

# Characterizing spring phenology of temperate broadleaf forests using Landsat and Sentinel-2 time series

Katja Kowalski<sup>a,\*</sup>, Cornelius Senf<sup>cb</sup>, Patrick Hostert<sup>a,c</sup>, Dirk Pflugmacher<sup>a</sup>

<sup>a</sup> Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

<sup>b</sup> Ecosystem Dynamics and Forest Management Group, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany

<sup>c</sup> Integrative Research Institute of Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

## ARTICLE INFO

### Keywords:

Land surface phenology  
Time series  
Logistic function  
Splines  
Data density  
Observation frequency

## ABSTRACT

Vegetation phenology has a great impact on land-atmosphere interactions like carbon cycling, albedo, and water and energy exchanges. To understand and predict these critical land-atmosphere feedbacks, it is crucial to measure and quantify phenological responses to climate variability, and ultimately climate change. Coarse-resolution sensors such as MODIS and AVHRR have been useful to study vegetation phenology from regional to global scales. These sensors are, however, not capable of discerning phenological variation at moderate spatial scales. By offering increased observation density and higher spatial resolution, the combination of Landsat and Sentinel-2 time series might provide the opportunity to overcome this limitation.

In this study, we analyzed the potential of combined Sentinel-2 and Landsat time series for estimating start of season (SOS) of broadleaf forests across Germany for the year 2018. We tested two common statistical modeling approaches (logistic and generalized additive models using thin plate splines) and the two most commonly used vegetation indices, the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI).

We found strong agreement between SOS estimates from logistic and spline models ( $r_{EVI} = 0.86$ ;  $r_{NDVI} = 0.65$ ), whereas agreement was higher for EVI than for NDVI ( $RMSE_{EVI} = 3.07$ ,  $RMSE_{NDVI} = 5.26$  days). The choice of vegetation index thus had a higher impact on the results than the fitting method. The EVI-based SOS also showed higher correlation with ground observations compared to NDVI ( $r_{EVI} = 0.51$ ,  $r_{NDVI} = 0.42$ ). Data density played an important role in estimating land surface phenology. Models combining Sentinel-2A/B, with an average cloud-free observation frequency of 12 days, were largely consistent with the combined Landsat and Sentinel-2 models, suggesting that Sentinel-2A/B may be sufficient to capture SOS for most areas in Germany in 2018. However, in non-overlapping swath areas and mountain areas, observation frequency was significantly lower, underlining the need to combine Landsat and Sentinel-2 for consistent SOS estimates over large areas. Our study demonstrates that estimating SOS of temperate broadleaf forests at medium spatial resolution has become feasible with combined Landsat and Sentinel-2 time series.

## 1. Introduction

Vegetation phenology, the timing of seasonal plant development through phases of active growth and dormancy (Hänninen, 1990), is a widely used indicator of climate change impacts on terrestrial ecosystems (Parmesan and Yohe, 2003). Climate change is expected to alter ecosystem functions and processes globally, including spatial shifts of species ranges and temporal shifts of phenological events (Jeong et al., 2011; Rosenzweig et al., 2008). It is therefore important to improve our understanding of current phenological processes to further assess the

response of ecosystems to climatic changes.

Remote sensing has become an essential tool for studying phenology at various spatial scales. Coarse-resolution data from the Moderate-resolution Imaging Spectroradiometer (MODIS; e.g. Zhang et al., 2003) and the Advanced Very High Resolution Radiometer (AVHRR; e.g. White et al., 2009) have been used for mapping phenological dynamics of forests globally. However, their spatial resolution is not sufficient to discern phenological patterns in heterogeneous landscapes (Fisher and Mustard, 2007; Hufkens et al., 2012b). Conversely, medium resolution sensors have been lacking the temporal resolution to accurately

\* Corresponding author.

E-mail address: [katja.kowalski@geo.hu-berlin.de](mailto:katja.kowalski@geo.hu-berlin.de) (K. Kowalski).

<https://doi.org/10.1016/j.jag.2020.102172>

Received 11 November 2019; Received in revised form 3 April 2020; Accepted 2 June 2020

Available online 11 June 2020

0303-2434/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

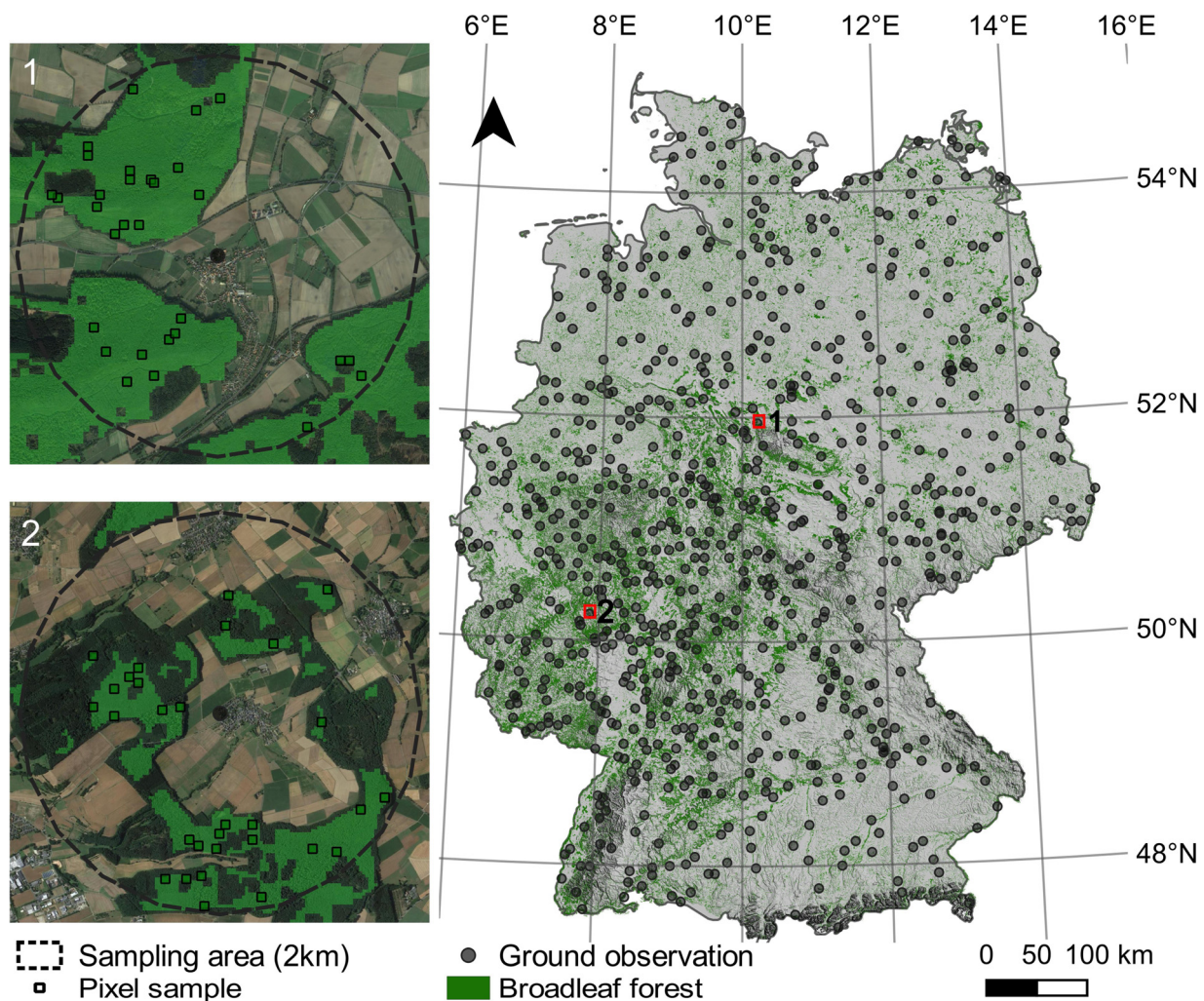


Fig. 1. Study area of Germany with phenological ground observations (right; background map: SRTM digital elevation model; NASA JPL, 2013). Examples of the sampling design with pixel samples in a 2 km buffer zone around each ground observation (left; background map: ©2020 Google, GeoBasis-De/BKG, GeoContent, Maxar Technologies).

characterize vegetation phenology. To circumvent this issue, studies have estimated inter-annual variations of spring phenology by pooling several decades of Landsat data (Melaas et al., 2013; Senf et al., 2017). To obtain annual SOS from Landsat data, Melaas et al. (2018) made use of increased observation densities in areas where acquisition paths overlapped. However, this method does not allow for spatially continuous phenology estimations.

Estimating spring phenology over shorter time periods has only recently become feasible with the launch of the two Sentinel-2 satellites in 2013 and 2017. Landsat 8 and Sentinel-2 combined offer a global average observation interval of 2.9 days, disregarding cloud coverage (Li and Roy, 2017), which substantially increases the potential for capturing annual phenology at the landscape scale. Underlining this, a recent study in the U.S. by Bolton et al. (2020) showed the increased value of phenological estimates from combined Landsat 8 and Sentinel-2 time series for resolving landscape-scale phenological patterns compared to MODIS-based estimates.

A variety of different methods have been developed for estimating phenological parameters from satellite time series. A common approach is to fit a single (global) function to a vegetation index time series, and subsequently use the fitted model parameters to describe the phenological response (Verma et al., 2016). For example, logistic functions have been widely used for deriving spring and autumn phenology from regional to global scales (Cai et al., 2017; Verma et al., 2016; Zhang et al., 2003). Asymmetric signals are, however, poorly captured by logistic

functions (Verma et al., 2016). Alternatively, local smoothing methods such as spline-based methods are more flexible to capture various phenological shapes, but they are more sensitive to noise and observation gaps (Buitenwerf et al., 2015; Melaas et al., 2016). Since the performance of a method is intrinsically linked to the type and quality of the vegetation index time series, it seems logical to also test the most common methods and vegetation indices on combined Landsat and Sentinel-2 time series.

The objective of this study was to assess the potential of combined Landsat and Sentinel-2 time series for characterizing spring phenology in temperate broadleaf forests. Specifically, we addressed the following research questions:

- 1) How do start of season (SOS) estimates from Landsat and Sentinel-2 time series vary with commonly used models and vegetation indices? To address this question, we assessed the applicability of the frequently used logistic model and an approach based on generalized additive models using thin plate regression splines to derive SOS from the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI).
- 2) How do SOS estimates from Landsat and Sentinel-2 compare to ground observations? To address the challenge of validating satellite-based land surface phenology, we evaluated our estimates against ground observations of leaf unfolding.
- 3) How do SOS estimates from combined Landsat and Sentinel-2 time

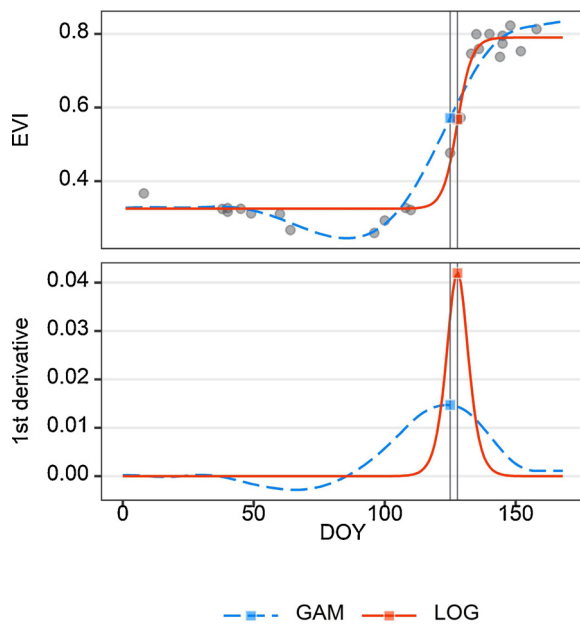


Fig. 2. Example of start of season (SOS) estimates (squares and vertical lines) from Generalized Additive Model (GAM) and logistic model (LOG) using EVI time series (top). SOS was defined as the point in time where the slope of the model fit is highest (bottom).

series compare to SOS estimates from single-sensor time series? To address this question, we estimated SOS from Landsat and Sentinel-2 time series separately and compared these estimates to SOS from combined time series.

## 2. Methods

### 2.1. Study area

We selected Germany as our study site as it covers a wide climate gradient within the temperate forest biome, and because it is one of the European countries with the largest number of phenological observations (Templ et al., 2018). About one third of Germany's total land area is covered by forests, of which nearly 42 % are broadleaf forests (Federal Ministry of Food and Agriculture, 2018). Germany lies in the temperate forest biome, where the annual cycle of plant phenology is mainly driven by temperature. The eastern part of Germany is characterized by a continental climate with high temperature extremes (average annual summer temperature of 18.2 °C, average annual winter temperature of 0.7 °C; meteorological station: Lübben, 51°55'36"N, 13°52'47"E). The western part is under oceanic influence, with milder winters and summers (average annual summer temperature of 17.6 °C, average annual winter temperature of 2.6 °C; meteorological station: Werl, 51°34'35"N, 7°53'16"E; DWD Climate Data Center (CDC), 2019). In mountainous regions and towards the Alps, temperatures are lower, varying with elevation (Zöller et al., 2017).

### 2.2. Sampling design

Our analysis is based on a large sample of Landsat and Sentinel-2 pixels close to phenological ground observations (Fig. 1). Because it is not feasible to directly match a single forest pixel to ground-based observations of leaf unfolding, we chose a two-stage sampling design to capture the phenological variability in the vicinity of the ground observations. The primary sampling units of the first sampling stage were the locations of the ground observations (*plots*) and a 2 km buffer zone. For each plot, up to six species-specific phenological observations were recorded. The exact location of each individual observation is unknown

but restricted to a distance of 2 km from the reported center location of the plot (German Weather Service, 2018). Within each 2 km buffer, we masked all non-broadleaf forest pixels using the Landsat-based land cover classification by Pflugmacher et al. (2019), and further masked pixels disturbed within the last 32 years using the forest disturbance map by Pflugmacher et al. (2020). Finally, we imposed a minimum mapping unit of 11 contiguous pixels, i.e. an area of 9900 m<sup>2</sup>, to the remaining broadleaf forest pixels. In the second sampling stage, we selected thirty overlapping Landsat and Sentinel-2 pixels within each buffer (Fig. 1). Overall, we selected 767 plots (primary sampling units) and 23,010 pixels (secondary sampling units).

### 2.3. Landsat and Sentinel-2 data processing

We used all available Landsat 7, Landsat 8, and Sentinel-2A/B images acquired in 2018. We downloaded the Landsat Tier-1 Collection 1 data from the USGS archive, and the Sentinel-2 Level-1C data from the Copernicus Open Access Hub. To build harmonized time series from different sensors, we used the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE v.2.0; Frantz, 2019). In FORCE, the atmospheric correction is based on radiative transfer theory using a combined image-, database- and object-based estimation of aerosol optical depth (Tanre et al., 1979). The MODIS precipitable water product is used to characterize gaseous absorption (Gao and Kaufman, 2003). For the topographic correction, FORCE uses a modified image-based C-correction with a 1-arc-second digital elevation model (DEM) derived from the Shuttle Radar Topography Mission (SRTM; NASA JPL, 2013). To minimize the effects of sensor differences, we adjusted the blue, red and near-infrared spectral bands of Landsat 7 and 8 to match Sentinel-2 bands using reduced major axis regression (RMA; Smith, 2009). RMA models were built using a subset of pixel samples where data pairs from Landsat 7/8 and Sentinel-2A/B were acquired on the same day during 2016-2019. Model parameters (slope and intercept) were used to adjust the blue, red and near-infrared bands of Landsat 7/8. For each sample, we calculated EVI (Liu and Huete, 1995) and NDVI (Tucker, 1979) time series. We excluded all observations flagged as cloud, cloud shadow, or snow in the FORCE pixel quality layer.

### 2.4. Landsat and Sentinel-2 phenological models

We fitted two models for each of the 23,010 pixel samples: (i) a parametric logistic model (LOG), and (ii) a Generalized Additive Model (GAM) using thin plate splines (Fig. 2). Following Melaas et al. (2013), we used a logistic function of the form:

$$VI(t) = v_{min} + \left( \frac{v_{max}}{1 + e^{-g(t-s)}} \right) \quad (1)$$

where  $VI$  is the vegetation index (NDVI or EVI) at day of year (DOY)  $t$ .  $v_{min}$  and  $v_{max}$  are the minimum and maximum of the function, respectively. The parameter  $g$  is the rate of change (i.e. vegetation green-up rate), and  $s$  is the location of the inflection point. Following previous studies (Melaas et al., 2013; Zhang et al., 2003), we used the parameter  $s$  as SOS estimate (Fig. 2). SOS was defined as the point in time where the modeled change rate of the vegetation index was highest (Verma et al., 2016). The logistic model was fitted using the *nls.multstart* package in R (Padfield and Matheson, 2018; R Core Team, 2017). The implemented algorithm searches combinations of starting parameters and uses the Levenberg-Marquardt algorithm to identify the best model based on the Akaike Information Criterion (AIC). As initial search values, we set  $v_{min}$  and  $v_{max}$  to the minimum and maximum of the respective vegetation index time series and the initial SOS value  $s$  to the mean of all DOYs of observations where the vegetation index value was larger than the median. For the rate of change  $g$ , initializing models based on a range of starting parameters (intervals of 0.1 between 0 and 1) increased convergence of the models considerably.

The GAM was fitted using the *mgcv* package in R (Wood, 2003). The generic formula of a one-dimensional GAM is:

$$VI(t) = \beta_0 + s(t) + \varepsilon_t \quad (2)$$

where  $VI$  is the observed vegetation index value at time  $t$ ,  $\beta_0$  is the intercept of the function at  $t = 0$ ,  $s$  represents a smoothing function of covariate  $t$ , in this case the DOY, and  $\varepsilon_t$  is a random error term with  $\varepsilon_t \sim N(0, \sigma^2)$ . The smoothing function is defined by

$$s(t) = \sum_{k=1}^K \beta_k b_k(t) \quad (3)$$

where the final smoothing function  $s(t)$  is the sum of all  $K$  basis functions  $b_k(t)$  multiplied by its corresponding weight  $\beta_k$ . We used thin plate regression splines as smoothing function, following recommendations by Wood (2003). Like the logistic function, the SOS of the GAM is derived by finding the DOY where the slope (i.e. the rate of change) of the function is highest (Fig. 2). The rate of change of  $VI(t)$  at a given point  $P(t|VI)$  is found by approximating the first derivative using finite differences. To decrease a potential negative bias of observations influenced by unfavorable atmospheric conditions, we iteratively assigned lower weights to observations lying below the previous model fit (Fig. 3). Weights were assigned as the quotient of observed and predicted EVI or NDVI value, raised to the power 4. All weights larger than 1, i.e., lying above the previous model fit, were set to 1. We found that 3 iterations removed the effects of outliers sufficiently (Fig. 3).

We fitted both models to NDVI and EVI time series between January 1<sup>st</sup> and the time of maximum vegetation greenness in summer. We included additional observations within 20 days following the vegetation index maximum or, if no observation followed within 20 days, the next observation to account for variation in the data and to improve model fit. To minimize the effect of outliers, which can be caused by, e.g., snow cover during winter months (Beck et al., 2006), we calculated the mean base value of observations between DOY 1 to DOY 50 and replaced all values during that time with the base value.

We obtained four SOS estimates for each sample (i.e.,  $GAM_{NDVI}$ ,  $GAM_{EVI}$ ,  $LOG_{NDVI}$ , and  $LOG_{EVI}$ ) by estimating SOS with both models

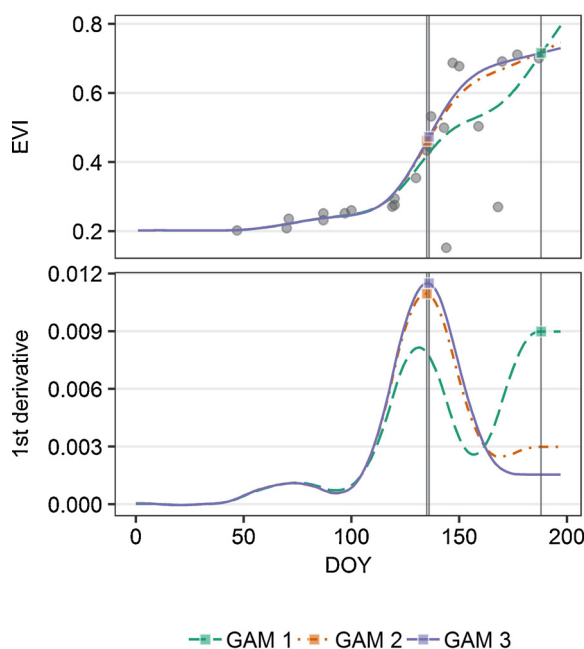


Fig. 3. Example of the iterative fitting process for Generalized Additive Models (GAM) using EVI time series. To reduce the effect of outliers on SOS estimates (GAM 1), GAMs were iteratively fitted to the upper envelope (GAM 2, GAM 3). Squares and vertical lines mark SOS.

and indices. In the same manner, we estimated SOS from single-sensor time series of Landsat and Sentinel-2, resulting in 8 additional estimates per sample. We calculated the mean difference between SOS estimates (Bias), the root-mean-square deviation (RMSD), and the Pearson correlation coefficient to evaluate the agreement between model results.

### 2.5. Evaluation of SOS estimates

For evaluating the satellite-based SOS estimates, we used phenological observations of the main broadleaf tree species (*Aesculus hippocastanum*, *Alnus glutinosa*, *Betula pendula*, *Fagus sylvatica*, *Fraxinus excelsior*, and *Quercus robur*) in Germany as in-situ reference. Data were provided by the Pan European Phenology project (Templ et al., 2018). We derived the average date of leaf unfolding (BBCH-code = 11; Meier, 2018) from the phenological observation (Fig. 1). In total, we obtained ground-based SOS estimates for 767 plots. For comparing SOS estimates from Landsat and Sentinel-2 time series to the phenological ground observations, we averaged all 30 pixel samples for each plot.

## 3. Results

### 3.1. Effect of model and vegetation index on SOS estimates

We found a moderate to strong correlation between SOS from logistic models and GAMs ( $r_{EVI} = 0.86$  and  $r_{NDVI} = 0.65$ ; Fig. 4). Estimates from different models were highly consistent using EVI, whereas NDVI-based SOS estimates varied more strongly with the choice of model. Overall, the choice of vegetation index had the strongest impact on SOS estimates. SOS from NDVI and EVI correlated moderately ( $r_{GAM} = 0.65$ ,  $r_{LOG} = 0.7$ ) but they were less consistent than the EVI-based estimates from different models as indicated by the higher RMSD and Bias (Fig. 4). Notably, NDVI-based models estimated SOS on average 3 days earlier than EVI-based models.

The majority of SOS estimates (0.05<sup>th</sup> – 0.95<sup>th</sup> quantile) ranged between DOY 98 and DOY 131, corresponding to SOS estimates in the time period from the beginning of April to mid-May. Spatially, SOS estimates varied along a North-South gradient with later SOS estimates in northern than in central and southern regions of the study area (Fig. 5). In regions associated with lower temperatures such as mountain ranges, SOS tended to be later than in surrounding areas.

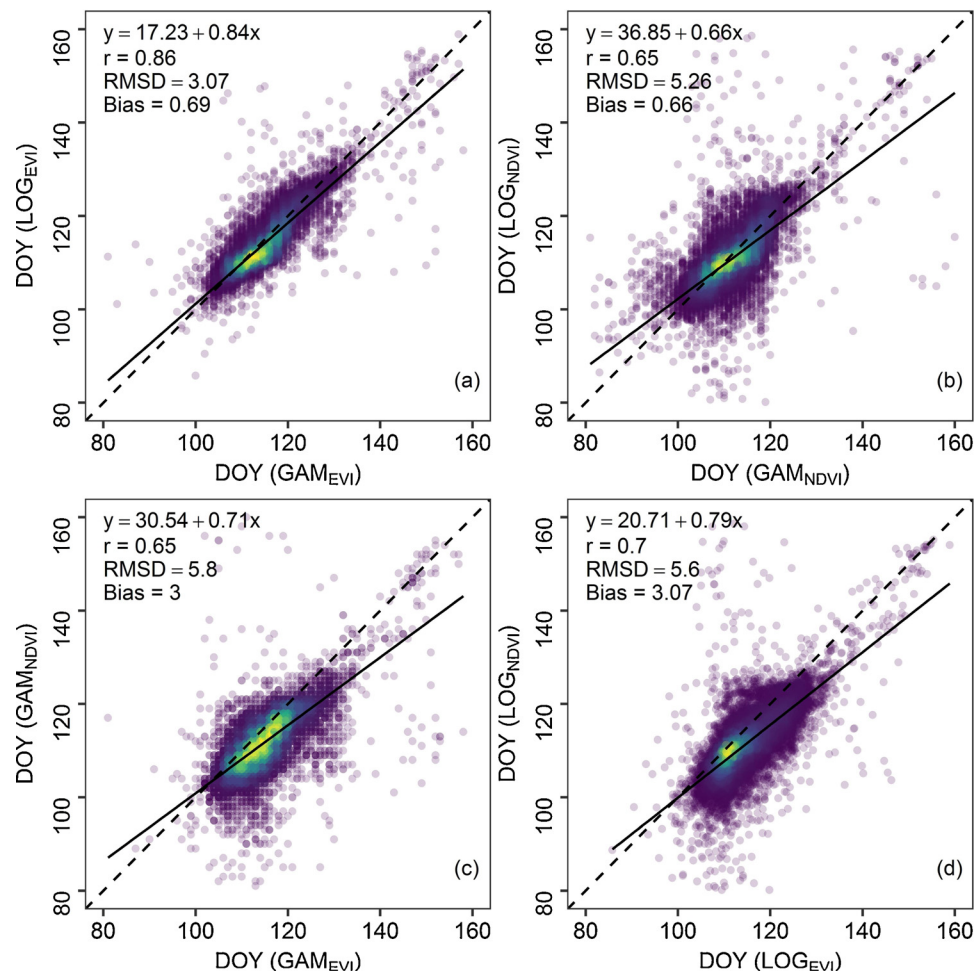
### 3.2. Evaluation of phenological estimates

Model choice influenced SOS estimates less than the two vegetation indices. We therefore report results for evaluation of SOS estimates from one model (GAMs).

The correlation between satellite-based SOS estimates and ground-based observations of leaf unfolding was moderate (Fig. 6). EVI-based SOS estimates showed the highest correlations with ground observations ( $r = 0.51$ ). In comparison, NDVI-based SOS estimates showed a weaker correlation with ground observations ( $r = 0.42$ ). SOS estimates from ground observation were on average 6 days earlier than the EVI models and 3 days earlier than the NDVI models. After removing the systematic bias, the average difference (RMSD) between ground observations and satellite observations was 4.51 days for the EVI model and 4.84 days for the NDVI models, suggesting that EVI models were slightly more consistent with ground observations than NDVI models.

### 3.3. Comparison of single-sensor and multi-sensor SOS estimates

SOS estimates using only Sentinel-2 data were moderately to strongly correlated with SOS estimates from combined time series ( $r_{EVI} = 0.81$ ,  $r_{NDVI} = 0.58$ ). Sentinel-2 time series had an average observation frequency of one cloud-free observation every 12 days, whereas the combined Landsat and Sentinel-2 time series had an average observation frequency of 8 days. From Landsat data, one cloud-



**Fig. 4.** Comparison of start of season (SOS) estimates from different models with (a) EVI and (b) NDVI time series, and SOS estimates from different vegetation indices with (c) Generalized Additive Model (GAM) and (d) logistic model (LOG). Dashed lines represent the 1:1 lines and solid lines are the regression lines.

free observation was available every 21 days on average, which substantially diminished the correlation with the combined time series estimates ( $r_{\text{EVI}} = 0.37$ ,  $r_{\text{NDVI}} = 0.23$ ). In line with these results, the comparison of correlation coefficients, where SOS estimates were grouped by observation frequency, revealed that the agreement between SOS estimates weakened as observation frequency decreased (Fig. 7). For Sentinel-2, correlations to SOS estimates from combined time series were strong ( $r > 0.9$ ) for observation frequencies higher than 10 days. The highest observation frequencies of 5 days led to the most consistent SOS estimates.

SOS estimates from Landsat were, on average, 3 days later than estimates from combined time series. Sentinel-2-based SOS estimates were unbiased for EVI time series and 1 day earlier for NDVI time series. Comparing aggregated SOS estimates from either Landsat or Sentinel-2 to ground observations of leaf unfolding, we found weak to moderate correlations ( $r = 0.15\text{--}0.47$ ). SOS estimates from Landsat showed weaker correlations ( $r = 0.15\text{--}0.24$ ) than Sentinel-2-based estimates ( $r = 0.37\text{--}0.47$ ). SOS estimates from Sentinel-2 and estimates from combined time series had nearly identical bias and RMSD when comparing them to ground observations (Bias = 2.8–6.23 days; RMSD = 6.5–7.86 days). For estimates based on Landsat time series, both measures were consistently higher (Bias = 6.52–9.01 days; RMSD = 10.29–11.05 days).

The results indicate that the high observation frequency of Sentinel-2 may be sufficient to estimate SOS from EVI time series in Germany for 2018. However, there are several areas that seem to benefit from adding Landsat observations. For example, observation frequency was

considerably lower outside swath overlap areas of Sentinel-2 and in mountain ranges in central and southern Germany. In the latter case, this is related to increased cloud cover which resulted in observation frequencies of less than 15 days for the study period from January to June (Fig. 8).

## 4. Discussion

### 4.1. Estimating spring phenology from Landsat and Sentinel-2 time series

We evaluated two models (logistic models and GAMs using thin plate regression splines) and vegetation indices (EVI and NDVI) to estimate spring phenology of temperate broadleaf forests from combined Landsat and Sentinel-2 time series. The results show that the logistic model and the GAM produced comparable SOS estimates, and that the choice of vegetation index had a higher influence on SOS estimates than the choice of model. The finding that two very different modeling approaches, with different sensitivities to data gaps and noise, come to the same conclusion speaks for the robustness of SOS estimates from combined Landsat and Sentinel-2 data. In other words, the choice of model becomes less important with very dense time series, at least in temperate forests.

The comparison of EVI and NDVI time series provides a useful frame of reference for the comparison of the model types, but it also confirms other studies that the choice of vegetation index matters. Using MODIS data, Hufkens et al. (2012a) demonstrated the impact of vegetation indices for retrieving phenological parameters from remote sensing

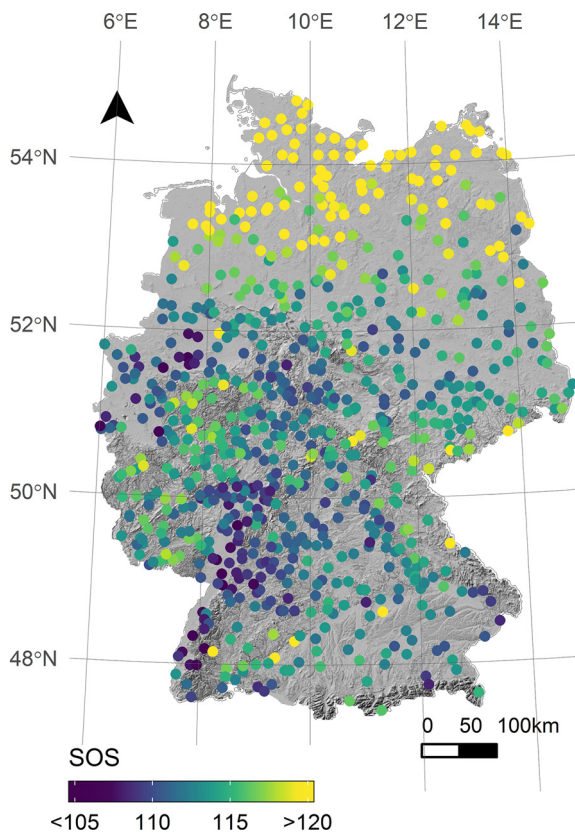


Fig. 5. Start of season (SOS) dates of 2018 estimated using generalized additive models with EVI time series at phenological observations plots.

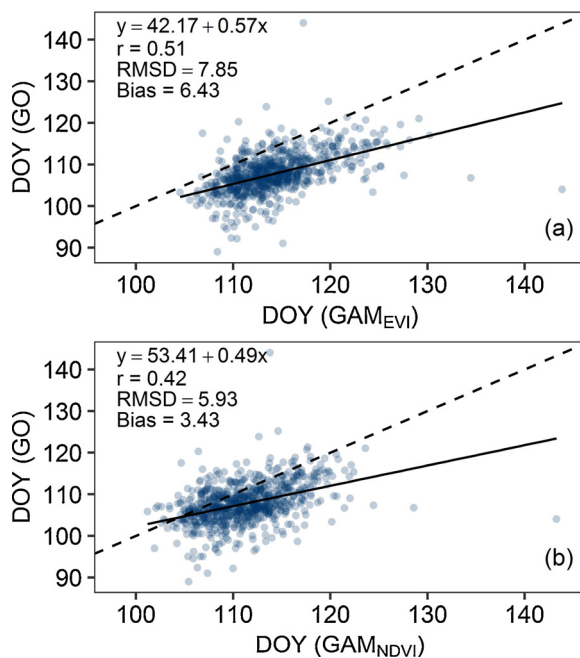


Fig. 6. Start of season (SOS) from Generalized Additive Model (GAM) using (a) EVI and (b) NDVI time series compared to SOS from ground observations (GO; see Fig. S1 for the comparison of satellite-based SOS to GO of single species). Dashed lines represent the 1:1 lines and solid lines are the regression lines.

time series. A great majority of studies relied on NDVI for deriving spring phenology (e.g. Beck et al., 2006; Jeong et al., 2011; Jönsson et al., 2018). Here, we found a higher suitability of EVI for deriving phenological estimates in temperate broadleaf forests. This result is in

agreement with MODIS-based studies (Klosterman et al., 2014; Liang et al., 2011; Tan et al., 2008), but we showed that it also holds true for combined Landsat and Sentinel-2 time series. The NDVI has well-studied drawbacks compared to EVI such as its earlier saturation over areas with high biomass and its sensitivity to snowmelt dynamics. Additionally, the EVI is less sensitive to soil background (Huete et al., 2002; Tan et al., 2008). These factors might have influenced phenological estimates from NDVI time series in our study area. Besides NDVI and EVI, there might be additional vegetation indices better suited for estimating phenology. Jin and Eklundh (2014), for example, developed a plant phenology index that linearizes the relationship with canopy development. However, NDVI and EVI are easy to calculate and available as ready-to-use data, making them widely applicable.

#### 4.2. Evaluation of phenological estimates

We found a moderate agreement between satellite-based SOS and ground observations and a consistent offset between the two data sets. Moderate agreements between satellite-based SOS and ground observations have been reported also by other studies (Melaas et al., 2016; Rodriguez-Galiano et al., 2015), albeit the reported offsets vary. Analyzing different levels of leaf development at a broadleaf site in North America, Melaas et al. (2016) found that satellite-based SOS, also derived from the inflection point in the transition period, were most highly correlated to ground observations when the leaf length reached 25%. In this study, we only had ground observations of the unfolding of the first leaves, which explains why the satellite-based SOS estimates were systematically later. However, this is not problematic as these systematic differences can be accounted for.

The lack of a strong agreement between satellite and ground-based SOS is not entirely surprising, as ground observations target single trees and specific phenophases, which are not representative of the phenological dynamics of an entire forest community as observed from space (Fisher and Mustard, 2007; White et al., 2009). Our results, with support of these studies, suggest that it is challenging to evaluate satellite-based SOS estimates with ground-based observations of species-specific phenophases. Underlining this, Liang et al. (2011) found that an up-scaling approach from detailed ground observations to landscape-level phenology enhanced the comparability to satellite-based land surface phenology. Scientists and managers planning future phenological observation sites should recognize challenges related to scalability and representativeness of ground observations to improve their integration with satellite-based phenological monitoring networks.

The spatial variation of spring phenology related to regional-scale meteorological gradients was reproduced by estimates from combined Landsat and Sentinel-2 time series. In agreement with spatial patterns reported here, other studies (e.g. Doktor et al., 2005; Scheifinger et al., 2002) showed influences of prevalent large-scale weather conditions during spring on phenological timing in our study area. In central and southern Germany, SOS varied spatially with topography displaying temperature-elevation gradients and atmospheric inversions in mountain valleys (Scheifinger et al., 2002; Senf et al., 2017).

Our results suggest that combining Landsat and Sentinel-2 is a crucial step for estimating annual spring phenology of temperate forests from medium resolution sensors. Estimates from Landsat alone were generally poor as a result of the comparably low observation frequency. Melaas et al. (2016) also found that observation frequency of Landsat alone is not sufficient to map annual land surface phenology in temperate forests. In comparison, our SOS estimates from Sentinel-2 were generally consistent with the combined time series indicating that Sentinel-2 alone may be sufficient for annual SOS estimates. However, observation frequency was not evenly distributed across space. Outside swath overlap areas of Sentinel-2 and in mountain ranges, observation frequency was considerably lower, even from combined time series. Therefore, the inclusion of Landsat data added valuable information to the time series, especially in more data-sparse areas. Kovalskyy and Roy

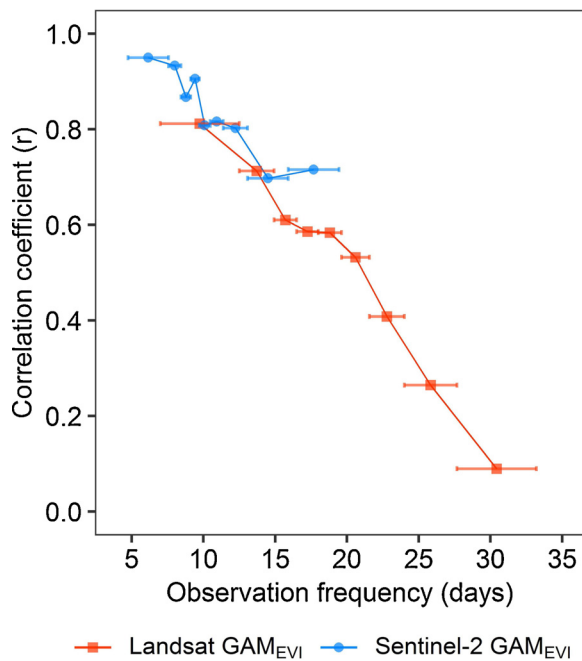


Fig. 7. Effect of cloud-free observation frequency on start of season (SOS) estimates shown by a decreasing correlation between SOS from single Landsat and Sentinel-2 time series and SOS estimates from combined time series. Observation frequency was binned by deciles.

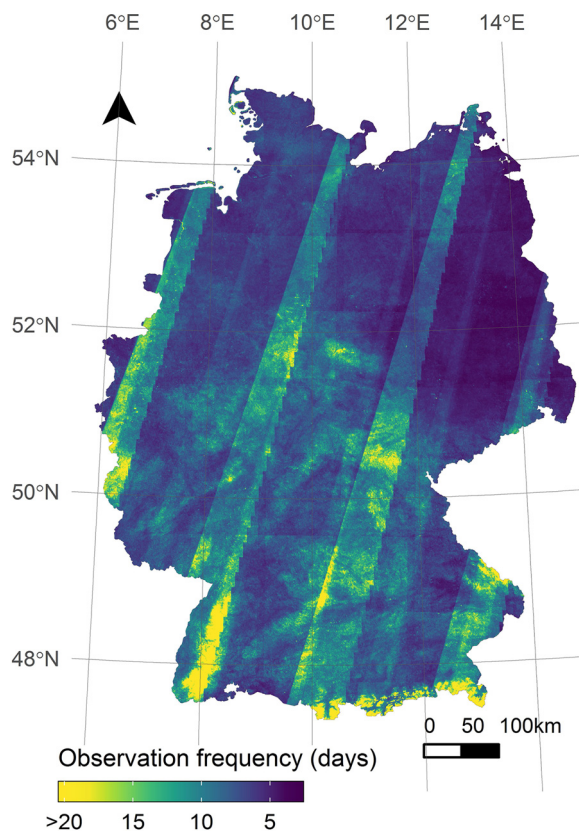


Fig. 8. Average frequency of cloud-free observations of Landsat and Sentinel-2 in Germany from January to June 2018.

(2015) showed that cloud cover substantially decreased available observations, e.g., across the United States. This was also the case for our study area as average observation frequencies were substantially lower than theoretical revisit intervals (Li and Roy, 2017). These results

highlight the need for dense time series to estimate SOS annually.

The frequency of cloud-free observations usually not only varies spatially but also from year to year. In this study, we analyzed only a single year, since 2018 was the first year for which both Sentinel-2A and 2B were available. Since 2018 was a hot and dry year with exceptionally low precipitation and cloud cover, we expect the results to be more optimistic than for more cloudy years. The year 2018 had an average of 2015 sun hours, the highest average since 1951 (thirty-year average = 1600 sun hours; German Weather Service, 2020). Also, the following year 2019 showed substantially higher than average sun hours in Germany (1834 sun hours). To understand how our results translate into future medium resolution phenology observations remains to be seen as Sentinel-2 time series accumulate, but we are also likely to see an increase in the capacity of medium resolution sensor acquisitions in the coming years (e.g., with the launch of Sentinel-2C).

## 5. Conclusion

Phenological shifts caused by climate change are altering structure, composition and function of ecosystems worldwide (Chuine, 2010; Parmesan and Yohe, 2003). Apart from ground observations of phenological stages, satellite remote sensing has become an important data source for deriving information on the phenological development of vegetation (Tang et al., 2016). In this study, we used dense time series of medium resolution data from Landsat and Sentinel-2 to model and characterize spring phenology of temperate broadleaf forests in Germany for a single year. By comparing two commonly used models and vegetation indices, we showed that the choice of vegetation index had a higher impact on the resulting SOS estimates than choice of model. We found a preference for EVI when comparing estimates to ground observations, even though we note that comparing land surface phenology to species-specific phenophases is challenging. We further showed that decreasing data density altered SOS estimates considerably, emphasizing the need for dense observation intervals and multi-sensor time series. Our results demonstrate the potential of combined Landsat and Sentinel-2 time series for estimating SOS annually.

## CRedit authorship contribution statement

**Katja Kowalski:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization.  
**Cornelius Senf:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing.  
**Patrick Hostert:** Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.  
**Dirk Pflugmacher:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research contributes to the Landsat Science Team 2018-2023 (<https://www.usgs.gov/land-resources/nli/landsat/2018-2023-landsat-science-team>). The research was supported by Geo. X, the Research Network for Geosciences in Berlin and Potsdam. We acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin. We would like to thank the anonymous reviewers for their constructive suggestions which helped to improve the manuscript. We thank David Frantz for facilitating preprocessing of Landsat and Sentinel-2 data with

FORCE. We also thank Markus Ungersböck for providing the phenological ground observations through the PEP725 project.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jag.2020.102172>.

## References

- Beck, P.S.A., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Remote Sens. Environ.* 100, 321–334. <https://doi.org/10.1016/j.rse.2005.10.021>.
- 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>.
- Buitenwerf, R., Rose, L., Higgins, S.I., 2015. Three decades of multi-dimensional change in global leaf phenology. *Nat. Clim. Change* 5, 364–368. <https://doi.org/10.1038/nclimate2533>.
- 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>.
- Chuiue, I., 2010. Why does phenology drive species distribution? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 3149–3160. <https://doi.org/10.1098/rstb.2010.0142>.
- Doktor, D., Badeck, F.W., Hattermann, F., Schaber, J., McAllister, M., 2005. Analysis and modelling of spatially and temporally varying phenological phases. *Geostatistics for Environmental Applications*. Springer-Verlag, Berlin/Heidelberg, pp. 137–148. [https://doi.org/10.1007/3-540-26535-X\\_12](https://doi.org/10.1007/3-540-26535-X_12).
- DWD Climate Data Center (CDC), 2019. Multi-Annual Station Means for the Climate Normal Reference Period 1981–2010, for Current Station Location and for Reference Station Location, Version V0.x.
- Federal Ministry of Food and Agriculture, 2018. *Der Wald in Deutschland - Ausgewählte Ergebnisse der dritten Bundeswaldinventur, 3rd ed.* .
- Fisher, J., Mustard, J., 2007. Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sens. Environ.* 109, 261–273. <https://doi.org/10.1016/j.rse.2007.01.004>.
- Frantz, D., 2019. FORCE—Landsat + Sentinel-2 analysis ready data and beyond. *Remote Sens.* 11, 1124. <https://doi.org/10.3390/rs11091124>.
- Gao, B.-C., Kaufman, Y.J., 2003. Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels. *J. Geophys. Res. Atmos.* 108. <https://doi.org/10.1029/2002JD003023>.
- German Weather Service, 2020. Time Series and Trends [WWW Document]. URL <https://www.dwd.de/EN/ourservices/zeitreihen/zeitreihen.html> (Accessed 4.2.20).
- German Weather Service, 2018. *Vorschriften und Betriebsunterlagen für die phänologischen Beobachter des Deutschen Wetterdienstes. Offenbach.*
- Hänninen, H., 1990. Modelling bud dormancy release in trees from cool and temperate regions. *Acta For. Fenn.* 213, 1–47. <https://doi.org/10.14214/aff.7660>.
- 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](https://doi.org/10.1016/S0034-4257(02)00096-2).
- Hufkens, K., Friedl, M., Sonntag, O., Braswell, B.H., Milliman, T., Richardson, A.D., 2012a. Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sens. Environ.* 117, 307–321. <https://doi.org/10.1016/j.rse.2011.10.006>.
- Hufkens, K., Friedl, M.A., Keenan, T.F., Sonntag, O., Bailey, A., O'Keefe, J., Richardson, A.D., 2012b. Ecological impacts of a widespread frost event following early spring leaf-out. *Glob. Change Biol.* 18, 2365–2377. <https://doi.org/10.1111/j.1365-2486.2012.02712.x>.
- 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. Change 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>.
- Jönsson, P., Cai, Z., Melaas, E., Friedl, M., 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>.
- Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonntag, 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>.
- Kovalsky, V., Roy, D., 2015. A one year Landsat 8 continuous United States study of cirrus and non-cirrus clouds. *Remote Sens.* 7, 564–578. <https://doi.org/10.3390/rs70100564>.
- Li, J., Roy, D.P., 2017. A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* 9, 902. <https://doi.org/10.3390/rs9090902>.
- Liang, L., Schwartz, M.D., Fei, S., 2011. Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. *Remote Sens. Environ.* 115, 143–157. <https://doi.org/10.1016/j.rse.2010.08.013>.
- Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* 33, 457–465. <https://doi.org/10.1109/36.377946>.
- Meier, U., 2018. Growth Stages of Mono- and Dicotyledonous Plants: BBCH Monograph. Open Agrar Repositorium <https://doi.org/10.5073/20180906-074619>.
- Melaas, E.K., Friedl, M.A., Zhu, Z., 2013. Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sens. Environ.* 132, 176–185. <https://doi.org/10.1016/j.rse.2013.01.011>.
- Melaas, E.K., Sulla-Menashe, D., Friedl, M.A., 2018. Multidecadal changes and inter-annual variation in springtime phenology of North American temperate and boreal deciduous forests. *Geophys. Res. Lett.* 45, 2679–2687. <https://doi.org/10.1002/2017GL076933>.
- 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>.
- NASA, J.P.L., 2013. NASA Shuttle Radar Topography Mission Global 1 Arc Second [Data Set]. <https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003>.
- Padfield, D., Matheson, G., 2018. nls.multstart: Robust Non-Linear Regression Using AIC Scores.
- Parnes, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421, 37. <https://doi.org/10.1038/nature01286>.
- Pflugmacher, D., Rabe, A., Peters, M., Hostert, P., 2019. Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. *Remote Sens. Environ.* 221, 583–595. <https://doi.org/10.1016/j.rse.2018.12.001>.
- Pflugmacher, D., Senf, C., Yang, Z., Seidl, R., Hostert, P., 2020. Characterizing Disturbance Regimes of Central European Forests Using Remote Sensing. *Manuscr. Prep.* .
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing.*
- Rodriguez-Galiano, V.F., Dash, J., Atkinson, P.M., 2015. Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. *Geophys. Res. Lett.* 42, 2253–2260. <https://doi.org/10.1002/2015GL063586>.
- Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T.L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., Imeson, A., 2008. Attributing physical and biological impacts to anthropogenic climate change. *Nature* 453, 353–357. <https://doi.org/10.1038/nature06937>.
- Scheifinger, H., Menzel, A., Koch, E., Peter, C., Ahas, R., 2002. Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central Europe. *Int. J. Climatol.* 22, 1739–1755. <https://doi.org/10.1002/joc.817>.
- Senf, C., Pflugmacher, D., Heurich, M., Krueger, T., 2017. A Bayesian hierarchical model for estimating spatial and temporal variation in vegetation phenology from Landsat time series. *Remote Sens. Environ.* 194, 155–160. <https://doi.org/10.1016/j.rse.2017.03.020>.
- Smith, R.J., 2009. Use and misuse of the reduced major axis for line-fitting. *Am. J. Phys. Anthropol.* 140, 476–486. <https://doi.org/10.1002/ajpa.21090>.
- Tan, B., Morissette, J.T., Wolfe, R.E., Gao, F., Ederer, G.A., Nightingale, J., Pedetty, J.A., 2008. Vegetation phenology metrics derived from temporally smoothed and gap-filled MODIS data. In: *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. <https://doi.org/10.1109/IGARSS.2008.4779417>. pp. III-593-III-596.
- Tang, J., Körner, C., Muraoka, H., Piao, S., Shen, M., Thackeray, S.J., Yang, X., 2016. Emerging opportunities and challenges in phenology: a review. *Ecosphere* 7. <https://doi.org/10.1002/ecs2.1436>.
- Tanre, D., Herman, M., Deschamps, P.Y., de Lefte, A., 1979. Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties. *Appl. Opt.* 18, 3587–3594. <https://doi.org/10.1364/AO.18.003587>.
- Templ, B., Koch, E., Bolmgren, K., Ungersböck, M., Paul, A., Scheifinger, H., Rutishauser, T., Busto, M., Chmielewski, F.-M., Hájková, L., Hodzic, S., Kaspar, F., Pietragalla, B., Romero-Fresneda, R., Tolvanen, A., Vučetić, 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>.
- Tucker, C.T., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150.
- Verma, M., Friedl, M.A., Finzi, A., Phillips, N., 2016. Multi-criteria evaluation of the suitability of growth functions for modeling remotely sensed phenology. *Ecol. Modell.* 323, 123–132. <https://doi.org/10.1016/j.ecolmodel.2015.12.005>.
- 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., van Leeuwen, W.J.D., Brown, J.F., de Wit, A., Schaefer, 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. Change Biol.* 15, 2335–2359. <https://doi.org/10.1111/j.1365-2486.2009.01910.x>.
- Wood, S.N., 2003. Thin plate regression splines. *J. R. Stat. Soc.* 65, 95–114. <https://doi.org/10.1111/1467-9868.00374>.
- 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](https://doi.org/10.1016/S0034-4257(02)00135-9).
- Zöller, L., Beierkuhnlein, C., Samimi, C., Faust, D., 2017. *Die Physische Geographie Deutschlands*. Wissenschaftliche Buchgesellschaft (WBG), Darmstadt.