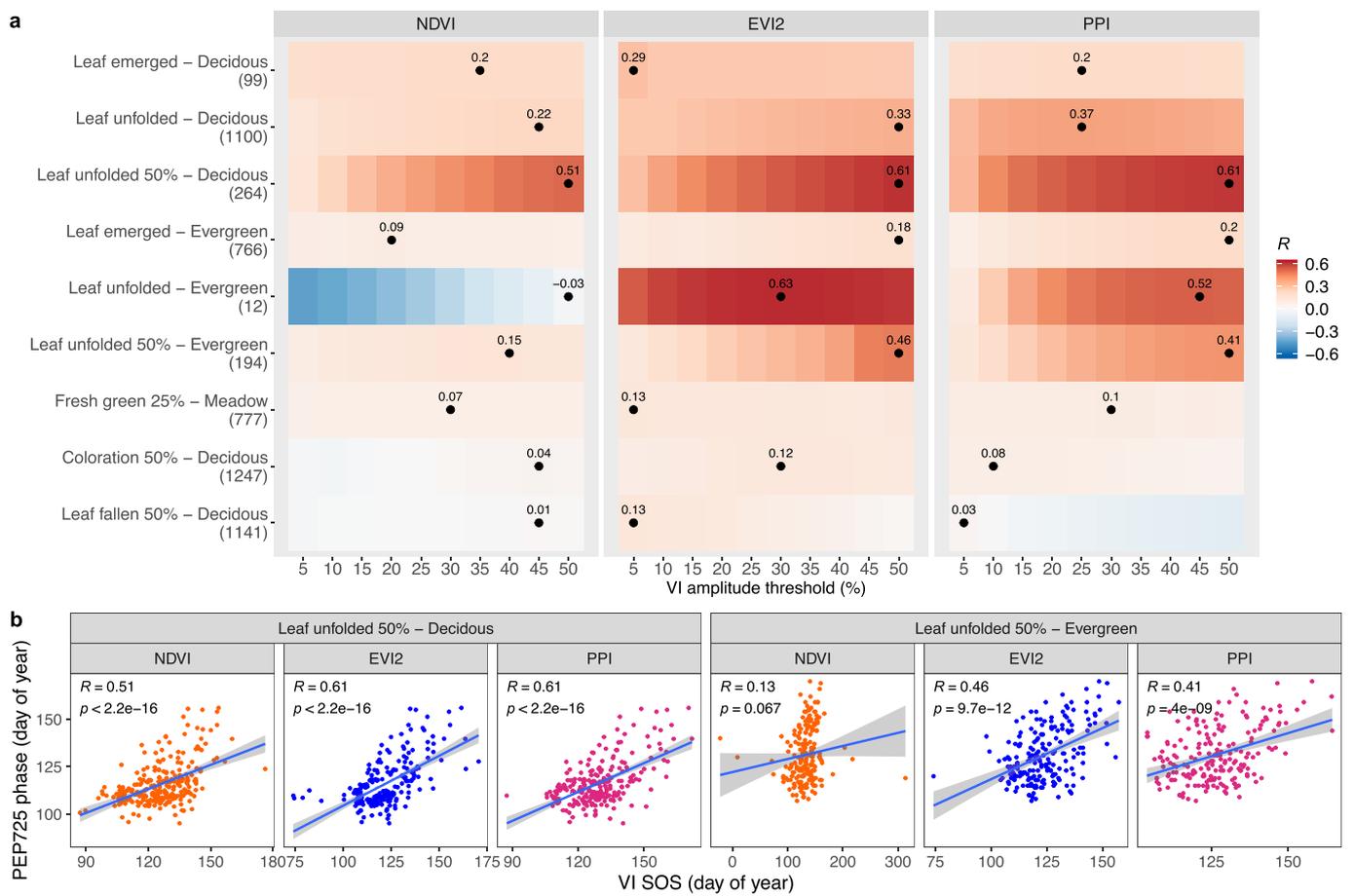
Fig. 8. Spatial correlation between paired PEP725 and VI phases. (a) Heatmap matrix showing the R values at varying VI thresholds. The threshold with the maximum R value for each paired phase is marked by a solid black dot, and the R values are labeled above the dot. The number of sites containing each paired PEP725 and Sentinel-2 phase is labeled on the y-axis. (b) Scatterplots with linear regression lines of the leaf unfolded 50% phase for deciduous broad-leaved trees and evergreen coniferous trees at VI threshold 50% (for illustration purpose).

4. Discussion

4.1. Performance of VIs and thresholds for deriving phenometrics

In this study, we used the spatial correlation analysis primarily as a way of selecting the most suitable VI for Sentinel-2 phenology mapping. For the comparison to GPP and GCC data, we obtained R values >0.8 for the best VIs (Fig. 6), which are higher than those derived from MODIS VIs (<0.7) as reported by Liu and Wu (2020), although from different sites, years, and methods. The higher spatial resolution of Sentinel-2 imagery could contribute to these stronger relationships, although Cai et al. (2021) did not report stronger relationships with eddy covariance-based GPP for Sentinel-2 than for MODIS in Scandinavia. We attribute the stronger relationships with GPP seasonality and phenology parameters in our study to the use of PPI. The remaining unexplained variability between the calibration data and the remotely sensed data could be related to several factors, including the limited ability of satellite signals to pick up phenological variability at ground level, data processing methods, inaccuracy of ground observations, merging of different land cover classes, and the functional relationship (e.g. linearity) between the calibration and the satellite data.

The use of either GPP or GCC as a reference to evaluate the VI performances in tracking vegetation seasonal variations makes a difference due to their different functional meanings (GPP expressing photosynthetic productivity vs. GCC expressing canopy colors). PPI is best aligned with GPP seasonal variations for all the land cover classes studied

(Fig. 5c). This temporal alignment between PPI and GPP results in the highest spatial consistency between their phenometrics (smallest bias and largest R values for both SOS and EOS, Fig. 6). For the GCC trajectories, none of the three VIs shows the highest correlations across all the land cover classes, but EVI2 shows the overall best performance (Fig. 5). Regarding the comparison with GCC phenometrics, all three VIs show comparable correlations for GCC SOS, while biases are lowest for PPI and EVI2. For GCC EOS, NDVI outperforms PPI and EVI2 (Fig. 6). The strong relationship between NDVI and GCC EOS is attributable to their similar sensitivity to leaf pigmentation (Filippa et al., 2018; Luo et al., 2018). PPI is near-linearly related to green LAI by definition (Jin and Eklundh, 2014), thus showing more variability than the other indices during the growing season, and strong relationship with GPP.

The VIs' performances are determined not only by their responses to vegetation changes (e.g., leaf area and leaf color), but also by their sensitivity to background noise (e.g., snow cover and soil color). As seen in the supplementary Fig. S2 and S3, both NDVI and EVI2 showed outliers in the snowy seasons, particularly for the coniferous forest sites (Fig. S2b and S3b), although we applied quality control based on the SCL information. The outliers in snowy seasons can sometimes introduce false identification of growing season, which in the current analysis was eliminated by retaining only one SOS/EOS for each year to facilitate the comparison with ground calibration data. However, these outliers can complicate the operational production of the Sentinel-2 phenology dataset, given the imperfect accuracy of the state-of-the-art masking algorithms for Sentinel-2 imagery (Zekoll et al., 2021). In contrast, PPI is
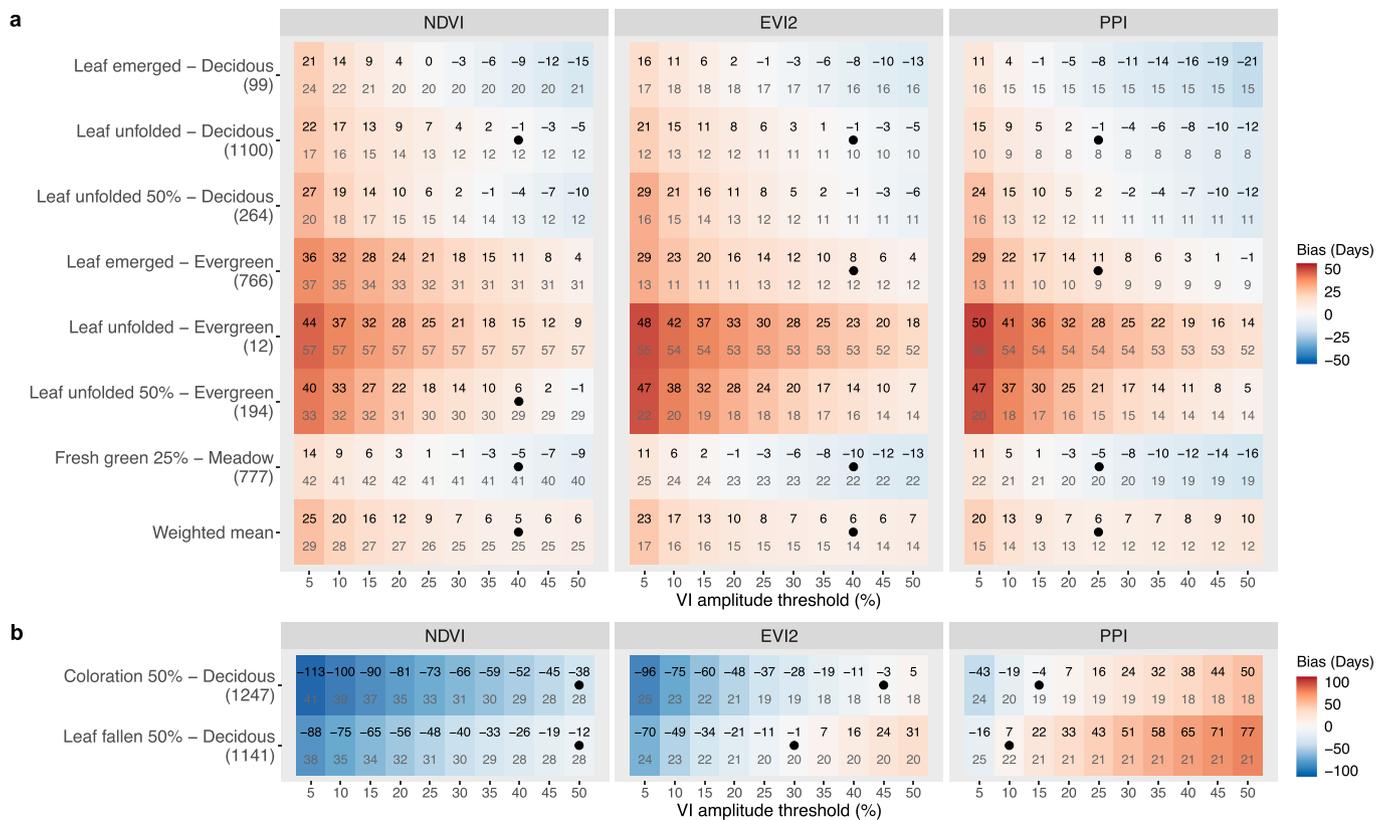
Fig. 9. Statistics of biases and standard deviations between paired PEP725 and VI at (a) spring and (b) autumn phases (Table 2) for different vegetation groups (Table 1) at varying VI amplitude thresholds from 5% to 50%. The mean bias and standard deviation values are labeled as black (upper) and grey (lower) digits respectively for each comparison. The spring phases with the best match for each vegetation group at the optimal threshold according to the weighted mean values is marked by a solid black dot. The number of sites containing each paired PEP725 and Sentinel-2 phase is labeled on the y-axis.

insensitive to snow cover (Supplementary Fig. S2 and S3), which is an advantage for phenology mapping in the northern hemisphere.

The spatial correlations of SOS/EOS derived from VIs and GPP/GCC are very similar across thresholds varying from 5% to 50%, and are relatively high (~ 0.8, Figs. 6 and 7). Therefore, it is acceptable to use a uniform threshold to quantify SOS/EOS across different land cover types in Europe, which is important for operational production of a continental-scale dataset. It also indicates that as long as simple functions are used for modeling the growing seasons, the selection criteria for SOS and EOS (threshold, derivatives, inflection points etc.) generate relatively equal strength of relationships, but mainly affect the biases against phenological phases. Therefore, selecting a threshold that best matches with certain PEP725 phenophases is necessary for providing a phenological meaning to the mathematically defined VI phenometrics.

Although the PEP725 observations lack precise geographical locations and are mainly distributed in Germany, leading to generally weak correlations with ground data, we managed to link the threshold-based VI SOS/EOS with phenological events at the land cover type level. For certain phenophases (e.g., leaf unfolded 50%), we obtained significant spatial correlations between PEP725 and VI phenometrics with R values around 0.5 (Fig. 8), similar to those reported by Kowalski et al. (2020), thus verifying the feasibility of the methodology. Based on the statistics between paired PEP725 and VI phases, we found that the phenological biases were minimized for SOS at NDVI 40%, at EVI2 40%, and at PPI 25% and for EOS at NDVI 50%, at EVI2 40%, and at PPI 15%, with a generally smaller standard deviation (i.e., higher consistency) for PPI than for EVI2 and NDVI (Fig. 9). EVI2 and PPI phenometrics derived from these optimal thresholds are visually consistent with latitudinal and elevational variations along two transects (Fig. 10), suggesting their capability of reflecting spatial variabilities in vegetation phenology.

4.2. PPI calculation and BRDF correction

PPI is capable of clearly reflecting phenology cycles for dense canopies, like in boreal forests, due to the usage of a logarithmic design and sun-zenith-angle-related canopy light extinction efficiency (Jin and Eklundh, 2014). However, the logarithm comes with a cost of being highly sensitive to variability in the differences between *NIR* and *red* reflectance during the peak of the growing season, which can have an impact on the calculation of phenometrics via influencing the seasonal amplitude. This problem can be largely solved by outlier filtering. A disadvantage of PPI as compared to NDVI and EVI2 is its complexity of calculation, particularly the calculation of DVI_{max} for each pixel, which can be time-consuming at the continental-scale. Moreover, the DVI_{max} parameter is in principle a pixel-dependent theoretical value, which can be estimated from multi-year DVI observations (Jin and Eklundh, 2014). Given the currently short time series of Sentinel-2 data, the DVI_{max} parameter needs to be updated when more years of observations are available. Abdi et al. (2019) found that PPI showed a better performance for estimating GPP than other indices for African semi-arid ecosystems. However, PPI values appeared to be less reliable in the sparsely vegetated areas in Tibetan Plateau, where there is a large proportion of bare soil (Huang et al., 2021). This issue is likely attributable to the use of a constant DVI_{soil} value of 0.09, which is an empirical estimation from the soil reflectance spectral database (speclib.jpl.nasa.gov) and the field measurements by Huete et al. (1984). A promising solution to this issue is to introduce a scene-based soil reflectance parameter and is currently under investigation.

Our results show a small difference in the derived SOS/EOS values between Sentinel-2 NBAR and TOC reflectance. Also, it should be noted that the correction of Sentinel-2 BRDF effects based on MODIS
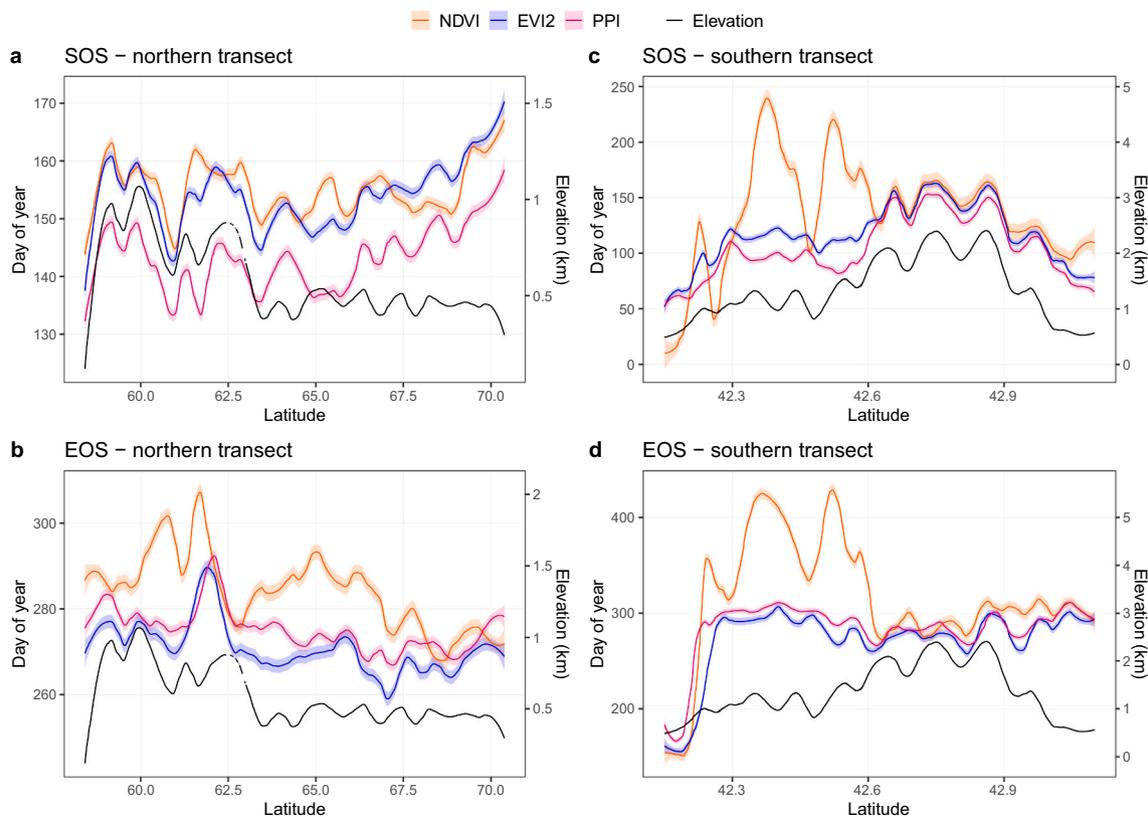
Fig. 10. The latitudinal and elevational variabilities of (a, c) SOS and (b, d) EOS along two transects in (a, b) northern and (b, d) southern Europe (Fig. 2). The amplitude thresholds used are 40% for NDVI/EVI2 SOS, 25% for PPI SOS, 50% for NDVI EOS, 45% for EVI2 EOS, and 15% for PPI EOS. The lines are the latitudinal trends of the per-pixel values using the LOESS smoothing method (span = 0.1) and the shaded areas are the standard errors.

parameters is not free from uncertainties, particularly for heterogeneous areas due to the scale difference between them (10 m vs. 500 m resolutions). Furthermore, using an external dynamic data source for the operational production of Sentinel-2 phenology dataset is impractical, considering the possibility of future deprecation of MODIS and the extra computation load. Therefore, based on our results (Section 3.6), applying BRDF correction to Sentinel-2 is a step that can be excluded for phenology mapping in Europe.

4.3. Implications and perspectives

Landsat was often used in combination with Sentinel-2 to gain more time series observations for phenology mapping, such as the MuSLI dataset for North America (Bolton et al., 2020) and the study by Kowalski et al. (2020) in Germany. While the data combination is beneficial for Landsat-based phenology mapping with long historical records back to the 1980s, it sacrifices the higher spatial resolution (10 m) of Sentinel-2, which may be crucial for certain applications like forest stand species mapping in heterogeneous areas (Grabska et al., 2020). Moreover, synthetic aperture radar (SAR) data from Sentinel-1 was proposed to integrate with Sentinel-2 for crop phenology mapping (Mercier et al., 2020). SAR provides more cloud-free observations and information on vegetation structural changes, which may be particularly helpful for near-real-time crop monitoring. But our preliminary analyses (not shown here) indicate that the SAR time series from Sentinel-1 is highly noisy for evergreen coniferous forests providing little information on phenology. With a proper selection of VI and the use of robust time series fitting methods, our results suggest that Sentinel-2 alone is capable to map vegetation phenology across a range of land cover types in Europe. Finally, with the forthcoming release of the Sentinel-2

HR-VPP dataset it will be of great importance to carry out wall-to-wall comparisons with existing datasets, such as the MODIS phenology product (Friedl et al., 2019).

5. Conclusions

In summary, PPI is shown to be the index with strongest relationships to photosynthetic phenology as represented by GPP, and it has the highest robustness against background noise in snowy conditions. For greenness phenology (as represented by GCC), no clear conclusion can be drawn, since different VIs perform best in different parts of the analysis. However, when considering all the analyses performed in the study, if only one VI is going to be used for Sentinel-2 phenology mapping across Europe we recommend using PPI. The PPI amplitude thresholds with best overall relationships with ground-observed PEP725 phenophases are 25% for SOS and 15% for EOS. It should be noted that all the ground datasets have limitations in geographical extent, and for regional and local applications or for adaptations to specific vegetation types or species, fine-tuning of the thresholds is probably necessary. Based on the results of this study we conclude that in the era of full Sentinel-2 operation with two satellites, we can accurately monitor continental-scale phenology at a 10 m spatial resolution.

CRediT authorship contribution statement

Feng Tian: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Zhanzhang Cai**: Methodology, Software, Formal analysis, Investigation, Funding acquisition. **Hongxiao Jin**: Methodology, Writing - review & editing. **Koen Hufkens**: Resources, Data
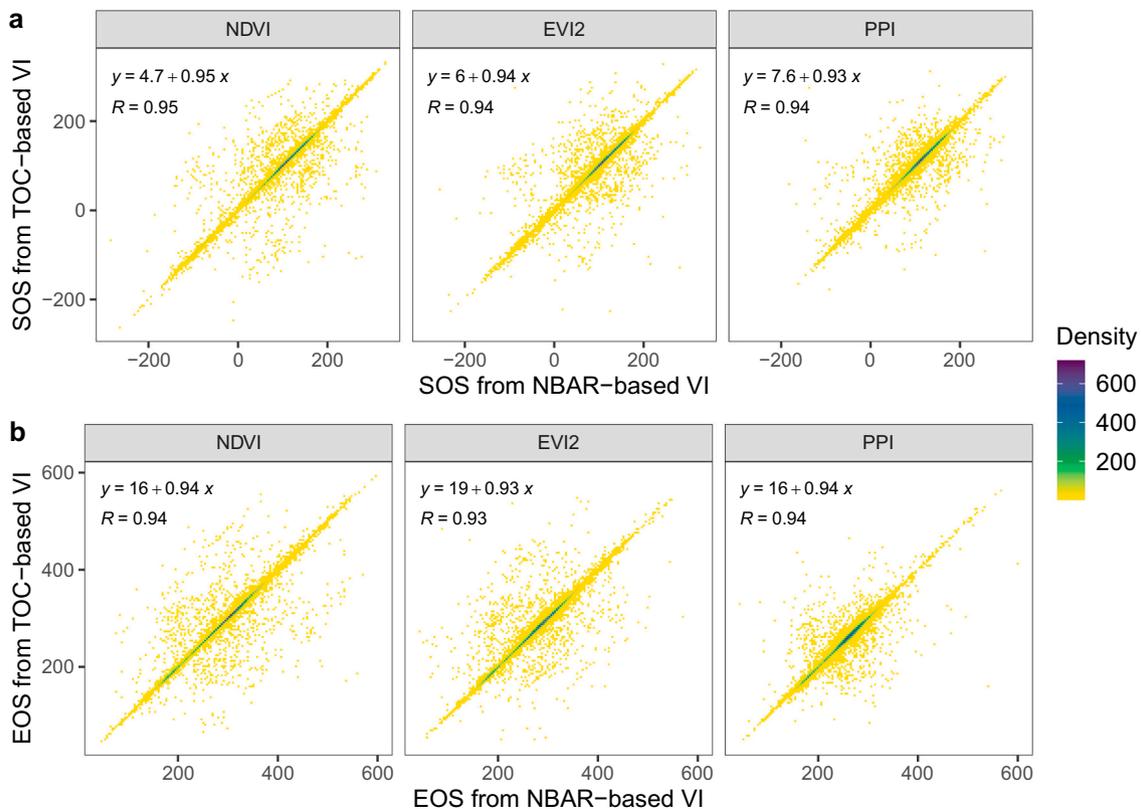
Fig. 11. Comparison of (a) SOS and (b) EOS values at amplitude threshold of 25% between NBAR VI and TOC (without BRDF correction) VI for all the 10,000 randomly selected pixels during 2018–2019.

Declaration of Competing Interest

The authors declare no competing interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112456.

References

Abdi, A.M., Boke-Olén, N., Jin, H., Eklundh, L., Tagesson, T., Lehsten, V., Ardö, J., 2019. First assessment of the plant phenology index (PPI) for estimating gross primary productivity in African semi-arid ecosystems. International Journal of Applied Earth Observation and Geoinformation 78, 249–260. https://doi.org/10.1016/j.jag.2019.01.018.

Archibald, S., Scholes, R.J., 2007. Leaf green-up in a semi-arid African savanna -separating tree and grass responses to environmental cues. J. Veg. Sci. 18, 583–594. https://doi.org/10.1111/j.1654-1103.2007.tb02572.x.

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sens. Environ. 240, 111685. https://doi.org/10.1016/j.rse.2020.111685.

Bórnez, K., Descals, A., Verger, A., Peñuelas, J., 2020. Land surface phenology from VEGETATION and PROBA-V data. Assessment over deciduous forests. Int. J. Appl. Earth Obs. Geoinf. 84 https://doi.org/10.1016/j.jag.2019.101974, 101974.

Brown, M.E., de Beurs, K., Vrieling, A., 2010. The response of African land surface phenology to large scale climate oscillations. Remote Sens. Environ. 114, 2286–2296. https://doi.org/10.1016/j.rse.2010.05.005.

Buitenwerf, R., Rose, L., Higgins, S.I., 2015. Three decades of multi-dimensional change in global leaf phenology. Nat. Clim. Chang. 5, 364–368. https://doi.org/10.1038/nclimate2533.

Cai, Z., 2019. Vegetation Observation in the Big Data Era: Sentinel-2 Data for Mapping the Seasonality of Land Vegetation (Ph.D. thesis). Lund University, Lund, Sweden.

Cai, Z., Jönsson, P., Jin, H., Eklundh, L., 2017. Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data. Remote Sens. 9, 1271. https://doi.org/10.3390/rs9121271.

Cai, Z., Junttila, S., Holst, J., Jin, H., Ardö, J., Ibrom, A., Peichl, M., Mölder, M., Jönsson, P., Rinne, J., Karamihalaki, M., Eklundh, L., 2021. Modelling daily gross primary productivity with Sentinel-2 data in the nordic region–comparison with data from MODIS. Remote Sens. 13, 469. https://doi.org/10.3390/rs13030469.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Cong, N., Wang, T., Nan, H., Ma, Y., Wang, X., Myneni, R.B., Piao, S., 2013. Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multimethod analysis. Glob. Chang. Biol. 19, 881–891. https://doi.org/10.1111/gcb.12077.

Copernicus, 2020. https://land.copernicus.eu/pan-european/biophysical-parameters/high-resolution-vegetation-phenology-and-productivity.

d'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., van der Velde, M., 2020. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series. Remote Sens. Environ. 239 (111660) https://doi.org/10.1016/j.rse.2020.111660.

de Beurs, K.M., Henebry, G.M., 2004. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sens. Environ. 89, 497–509. https://doi.org/10.1016/j.rse.2003.11.006.

de Beurs, K.M., Henebry, G.M., 2010. Spatio-temporal statistical methods for Modelling land surface phenology. In: Hudson, I.L., Keatley, M.R. (Eds.), Phenological Research: Methods for Environmental and Climate Change Analysis. Springer Netherlands, Dordrecht, pp. 177–208. https://doi.org/10.1007/978-90-481-3335-2_9.

Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., Picard, G., 2005. Determination of phenological dates in boreal regions using normalized difference water index. Remote Sens. Environ. 97, 26–38. https://doi.org/10.1016/j.rse.2005.03.011.

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., Vanhellemont, Q., 2018. Atmospheric correction inter-comparison exercise. Remote Sens. 10, 352. https://doi.org/10.3390/rs10020352.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution Mission for GMES operational services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026. The Sentinel Missions - New Opportunities for Science.

Eklundh, L., Jönsson, P., 2015. TIMESAT: A software package for time-series processing and assessment of vegetation dynamics. In: Kuenzer, C., Dech, S., Wagner, W. (Eds.), Remote Sensing Time Series: Revealing Land Surface Dynamics, Remote Sensing and Digital Image Processing. Springer International Publishing, Cham, pp. 141–158. https://doi.org/10.1007/978-3-319-15967-6_7.

Filippa, G., Cremonese, M., Galvagno, M., Sonnentag, O., Humphreys, E., Hufkens, K., Ryu, Y., Verfaillie, J., Morra di Cella, U., Richardson, A.D., 2018. NDVI derived from near-infrared-enabled digital cameras: Applicability across different plant functional types. Agric. For. Meteorol. 249, 275–285. https://doi.org/10.1016/j.agrformet.2017.11.003.

Fisher, J.I., Mustard, J.F., Vadeboncoeur, M.A., 2006. Green leaf phenology at Landsat resolution: scaling from the field to the satellite. Remote Sens. Environ. 100, 265–279. https://doi.org/10.1016/j.rse.2005.10.022.

Friedl, M., Gray, J., Sulla-Menashe, D., 2019. MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD12Q2.006.

Ganguly, S., Friedl, M.A., Tan, B., Zhang, X., Verma, M., 2010. Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. Remote Sens. Environ. 114, 1805–1816. https://doi.org/10.1016/j.rse.2010.04.005.

Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sens. Environ. 251, 112103. https://doi.org/10.1016/j.rse.2020.112103.

Helman, D., 2018. Land surface phenology: what do we really 'see' from space? Sci. Total Environ. 618, 665–673. https://doi.org/10.1016/j.scitotenv.2017.07.237.

Huang, K., Zhang, Y., Tagesson, T., Brandt, M., Wang, L., Chen, N., Zu, J., Jin, H., Cai, Z., Tong, X., Cong, N., Fensholt, R., 2021. The confounding effect of snow cover on assessing spring phenology from space: a new look at trends on the Tibetan Plateau. Sci. Total Environ. 756, 144011. https://doi.org/10.1016/j.scitotenv.2020.144011.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2. The Moderate Resolution Imaging Spectroradiometer (MODIS): A New Generation of Land Surface Monitoring.

Huete, A.R., Post, D.F., Jackson, R.D., 1984. Soil spectral effects on 4-space vegetation discrimination. Remote Sens. Environ. 15, 155–165. https://doi.org/10.1016/0034-4257(84)90043-9.

Hufkens, K., Filippa, G., Cremonese, E., Migliavacca, M., D'Odorico, P., Peichl, M., Gielen, B., Hortnagl, L., Soudani, K., Papale, D.B., Rebmann, C., Brown, T., Wingate, L., 2018. Assimilating phenology datasets automatically across ICOS ecosystem stations. Int. Agrophys. https://doi.org/10.1515/intag-2017-0050.

Jeong, S.-J., Ho, C.-H., Gim, H.-J., Brown, M.E., 2011. Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008. Glob. Chang. Biol. 17, 2385–2399. https://doi.org/10.1111/j.1365-2486.2011.02397.x.

Jin, H., Eklundh, L., 2014. A physically based vegetation index for improved monitoring of plant phenology. Remote Sens. Environ. 152, 512–525. https://doi.org/10.1016/j.rse.2014.07.010.

Jin, H., Jönsson, A.M., Bolmgren, K., Langvall, O., Eklundh, L., 2017. Disentangling remotely-sensed plant phenology and snow seasonality at northern Europe using MODIS and the plant phenology index. Remote Sens. Environ. 198, 203–212. https://doi.org/10.1016/j.rse.2017.06.015.

Jin, H., Jönsson, A.M., Olsson, C., Lindström, J., Jönsson, P., Eklundh, L., 2019. New satellite-based estimates show significant trends in spring phenology and complex

sensitivities to temperature and precipitation at northern European latitudes. Int. J. Biometeorol. 63, 763–775. https://doi.org/10.1007/s00484-019-01690-5.

Jönsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40, 1824–1832. https://doi.org/10.1109/TGRS.2002.802519.

Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845. https://doi.org/10.1016/j.cageo.2004.05.006.

Jönsson, A.M., Eklundh, L., Hellström, M., Bärring, L., Jönsson, P., 2010. Annual changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow dynamics and tree phenology. Remote Sens. Environ. 114, 2719–2730. https://doi.org/10.1016/j.rse.2010.06.005.

Jönsson, P., Cai, Z., Melaas, E., Friedl, M.A., Eklundh, L., 2018. A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data. Remote Sens. 10, 635. https://doi.org/10.3390/rs10040635.

Justice, C.O., Townshend, J.R.G., Holben, B.N., Tucker, C.J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data. Int. J. Remote Sens. 6, 1271–1318. https://doi.org/10.1080/01431168508948281.

Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P., Morisette, J.T., 2002. An overview of MODIS Land data processing and product status. Remote Sens. Environ. 83, 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.

Karkauskaite, P., Tagesson, T., Fensholt, R., 2017. Evaluation of the Plant Phenology Index (PPI), NDVI and EVI for start-of-season trend analysis of the northern hemisphere boreal zone. Remote Sens. 9, 485. https://doi.org/10.3390/rs9050485.

Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M.A., Richardson, A.D., 2014. Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. Biogeosciences 11, 4305–4320. https://doi.org/10.5194/bg-11-4305-2014.

Kowalski, K., Senf, C., Hostert, P., Pflugmacher, D., 2020. Characterizing spring phenology of temperate broadleaf forests using Landsat and Sentinel-2 time series. Int. J. Appl. Earth Obs. Geoinf. 92, 102172. https://doi.org/10.1016/j.jag.2020.102172.

Li, X., Strahler, A.H., 1992. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. IEEE Trans. Geosci. Remote Sens. 30, 276–292. https://doi.org/10.1109/36.134078.

Lieth, H., 1974. Phenology and Seasonality Modeling. Springer-Verlag, New York.

Liu, Y., Wu, C., 2020. Understanding the role of phenology and summer physiology in controlling net ecosystem production: a multiscale comparison of satellite, PhenoCam and eddy covariance data. Environ. Res. Lett. 15, 104086. https://doi.org/10.1088/1748-9326/abb32f.

Liu, R., Shang, R., Liu, Y., Lu, X., 2017. Global evaluation of gap-filling approaches for seasonal NDVI with considering vegetation growth trajectory, protection of key point, noise resistance and curve stability. Remote Sens. Environ. 189, 164–179. https://doi.org/10.1016/j.rse.2016.11.023.

Luo, Y., El-Madany, T.S., Filippa, G., Ma, X., Ahrens, B., Carrara, A., Gonzalez-Cascon, R., Cremonese, E., Galvagno, M., Hammer, T.W., Pacheco-Labrador, J., Martín, M.P., Moreno, G., Perez-Priego, O., Reichstein, M., Richardson, A.D., Römermann, C., Migliavacca, M., 2018. Using Near-Infrared-Enabled Digital Repeat Photography to Track Structural and Physiological Phenology in Mediterranean Tree–Grass Ecosystems. Remote Sens. 10 (8), 1293. https://doi.org/10.3390/rs10081293.

Malingreau, J.-P., 1986. Global vegetation dynamics: satellite observations over Asia. Int. J. Remote Sens. 7, 1121–1146. https://doi.org/10.1080/01431168608948914.

Medlyn, B.E., 1998. Physiological basis of the light use efficiency model. Tree Physiol. 18, 167–176. https://doi.org/10.1093/treephys/18.3.167.

Melaas, E.K., Richardson, A.D., Friedl, M.A., Dragoni, D., Gough, C.M., Herbst, M., Montagnani, L., Moors, E., 2013. Using FLUXNET data to improve models of springtime vegetation activity onset in forest ecosystems. Agric. For. Meteorol. 171–172, 46–56. https://doi.org/10.1016/j.agrformet.2012.11.018.

Melaas, E.K., Sulla-Menashe, D., Gray, J.M., Black, T.A., Morin, T.H., Richardson, A.D., Friedl, M.A., 2016. Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. Remote Sens. Environ. 186, 452–464. https://doi.org/10.1016/j.rse.2016.09.014.

Melaas, E.K., Sulla-Menashe, D., Friedl, M.A., 2018. Multidecadal changes and interannual variation in springtime phenology of North American temperate and boreal deciduous forests. Geophys. Res. Lett. 45, 2679–2687. https://doi.org/10.1002/2017GL076933.

Menzel, A., 2000. Trends in phenological phases in Europe between 1951 and 1996. Int. J. Biometeorol. 44, 76–81. https://doi.org/10.1007/s004840000054.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. ISPRS J. Photogramm. 163, 231–256. https://doi.org/10.1016/j.isprsjprs.2020.03.009.

Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G., Nemani, R.R., 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature 386, 698–702. https://doi.org/10.1038/386698a0.

Norris, J.R., Walker, J.J., 2020. Solar and sensor geometry, not vegetation response, drive satellite NDVI phenology in widespread ecosystems of the western United States. Remote Sens. Environ. 249, 112013. https://doi.org/10.1016/j.rse.2020.112013.

Peng, D., Zhang, X., Wu, C., Huang, W., Gonsamo, A., Huete, A.R., Didan, K., Tan, B., Liu, X., Zhang, B., 2017. Intercomparison and evaluation of spring phenology products using National Phenology Network and AmeriFlux observations in the

contiguous United States. Agric. For. Meteorol. 242, 33–46. https://doi.org/10.1016/j.agrformet.2017.04.009.

Peñuelas, J., Filella, I., 2001. Responses to a warming world. Science 294, 793–795. https://doi.org/10.1126/science.1066860.

Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., Zhu, X., 2019. Plant phenology and global climate change: current progresses and challenges. Glob. Chang. Biol. 25, 1922–1940. https://doi.org/10.1111/gcb.14619.

Richardson, A.D., Jenkins, J.P., Braswell, B.H., Hollinger, D.Y., Ollinger, S.V., Smith, M.-L., 2007. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. Oecologia 152, 323–334. https://doi.org/10.1007/s00442-006-0657-z.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. 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.

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., Frolking, S., 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. Sci. Data 5, 180028. https://doi.org/10.1038/sdata.2018.28.

Ross, J., 1981. The Radiation Regime and Architecture of Plant Stands. Artech House, MA, USA.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112, 3112–3130. https://doi.org/10.1016/j.rse.2008.03.009.

Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sens. Environ. 185, 57–70. https://doi.org/10.1016/j.rse.2015.12.024. Landsat 8 Science Results.

Roy, D.P., Li, J., Zhang, H.K., Yan, L., Huang, H., Li, Z., 2017. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. Remote Sens. Environ. 199, 25–38. https://doi.org/10.1016/j.rse.2017.06.019.

Scheifinger, H., Menzel, A., Koch, E., Peter, Ch., 2003. Trends of spring time frost events and phenological dates in Central Europe. Theor. Appl. Climatol. 74, 41–51. https://doi.org/10.1007/s00704-002-0704-6.

Schwartz, M.D., Reed, B.C., 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. Int. J. Remote Sens. 20, 3451–3457. https://doi.org/10.1080/014311699211499.

Singh, R.K., Svystun, T., AlDahmash, B., Jönsson, A.M., Bhalerao, R.P., 2017. Photoperiod- and temperature-mediated control of phenology in trees – a molecular perspective. New Phytol. 213, 511–524. https://doi.org/10.1111/nph.14346.

Tagesson, T., Fensholt, R., Huber, S., Horion, S., Guiro, I., Ehammer, A., Ardö, J., 2015. Deriving seasonal dynamics in ecosystem properties of semi-arid savannas using in situ based hyperspectral reflectance. Biogeosciences 12, 4621–4635. https://doi.org/10.5194/bg-12-4621-2015.

Templ, B., Koch, E., Bolmgren, K., Ungersböck, M., Paul, A., Scheifinger, H., Rutishauser, T., Busto, M., Chmielewski, F.-M., Hájková, L., Hodzić, S., Kaspar, F., Pietragalla, B., Romero-Fresneda, R., Tolvanen, A., Vučetič, V., Zimmermann, K., Zust, A., 2018. Pan European Phenological database (PEP725): a single point of access for European data. Int. J. Biometeorol. 62, 1109–1113. https://doi.org/10.1007/s00484-018-1512-8.

Tong, X., Tian, F., Brandt, M., Liu, Y., Zhang, W., Fensholt, R., 2019. Trends of land surface phenology derived from passive microwave and optical remote sensing systems and associated drivers across the dry tropics 1992–2012. Remote Sens. Environ. 232, 111307. https://doi.org/10.1016/j.rse.2019.111307.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang, T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M., 2018. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. Remote Sens. Environ. 215, 517–529. https://doi.org/10.1016/j.rse.2018.03.014.

White, M.A., Nemani, R.R., 2006. Real-time monitoring and short-term forecasting of land surface phenology. Remote Sens. Environ. 104, 43–49. https://doi.org/10.1016/j.rse.2006.04.014.

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

White, M.A., Beurs, K.M.D., Didan, K., Inouye, D.W., Richardson, A.D., Jensen, O.P., O'keefe, J., Zhang, G., Nemani, R.R., Leeuwen, W.J.D.V., Brown, J.F., Wit, A.D., Schaepman, M., Lin, X., Dettinger, M., Bailey, A.S., Kimball, J., Schwartz, M.D., Baldocchi, D.D., Lee, J.T., Lauenroth, W.K., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. Glob. Chang. Biol. 15, 2335–2359. https://doi.org/10.1111/j.1365-2486.2009.01910.x.

Wu, C., Wang, X., Wang, H., Ciais, P., Peñuelas, J., Myneni, R.B., Desai, A.R., Gough, C.M., Gonsamo, A., Black, A.T., Jassal, R.S., Ju, W., Yuan, W., Fu, Y., Shen, M., Li, S., Liu, R., Chen, J.M., Ge, Q., 2018. Contrasting responses of autumn-leaf senescence to daytime and night-time warming. Nat. Clim. Chang. 8, 1092–1096. https://doi.org/10.1038/s41558-018-0346-z.

Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. Vol. 185, 271–283. https://doi.org/10.1016/j.rse.2015.11.032. Landsat 8 Science Results.

Zekoll, V., Main-Knorn, M., Alonso, K., Louis, J., Frantz, D., Richter, R., Pflug, B., 2021. Comparison of masking algorithms for Sentinel-2 imagery. Remote Sens. 13, 137. https://doi.org/10.3390/rs13010137.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., 2004. Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. Glob. Chang. Biol. 10, 1133–1145. https://doi.org/10.1111/j.1529-8817.2003.00784.x.

Zhang, X., Wang, J., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y., Jayavelu, S., Gray, J., Liu, L., Yan, D., Henebry, G.M., 2017. Exploration of scaling effects on coarse resolution land surface phenology. Remote Sens. Environ. 190, 318–330. https://doi.org/10.1016/j.rse.2017.01.001.

Zhang, X., Jayavelu, S., Liu, L., Friedl, M.A., Henebry, G.M., Liu, Y., Schaaf, C.B., Richardson, A.D., Gray, J., 2018a. Evaluation of land surface phenology from VIIRS data using time series of PhenoCam imagery. Agric. For. Meteorol. 256–257, 137–149. https://doi.org/10.1016/j.agrformet.2018.03.003.

Zhang, X., Liu, L., Liu, Y., Jayavelu, S., Wang, J., Moon, M., Henebry, G.M., Friedl, M.A., Schaaf, C.B., 2018b. Generation and evaluation of the VIIRS land surface phenology product. Remote Sens. Environ. 216, 212–229. https://doi.org/10.1016/j.rse.2018.06.047.