Characterizing Spring Phenology of Broadleaf Forests Across Germany Combining Landsat/Sentinel Time Series and Phenological Models

Masterarbeit

Humboldt-Universität zu Berlin

Geographisches Institut

eingereicht von:

Gutachter: Dr. Dirk Pflugmacher

Dr. Cornelius Senf

Berlin, den 28.02.2019

Table of contents

List of figures

List of tables

Abstract

Vegetation phenology has a great impact on land-atmosphere interactions like carbon cycling, albedo, or 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. While a wide range of mechanistic phenological models have been applied to estimate key phenological parameters, it has been shown that these models fail to predict interannual variability in land surface phenology measured from global satellite archives. It is therefore necessary to improve our understanding of how phenological models relate to land surface phenology measured from various spaceborne remote sensing sources.

In this study, we analyzed the potential of multi-sensor, medium resolution time series for estimating spring phenology of broadleaf forests across Germany. We combined dense time series of Sentinel-2, Landsat 7, and Landsat 8 acquired in 2017. We fitted two phenological models, logistic models and generalized additive models, using Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) time series at 12,420 sampling sites, and compared their suitability for representing phenology across our study area. We further evaluated our estimates against state-of-the-art mechanistic models calibrated using literature-based species parameters and in-situ observations of leaf unfolding. This comparison also allowed to characterize the spatial variability in the drivers of spring phenology.

Our results show that estimates of spring phenology are feasible for single years by integrating multi-sensor time series. They also show, though, that the choice of vegetation index can have a strong impact on predictability. EVI time series outperformed NDVI time series in logistic models and generalized additive models across the study area. Moreover, results indicate that comparability of ground observations of leaf unfolding and start of season (SOS) estimates on landscape scales is limited. We further found that current mechanistic models were not able to reproduce spatial variability of SOS estimates. Local variations in thermal forcing, chilling and other environmental drivers such as built-up area and elevation explained large shares of differences between SOS dates from Landsat and Sentinel-2 time series and mechanistic models. The results underline the need for a better representation of spatial variability of spring phenology in process-based phenological models. The study shows that dense medium resolution time series from integrated Landsat and Sentinel-2 data can be used to advance our knowledge of phenological dynamics and their drivers across broader spatial scales. This is especially important in the light of climate change and increasing climate variability which impact phenology of forest ecosystems globally.

1 Introduction

In the last decades, it became evident that climate change has severe and diverse influences on ecosystems globally, encompassing spatial shifts of species ranges but also temporal shifts of phenological events (Jeong, Ho, Gim, & Brown, 2011; Parmesan & Yohe, 2003; Rosenzweig et al., 2008). A substantial body of research suggests that climate change, through warming, will extend the growing season of forests by facilitating earlier leaf unfolding in spring (Crabbe et al., 2016; Friedl et al., 2014) and changing dynamics of leaf senescence in autumn (Fu et al., 2018; Gill et al., 2015). In Europe, spring onset has advanced by approximately 2.5 days per decade in the last decades of the 20th century (Menzel et al., 2006). Therefore, vegetation phenology, the study of the timing of seasonal plant development through phases of active growth and dormancy (Hänninen, 1990), is a suitable and widely used indicator to measure impacts of climate change on ecosystems (Parmesan & Yohe, 2003; Rosenzweig et al., 2007). The consequences of a changing phenology on forest ecosystems are manifold. In addition to altering tree species distribution and ranges (Chuine, 2010; Chuine & Beaubien, 2001), changes in phenology are expected to impact ecosystem services such as the carbon uptake of temperate and boreal forests negatively (Han, Wang, Jiang, Fischer, & Li, 2018; Piao et al., 2008). In this regard, changing dynamics of spring and autumn phenology have received considerable attention in the field of phenological studies since these parameters determine growing season length which is a key factor controlling, e.g., ecosystem productivity (Crabbe et al., 2016; Richardson et al., 2010).

1.1 Phenological mechanisms of temperate broadleaf forests

Differences in phenological dynamics are often attributed to changing impacts of various environmental drivers such as temperature, photoperiod, and precipitation at different sites and points in time (Basler & Körner, 2012; Fu et al., 2018; Yun et al., 2018). However, there still is uncertainty about the influence of the different drivers on phenological responses of forests. While it has been widely assumed that a warming climate will extend the growing season in the future, by advancing spring leaf unfolding and delaying autumn senescence, a number of studies found contrasting evidence. Recently, a decrease in the sensitivity of leaf unfolding to warming has been detected in broadleaf tree species (Fu et al., 2015). Moreover, several studies (e.g. Chuine et al., 2016; Cook, Wolkovich, & Parmesan, 2012; Dantec et al., 2014; Fu et al., 2015) argue that chilling has been an underestimated factor in most phenological studies and models: Unfulfilled chilling requirements in certain tree species caused by warmer temperatures during winter can counterbalance early leaf-out dates induced by spring warming. The same might be

true for photoperiod, which has an increasing impact on spring phenology when leaf unfolding advances further. These factors are considered to become more important under future climate change and are not sufficiently represented in current phenological models (Chuine et al., 2016; Fu et al., 2015; Garonna et al., 2018). The underlying physiological plant mechanisms are not fully understood either. Adding to this, it has been suggested that daytime temperatures exert stronger controls than nighttime temperatures on both, spring and autumn phenology and should therefore be considered in future studies (Fu et al., 2016; Piao et al., 2015; Wu et al., 2018). Overall, it still remains challenging to conclude which environmental cues drive temperate forest phenology at which sites and to what extent.

Not all mechanisms and underlying processes related to phenological dynamics are understood in detail yet. However, it is known that the phenological cycle of temperate trees consists of different phases of active growth and dormancy to facilitate plant development during favorable environmental conditions and to ensure survival during temperatures below 0°C (Delpierre et al., 2016; Perry, 1971; Rohde & Bhalerao, 2007). In this study, we look specifically at spring leaf phenology of temperate broadleaf trees. Other plant parts of temperate and boreal trees also undergo a phenological cycle, e.g., wood, needles, fine roots and fruits which are, like leaves and flowers, sensible to environmental conditions (see e.g. Delpierre et al., 2016).

During late summer to early autumn, growth cessation is initiated in trees by different factors such as photoperiod, low temperatures and drought (Rohde & Bhalerao, 2007). In this time, broadleaf trees grow new buds which then go through different phases of dormancy, described e.g. by Perry (1971) and Lang, Early, Martin, & Darnell (1987): Firstly, paradormancy is a state where growth is inhibited not by the bud itself but signalized through other plant organs. After paradormancy, trees enter into endodormancy, where the perception of environmental or endogenous factors inhibiting growth appears within the buds themselves. To break endodormancy, the tree must be exposed to a certain amount of chilling temperatures, depending on tree species. Budburst occurs later in species with a higher chilling requirement and deeper rest, whereas tree species with a lower chilling requirement tend to start growth earlier. With spring warming, this delayed effect of chilling on budburst is expected to cause stronger advances of season onset in species with a low chilling requirement and smaller effects on species with a higher chilling requirement (Hänninen, 1990). Endodormancy break and transition into ecodormancy usually take place in late winter to early spring. During this last dormancy phase, unfavorable environmental conditions signal a growth restriction and prevent further development i.e. bud growth (Basler, 2016; Horvath, Anderson, Chao, & Foley, 2003). Because of a direct dependency of bud development on temperature, the rate of cell division

and bud growth is increasing with rising temperatures. The transition between different phases of dormancy are gradual and can be different among species, individuals and even buds of the same tree. After a sufficient amount of forcing temperatures have accumulated, bud break and leaf unfolding start (Dantec et al., 2014; Perry, 1971). Moreover, photoperiod has an effect on the timing of budburst of different species which could alleviate advancing budburst dates with climate warming (Basler & Körner, 2014; Linkosalo, Carter, Häkkinen, & Hari, 2000). It has been shown that chilling and forcing requirements are species-specific (Chuine, 2000) but also vary between individuals of the same species (Kramer et al., 2017). These intraspecies differences can mainly be explained by genetic variations (Wesołowski & Rowiński, 2006).

While genetic variations have to be considered as a factor driving spatial variability (Liang, 2016), other environmental gradients, mostly tied to temperature gradients, are known to impact spatial differences of spring phenology. Higher elevations are associated with lower temperatures and therefore later occurrence of leaf unfolding and green up. Accordingly, several studies found a delay in spring phenology with increasing elevation (Čufar, De Luis, Saz, Črepinšek, & Kajfež-Bogataj, 2012; Richardson, Bailey, Denny, Martin, & O'Keefe, 2006; Senf, Pflugmacher, Heurich, & Krueger, 2017). Besides, there is evidence, that elevation gradients differ among species (Hufkens, Friedl, Keenan, et al., 2012). However, Vitasse, Signarbieux, & Fu (2018) found that spatial variability of spring phenology decreased along elevation gradients. This was caused by opposing effects of spring warming on the fulfillment of chilling requirements at different elevations. In addition to elevation, urban heat islands are advancing spring phenology and extending the growing season of vegetation in the cities as well as in surrounding areas. Phenology along the urban-rural gradient has gained increasing attention in the last years with various studies targeting urban areas in, e.g., North America (X. Li et al., 2017; Melaas, Wang, Miller, & Friedl, 2016; Zipper et al., 2016), Europe (Dallimer, Tang, Gaston, & Davies, 2016) and Asia (Zhou, Zhao, Zhang, & Liu, 2016).

1.2 Research approaches of phenological studies

Two different research approaches are apparent within the study area of phenology. One approach relies on in-situ observations, which are used in process-based phenological models. The other approach uses remote sensing data to measure land surface phenology directly (Fisher, Richardson, & Mustard, 2007).

Comprehensive ground observations of phenological events have been registered throughout Europe since the 1950s. Over 12 million records are accessible through the PEP725 database,

which represents a valuable data source. It has been used in a variety of different studies related to, e.g., climate change and variability and the sensitivity of phenology to various environmental drivers (see Templ et al., 2018).

To explain and predict phenological responses to current and future climate, mechanistic (also "process-based") models take different combinations of drivers of phenology such as chilling and forcing temperatures, photoperiod, and vapor pressure into account to estimate leaf unfolding in spring (Basler, 2016). Most simple models, such as the Thermal Time model and the Sequential model make use of the concept of growing degree days (GDD) and chilling days (CD) (Chuine, Cour, & Rousseau, 1999). These parameters represent species- and possibly site-specific (Kramer et al., 2017) forcing and chilling requirements which determine endodormancy and ecodormancy break, respectively. Daily or hourly mean temperatures are used to assess the state of chilling and forcing accumulation during autumn, winter, and spring prior to the growing season (Chuine et al., 1999). These models are commonly calibrated with phenological ground observations (e.g. Basler, 2016; Linkosalo, Lappalainen, & Hari, 2008).

Basler (2016) compared and evaluated most common mechanistic phenological models and found that model performance was overall very similar, despite different model complexities and environmental drivers included in the model. Moreover, most of the models tended to underestimate inter-annual variability and errors increased when very late or early leaf-out dates were evident. In line with these results, Richardson et al. (2012) and Keenan et al. (2012) found that inter-annual variation of phenology was not sufficiently represented in terrestrial biosphere models and caused differences between modeled and observed terrestrial carbon fluxes. Model errors were especially large for the broadleaf forest type. This shows that current phenological models are not likely to represent phenology well under future climate conditions and need improvement to correctly express corresponding biosphere-atmosphere interactions. Recently, Wang et al. (2018) used remotely sensed phenology to improve modeling of gross primary productivity in Chinese forests. While ground observations have been the main source for calibrating process-based models locally, the potential of remote sensing data to calibrate these models to enable spatially and temporally flexible phenology estimates has not been exploited yet.

In recent years, remote sensing data have been used to measure land surface phenology (LSP) directly. LSP describes the biophysical dynamics of vegetation connected to its annual cycle of development and dormancy as measured by remote sensing (White & Nemani, 2006). Ground observations are visual interpretations of plant development by a human interpreter whereas

LSP is derived by analyzing the spectral information from a certain sensor which includes signals of the complete vegetation cover and background (e.g. soil, water) within a pixel (Fisher & Mustard, 2007; Misra, Buras, & Menzel, 2016).

During the last decades, most studies used coarse-resolution data from sensors such as the Moderate-resolution Imaging Spectroradiometer (MODIS) (Keenan & Richardson, 2015; Liang & Schwartz, 2009; L. Liu et al., 2015) and the Advanced Very High Resolution Radiometer (AVHRR) (e.g. White et al., 2009) to map phenological dynamics of forests. While the spatial resolution of the respective sensors is not sufficient to discern phenological patterns on landscape scales (Fisher & Mustard, 2007; Hufkens, Friedl, Keenan, et al., 2012), low observation density and cloud cover constrain the use of higher-resolution sensors such as the Landsat ensemble in certain areas (P. Jönsson, Cai, Melaas, Friedl, & Eklundh, 2018). The launch and operation of the two Sentinel-2 satellites in 2013 and 2017 significantly increases observation density. By combining Landsat 8, Sentinel-2A and Sentinel-2B, a median observation interval of 2.9 days could be achieved theoretically. In practice though, cloud cover and haze can decrease the observation density considerably (J. Li & Roy, 2017). Landsat 7 delivers only fragmentary data since 2003 when the scan line corrector failed (Markham, Storey, Williams, & Irons, 2004). However, current incomplete scenes can still be used and therefore can add valuable observations to time series analysis especially in areas with frequent cloud cover. The combination of Landsat and Sentinel-2 data therefore is a valuable data source for characterizing phenological dynamics with higher spatial and temporal resolution (Melaas, Friedl, & Zhu, 2013).

A variety of different methods for time series fitting and extraction of phenological parameters, most commonly start of season (SOS) and end of season (EOS), have been developed so far. There exist different methods to derive phenology estimates from time series through curve fitting: While some studies applied threshold-based approaches (e.g. Garonna et al., 2018; Keenan et al., 2014), it is also possible to use different fitted parameters directly, e.g., from a logistic function (Verma, Friedl, Finzi, & Phillips, 2016). Logistic functions have been commonly used by a range of studies to derive spring and autumn phenology from remote sensing time series (Verma et al., 2016): Fisher et al. (2007) used the logistic function to model spring phenology from MODIS data. Senf et al. (2017) applied a logistic function in a Bayesian modeling framework to characterize spatial and temporal variability of spring phenology using multiannual Landsat time series. P. Jönsson et al. (2018) used box-constrained fits on a double logistic function to derive start and end of season of several years from Normalized Difference Vegetation Index (NDVI) Landsat and Sentinel-2 time series over a test site in Sweden.

Similarly, Melaas, Friedl, & Zhu (2013) created pooled time series of Enhanced Vegetation Index (EVI) from Landsat data with two separate logistic functions to estimate longtime averages but also inter-annual variability of spring and autumn phenology. Other methods which have been used include the application of different filters and polynomial functions (e.g. Garonna et al., 2018; Jönsson & Eklundh, 2004). However, White et al. (2009) showed that estimates from different threshold-based algorithms do not consistently correlate for larger study areas. Verma et al., 2016 on the other hand, found a high agreement between results of four different growth functions, confirming the suitability of the logistic function to model spring phenology. Asymmetric signals in spring and autumn are poorly captured by logistic functions though (Verma et al., 2016). Therefore, Melaas, Sulla-Menashe, et al. (2016) used cubic splines to estimate historical phenological dynamics in several study sites across North America from Landsat time series. Moreover, Buitenwerf, Rose, & Higgins (2015) applied cubic splines to model various phenological parameters on a global scale. Compared to the logistic function, smoothing splines have been used infrequently to extract spring phenology from remote sensing time series.

While methods developed to extract phenology from satellite-based data are manifold, validation data for LSP is generally scarce and oftentimes disparities regarding scale, monitoring approach and spatial representation restrict their applicability. Phenology estimates of data sources such as near-surface remote sensing have been compared to remote sensing-based phenology and several studies found good agreement, especially regarding spring phenology (e.g. Klosterman et al., 2014; Melaas et al., 2016). Near-surface remote sensing in form of stationary sensors facilitates very dense observation intervals. Nonetheless, Hufkens, Friedl, Sonnentag, et al. (2012) emphasized substantial uncertainties regarding scalability and landscape representation when comparing near-surface remote sensing to satellite-based remote sensing.

Most commonly, phenological estimates from remote sensing have been compared to ground (also “in-situ”) observations (Misra et al., 2016). There are, however, several issues related to this: Firstly, observations on point level (ground observations) do not necessarily represent phenology on the pixel level as observed by satellite sensors. Since ground observations are mostly species-specific, it is unclear if single species can represent landscape scale phenology or phenology of the respective area of a single pixel (White et al., 2009). Adding to this, the biophysical development of the vegetation as observed by remote sensing does not resemble the same process as phenological phases observed on ground level. For example, the SOS estimate based on a vegetation index threshold might not be comparable to different stages of

leaf development observed on ground level (Fisher & Mustard, 2007; Misra et al., 2016; White et al., 2009). Due to these constraints, it is generally difficult to use ground observations as verification for LSP which explains contradictory results regarding the comparability of in-situ observations and remote sensing-derived LSP: Fisher et al. (2007) found a low agreement between ground observations and LSP from coarse resolution sensors as MODIS. By contrast, e.g. Melaas, Sulla-Menashe, et al. (2016) reported strong and consistent correlations between SOS observed on ground and SOS estimated from Landsat data. This indicates that in-situ observations might be suitable to validate LSP from medium resolution remote sensing time series.

1.3 Research objectives

In this study, we characterize spring phenology of temperate broadleaved forests in Germany by exploiting the potential of multi-sensor, medium resolution time series. We assess the applicability of the frequently used logistic model and generalized additive models using thin plate regression splines to derive spring phenology of temperate broadleaf forests. We then compare our remote sensing-based estimates of spring phenology with in-situ observations and two mechanistic models, the Thermal Time model and the Sequential model, to analyze the regional importance of meteorological drivers (such as thermal forcing and chilling) and to reveal potential limitations of current process-based models. In order to overcome those limitations, we finally use remote sensing estimates to derive GDD and CD from process-based models. Apart from forcing and chilling mechanisms, other environmental drivers such as elevation, cardinal directions and extent of urban areas are of interest because they are associated with spatial variability in spring phenology, which can be detected using medium resolution multi-sensor time series.

Specifically, we want to address the following questions:

- How does the choice of model and vegetation index influence the SOS estimates from Landsat/Sentinel-2 time series?
- How do SOS estimates from Landsat/Sentinel-2 compare to in-situ observations?
- How do SOS estimates from Landsat/Sentinel-2 compare to estimates from mechanistic models?
- How well do environmental factors explain the spatial variability in SOS estimates?

2 Methods

The study approach consists of three parts: We first derive spring phenology across Germany from remote sensing time series, built from Sentinel-2 and Landsat 7 and 8 data. We test two different models (logistic model and generalized additive model (GAM)) with two different vegetation indices (EVI and NDVI) to extract SOS dates for the year 2017 at 414 different locations across Germany and compare them to available in-situ observations of leaf unfolding. Further, we evaluate potential environmental drivers of spring phenology and their impact on the spatial variability of SOS estimates. In the second part, we apply two commonly used mechanistic phenological models, the Thermal Time (TT) model and the Sequential model, using meteorological data from the German Weather Service. We then assess the differences between SOS estimates from time series and SOS dates from mechanistic models. Adding to this, we obtain the amount of chilling (i.e. CD) prior to SOS. In a third step, we use SOS dates from remote sensing to derive GDD from the TT model for each plot in our study area to assess the spatial representation of chilling and thermal forcing from the TT model in Germany (Fig. 1). In the following sections, we describe details and specifications regarding the study area, remote sensing data, meteorological data, sampling design, time series models and mechanistic models.

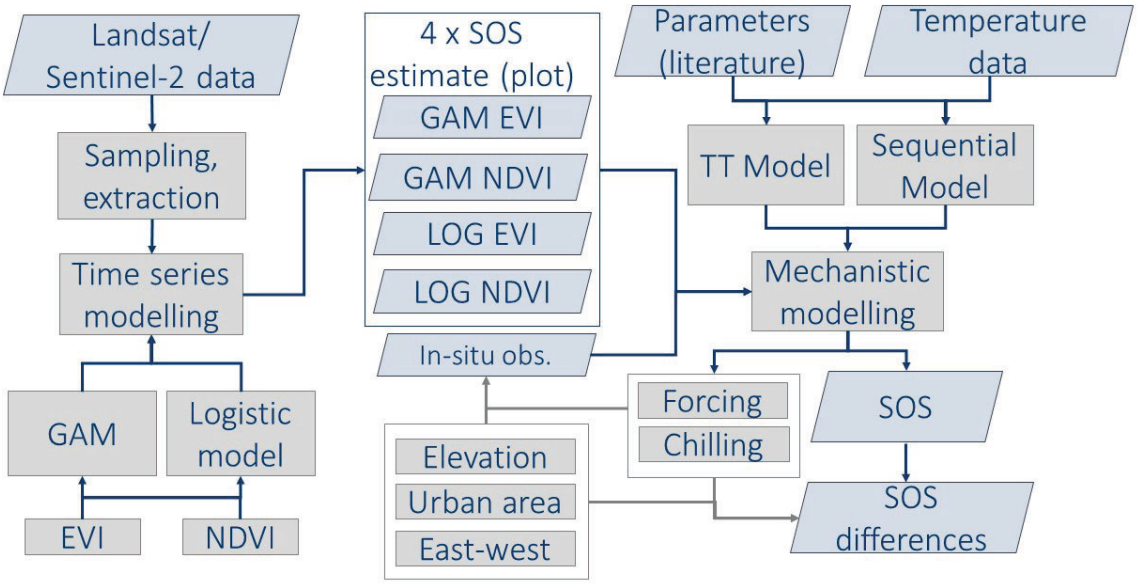


Fig. 1: Data, processing steps and modeling approaches to derive SOS and GDD estimates for Germany from Landsat and Sentinel-2 time series and process-based models.

2.1 Study area

Germany spans an area of 357.386 km^2 from which about a third is covered by forests belonging to the northern temperate forest biome in Central Europe. Nearly 42% are broadleaf forest and consist of commonly occurring broadleaf species including European beech (*Fagus sylvatica*), sessile oak (*Quercus petraea*), common oak (*Quercus robur*), silver birch (*Betula pendula*), European ash (*Fraxinus excelsior*) and sycamore maple (*Acer pseudoplatanus*). Most broadleaf tree species are endemic in Germany with non-native species accounting for only 5% of the total forest area. Broadleaved trees – such as oak or beech – also can reach a relatively high age (mean age of 102 and 100 years) (Bundesministerium für Ernährung und Landwirtschaft, 2018). Especially older forests enable various benefits such as increased microhabitats and carbon storage and are therefore of great ecological value (Luyssaert et al., 2008; Stephenson et al., 2014; Winter, Höfler, Michel, Böck, & Ankerst, 2015).

Climate, as one of the main factors driving phenology of forests, is characterized by an east-west gradient in Germany: The eastern regions of the country are primarily influenced by a continental climate whereas western Germany is dominated by oceanic influences. Overall, the temperature range is smaller in northwestern Germany, with mild winters and summers. Continentality increases in eastern Germany where winter temperatures are lower and summer temperatures are higher than in northwestern parts of the country. In central German mountain ranges and towards the Alps, temperatures are lower due to higher elevations (Zöller, Beierkuhnlein, Samimi, & Faust, 2017).

Generally, elevation increases in Germany from north to south. The northern German plains are lower than 200m a.s.l. and have been shaped by several glacial advances. Towards the central German uplands, elevation ranges from 200 to 1,500m a.s.l. The landscape is structured by valleys, basins and mountain ranges, e.g., the Harz mountains, the Rhenish Slate mountains, the Black Forest and the Thuringian and the Bavarian Forest. The Alpine foreland begins approximately south of the Danube river where the landscape is shaped by glacial sediments. Further south, a small portion of the Alps is part of Germany where mountains reach elevations of nearly 3,000m (Semmel, 1996; Zöller et al., 2017).

2.2 Sampling design

The sampling design is based on the network of the meteorological stations (from here on referred to as plots) in Germany operated by the German Weather Service. Sampling units are pixels of different sizes corresponding to each sensor (30m for Landsat 7 and 8, and 10m for

Sentinel-2). Thirty random samples were drawn in 5,000m distance around each plot in undisturbed broadleaf forest (Fig. 2). We included plots with at least 350 pixels of broadleaf forest within 5,000m distance. We used the land cover classification of Pflugmacher et al. (2019) and Pflugmacher, Senf, Yang, Seidl, & Hostert (2019) to identify undisturbed broadleaf forest with a minimum mapping unit of 11 pixels, corresponding to 9,900m^2 in total. Most plots with insufficient broadleaf forest area were situated in lower mountainous areas such as the Harz Mountains, the Bavarian Forest, and the Black Forest, where coniferous forest prevails. Other omitted plots lie in the north of Germany, mostly on islands or within mudflats where little to no broadleaf forest exist. Overall, 12,420 pixels were sampled for 414 plots. To ensure that not more than one sample falls within a single pixel, the minimum distance between each sample is 30m in X and Y direction. The design enables a meaningful comparison between the SOS estimates from remote sensing time series and the SOS estimates from different mechanistic models due to the spatial proximity to the plots where meteorological data was collected.

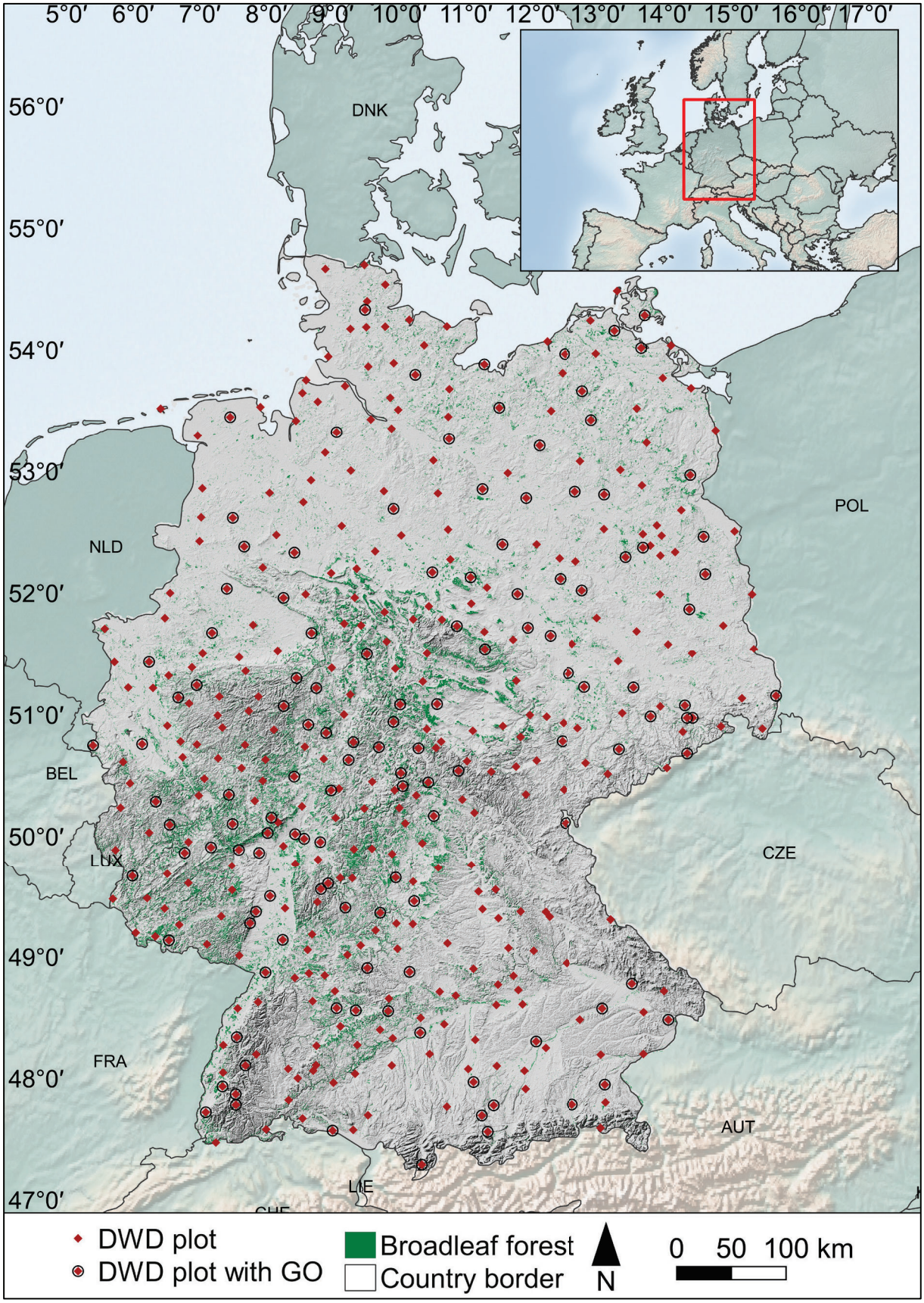


Fig. 2: Study area of Germany with meteorological plots, ground observations (GO), and broadleaf forest cover (background map: Natural Earth).

2.3 Data

2.3.1 Landsat/Sentinel-2 time series

We used optical remote sensing data from three different sensors to derive dense time series for the year 2017. The sensors include the Enhanced Thematic Mapper + (ETM+, aboard Landsat 7), the Operational Land Imager (OLI, aboard Landsat 8), and the Multi Spectral Instruments (MSI, aboard Sentinel-2A and Sentinel-2B).

For preprocessing of the data, we used the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE v.2.0). The algorithm is capable of transforming level 1 data to analysis-ready data ranging from level 2 to 4. To obtain level 2 data, preprocessing steps include cloud and cloud shadow detection, quality screening, atmospheric and topographic correction, adjacency effect correction and BRDF reduction. FORCE is able to integrate Landsat 4-8 and Sentinel-2A/2B as a Virtual Constellation (VC) (Frantz, 2018; Frantz, Röder, Stellmes, & Hill, 2016). To ease time series analysis, processed data is transferred into a gridded tile structure. For further information on FORCE please see Frantz et al. (2016).

To build time series for each sample, we extracted spectral reflectance data of bands in the visible, near infrared and short-wave infrared spectrum (excluding the 3 red edge bands of Sentinel-2) for the year 2017. Quality flags were used to exclude all observations with an indication of cloud cover, cloud shadow or snow.

In phenological studies, different vegetation indices have been used to assess the spatial and temporal variation of vegetation state and to extract phenological parameters. Most studies used the NDVI or the EVI in their analysis (Melaas, Sulla-Menashe, & Friedl, 2018). We therefore test both indices and assess their applicability. The NDVI was defined by Tucker (1979) as:

$$NDVI = \frac{(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + \rho_{red})} \quad (1)$$

where ρ_{NIR} and ρ_{red} are the respective reflectance values of the red and near-infrared band. The NDVI can take values between -1 and 1, with vegetation surfaces typically yielding values in the range of 0 to 1.

The EVI is similar to the NDVI, as it also makes use of the near infrared and red bands. By adding additional parameters and the blue band, the EVI takes atmospheric effects and soil

background into account which leads to a better differentiation of the vegetation signal (Huete et al., 2002). The EVI has been defined by H. Q. Liu & Huete (1995) as:

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \times \rho_{red} - C_2 \times \rho_{blue} + L} \quad (2)$$

where ρ_{NIR}, ρ_{blue} and ρ_{red} are the respective reflectance values of the near-infrared, blue and red band, C_1 and C_2 are coefficients for correcting aerosol scattering and L is a factor accounting for soil background. The coefficients take values of $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$ *and* $L = 1$ (Huete, Liu, Batchily, & van Leeuwen, 1997).

2.3.2 Meteorological data

The German Weather Service operates a network of approximately 500 meteorological stations in Germany from which daily weather data is available (German Weather Service, 2013). We obtained meteorological data for 414 plots across Germany for the time period from mid-August 2016 to mid-August 2018 (DWD Climate Data Center (CDC), 2018). The altitude of all plots ranges from -5 to 1485m a.s.l. We downloaded air temperature data as daily mean air temperature and filled data gaps in daily mean temperatures of up to 2 days using linear interpolation. We excluded plots with data gaps larger than 2 days from 1st January to 15th July 2017 for the TT model and in the time period from day of year (DOY) 245 (1st September) in 2016 to DOY 196 (15th July) in 2017 for the Sequential Model.

2.3.3 In-situ observations of leaf unfolding

We accessed observations of phenological parameters for available broadleaved tree species in Germany from the PEP725 database for 2017 (see App. A for number and share of observations per tree species). All observations are classified according to the BBCH scale which describes various phenological growth stages of plants over the year. We used all records with BBCH code 11, which notes a stage, where first leaves of a perennial plant are unfolded (Meier, 2018). Compliant with the sampling design, we derived the average date of leaf unfolding of all available observations and broadleaved tree species within 5,000m distance of overall 130 meteorological plots across Germany.

2.4 Landsat/Sentinel-2 time series models

For each of the 12,420 samples, two different types of models were fitted: (i) a GAM using smoothing splines; and (ii) a parametric logistic model (Fig. 1). As we were primarily interested in spring phenology, we only used observations from late winter, spring and summer. Hence, we selected all clear observations from DOY 1 to the day of year where the vegetation index reached its maximum, plus an extra 20 days to account for variation in the data and to improve model fit. For both models, we used two different vegetation indices, EVI and NDVI, as input variables. We found that unrealistic high NDVI values, most probably caused by snow during the first months of the year, led to problems with model fitting. We therefore calculated a base vegetation index (VI) (i.e. the mean of all observations in the time period from DOY 1 to DOY 50) and filled all VI values for DOY 1-50 with this base value.

By fitting each model and index combination, we derived four spring phenology estimates for each sample (Fig. 1). We aggregated all sample estimates from each model separately for every corresponding plot and evaluated model performance based on the mean squared error (MSE). The model results were compared by calculating the mean difference between the predicted SOS dates. Additionally, we assessed differences between SOS derived from time series models and available in-situ observations. We therefore used linear regression analysis, the corresponding root-mean-square error (RMSE) and Pearson product-moment correlation. In the following sections, both, the logistic model and the GAM are described in detail.

2.4.1 Logistic model

The logistic model is the most commonly used parametric model to derive different phenological parameters from remote sensing time series (Verma et al., 2016) and has been applied in various studies (e.g. Hufkens, Friedl, Sonnentag, et al., 2012; P. Jönsson et al., 2018; Melaas et al., 2013; Senf et al., 2017). Our model uses a logistic function of the form:

$$VI(t) = v_{min} + \left(\frac{v_{max}}{1 + e^{-g(t-s)}}\right) \quad (3)$$

Where VI is the vegetation index of choice at DOY t, v_{min} and v_{max} are the minimum and maximum of the function, respectively, g is the rate of change (i.e. vegetation green-up rate), t is time given in DOY, and s is the location of the inflection point, which we define as SOS based on previous studies (Melaas et al., 2013; Zhang et al., 2003). The logistic function is symmetrical around the inflection point, where the growth rate is at its maximum (Fig. 3).

Regarding vegetation phenology, this is the point in time, where the change rate in canopy development is highest (Verma et al., 2016).

The model was fitted using the *nls* function in R, which uses a Gauss-Newton algorithm to fit the function to the data (R Core Team, 2017). As prior guesses on the approximate location of model parameters are needed for faster convergence of the model, we set the starting values for v_{min} and v_{max} to the minimum and maximum of the vegetation index of the time series, the starting value for the rate of change g to 0.2, and the initial SOS value, s, to the mean of all values which are larger than the median.

2.4.2 Generalized additive model

In addition to the logistic model, we used a GAM to fit the time series of each sample. We used the *mgcv* package in R (Wood, 2003) which includes a variety of different smoothing options, e.g., cubic splines and thin plate regression splines. The general formula of a one-dimensional GAM is:

$$VI(t) = \beta_0 + s(t) + \varepsilon_t \tag{4}$$

where VI is the observed vegetation index value at DOY t, β_0 is the intercept of the function at $t = 0$, s represents a smoothing function of covariate t, in this case the DOY and ε_t is a random error term, with $\varepsilon_t \sim N(0, \sigma^2)$.

The smoothing function is defined by:

$$s(t) = \sum_{k=1}^{K} \beta_k b_k(t) \tag{5}$$

where the final smoothing function, $s(t)$, is the sum of all basis functions $b_k(t)$ multiplied by their corresponding weight β_k. As a smoothing function, we here use thin plate regression splines as recommended by Wood (2003). Thin plate regression splines of one dimension are generally comparable to cubic splines. However, knot locations do not need to be defined in thin plate regression splines (Wood, 2003). Basis functions are created for each data point. In reality, however, it is unlikely that complex basis functions like this are needed to approximate the true function. Therefore, the number of basis functions is reduced by an eigendecomposition where eigenvectors with the k largest eigenvalues are kept. The weight parameter, β_k, is estimated for each basis function using least square regression. The final smoothing function is

obtained by summing up the values of the scaled basis functions at each value of t (Simpson, 2018; Wood, 2003). Similar to 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. 3). The rate of change of $VI(t)$ at a given point $P(t|VI)$ is found by approximating the first derivative using finite differences.

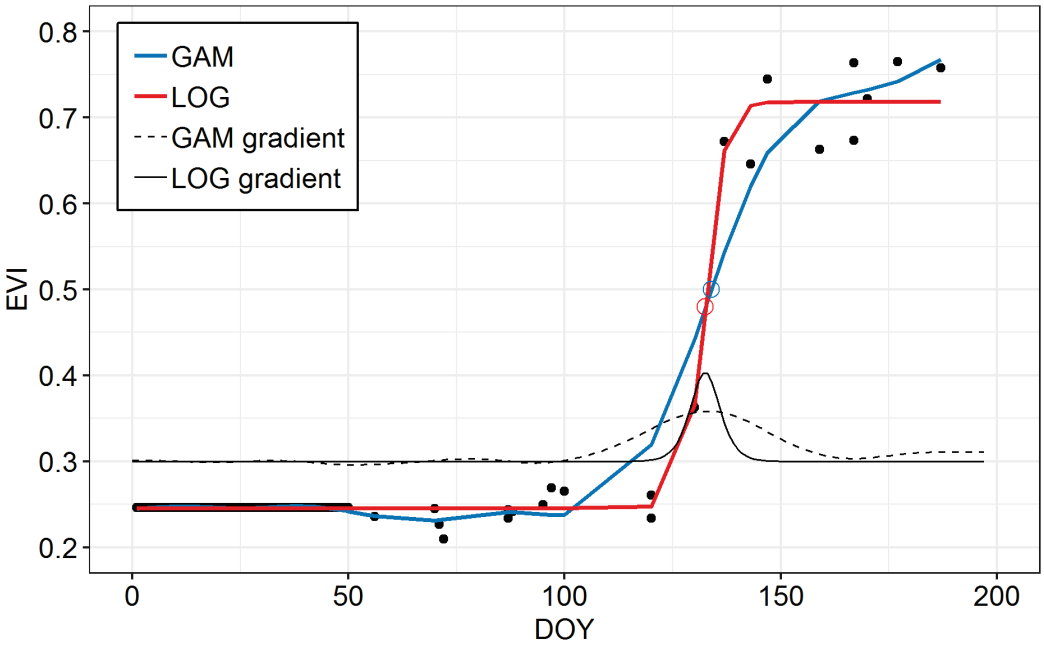


Fig. 3: Exemplary model fits of the logistic model (LOG) and the GAM using EVI as input variable with corresponding gradient functions, multiplied by 2 and 4, respectively (black points: observations, red and blue circles: SOS estimate of logistic model and GAM).

2.5 Mechanistic models

Mechanistic models have been commonly used to simulate the timing of phenological events such as budburst and leaf unfolding (Fu et al., 2015). The simplest models simulate ecodormancy break by taking only forcing temperatures into account. More advanced models additionally model endodormancy release by including a chilling response and different combinations of other factors such as photoperiod and vapor pressure (Basler, 2016). We decided to test two different models: The TT model and the Sequential model. We used comparably simple models and temperature responses because it has been shown that more complex phenological models do not necessarily perform better (Basler, 2016; Linkosalo et al., 2008).

Additionally, we calculated a linear regression of SOS depending on mean spring temperatures. We therefore used mean temperature in the time period from 1st February 2017 to 31st May

2017. We assessed spatial differences between spring phenology estimates from time series models and the process-based models by calculating differences between results of each model type and combination. Finally, we derived GDD for each plot by calibrating the TT model with our SOS estimates from remote sensing time series.

2.5.1 Thermal Time model

The TT model has been described by several authors including Basler (2016), Chuine (2000), and Fu et al. (2015). It models the timing of leaf unfolding by accounting for the impact of forcing temperatures on bud development:

$$S_f(t) = \sum_{t_f}^{t_{GDD}} R_f(x_t) \tag{6}$$

where S_f describes the state of forcing and x_t is the daily mean temperature. Forcing temperatures start to accumulate at DOY t_f and are completed when reaching t_{GDD}. We use a linear temperature response. Thus, R_f is defined by:

$$R_f(x_t) = \begin{cases} 0 & if\ x_t < T_{b1} \\ x_t - T_{b1} & if\ x_t \geq T_{b1} \end{cases} \tag{7}$$

where T_{b1} is a temperature threshold controlling forcing temperature accumulation. We used a linear representation of forcing temperatures. The forcing requirement is fulfilled if S_f is equal to or larger than the forcing threshold F^*. T_{b1} and F^* are set to 5°C and 250 GDD as used and determined by Dantec et al. (2014). We assessed differences between spring phenology estimates from time series models, the TT model and ground observations by calculating differences between SOS dates of each model.

In addition to calibrating the TT model with literature-based values to derive the date of leaf unfolding for each plot, we also used the SOS estimates from time series modeling to obtain GDD for the respective plot by setting $t_{GDD} = SOS$. We evaluated differences between GDD estimates from models based on leaf unfolding dates from ground observations and from SOS dates obtained from remote sensing time series. We calculated differences and further used linear regression and corresponding RMSE and Pearson correlation coefficients to compare the GDD estimates.

2.5.2 Sequential model

The Sequential model is an extended version of the TT model. Besides forcing temperatures, it also considers chilling as an additional factor influencing the timing of leaf unfolding. The name of the model points to the implementation of chilling and forcing responses: The chilling requirement has to be fulfilled before forcing accumulation starts (Chuine, 2000). The forcing response differs in the parameter t_f which is here the DOY when $S_c \geq C^*$. In addition to the forcing response specified in equation (6) and (7), chilling is represented by:

$$S_c(t) = \sum_{t_c}^{t_{CD}} R_c(x_t) \tag{8}$$

where S_c is the state of chilling at DOY t and x_t is the daily mean temperature. Chilling temperatures start to accumulate at t_c. R_c is defined by:

$$R_c(x_t) = \begin{cases} 0 \; if \; x_t > T_{b2} \\ 1 \; if \; x_t \leq T_{b2} \end{cases} \tag{9}$$

where T_{b2} is a temperature threshold controlling CD accumulation. t_{CD} is the point in time when R_c is equal to or larger than C^*. Ecologically, this is the timing of endodormancy break. Additionally, we derived the amount of chilling of each model and index combination occurring in the time period from 1st Sep 2016 until the SOS from remote sensing time series for each plot in 2017. To determine a suitable value of C^*, we first tested chilling requirement values for *Q. robur* and *F. sylvatica* specified by Dantec et al. (2014). However, we found the chilling requirement of 80 CD to be too high (i.e. not to be fulfilled during winter) for our study area. This was especially the case for northwestern parts of Germany which are more influenced by an oceanic climate with a smaller annual temperature range. We therefore used the mean chilling requirement of all plots in 2017 (64 CD) derived from remote sensing-based SOS.

Tab. 1: Parameter values used for the TT model and the Sequential model.

	Parameter	Value
TT Model	t_f	DOY 1 (1st Jan 2017)
	T_{b1}	5 °C
	F^*	250
Sequential Model	t_c	DOY 245 (1st Sep 2016)
	T_{b1}/T_{b2}	5 °C
	C^*	64
	t_f	DOY where $S_c \geq C^*$
	F^*	250

2.6 Environmental drivers of spring phenology

We evaluated different potential environmental drivers of spring phenology in our study area. As described in the previous section, we obtained chilling and forcing from SOS estimates for 2017. We examined three additional factors which have been described to impact spring phenology (see section 1.1): Elevation, east-west gradient and amount of urban area. We extracted the elevation of each plot from a digital elevation model (DEM) with 30m spatial resolution (SRTM-1). The location along the east-west gradient was determined from the x-coordinate of each plot. Lastly, we derived the amount of built-up area within 5,000m distance around each plot based on the land cover class 'artificial land' by Pflugmacher et al. (2019). We used the Pearson correlation coefficient and simple linear regressions to evaluate the impact of different drivers on SOS dates, model differences, and GDD.

3 Results

3.1 SOS estimates from integrated Landsat/Sentinel-2 time series

We applied two different types of models to estimate SOS from Landsat/Sentinel-2 time series: The logistic model specified in section 2.4.1 and the GAM using thin plate regression splines specified in section 2.4.2. Due to their higher flexibility when fitting datasets, GAMs converged for nearly all sample time series while the logistic model converged in 53.81% to 74.80% of the cases, using NDVI and EVI, respectively. Samples were fitted with a mean of 14 observations. We were able to aggregate SOS dates for all plots (except for one plot from the logistic model using NDVI), derived from 24.61 samples per plot on average.

We found a moderate correlation between models on the sample level (r_{EVI} = 0.75 and r_{NDVI} = 0.67) which increased to r_{EVI} = 0.88 and r_{NDVI} = 0.83 after aggregation to plot level. Agreement between models was higher when using EVI instead of NDVI (Tab. 2). In line with this, differences between the two models using either EVI or NDVI were smaller than differences between the two vegetation indices applied to the same model. Accordingly, mean differences between models were 3.76 days (EVI) and 1.85 days (NDVI). Mean differences of vegetation indices amounted to -8.32 days for the logistic model and -10.11 days for the GAMs. This indicates that the GAM estimated later SOS than the logistic model and that models using EVI estimated later SOS than models using NDVI (Tab. 2). SOS estimates for all plots ranged from DOY 70 to DOY 153 for all four model and index combinations. However, the majority of values (0.05th - 0.95th quantile) ranged from DOY 103 to DOY 142, corresponding to SOS estimates in the time period from mid-April to the end of May across Germany (Tab. 3).

Tab. 2: Correlations and mean differences for four different model and vegetation index combinations.

	Model		**Index**	
	LOG_{EVI}	LOG_{NDVI}	LOG_{NDVI}	GAM_{NDVI}
	GAM_{EVI}	GAM_{NDVI}	LOG_{EVI}	GAM_{EVI}
r_{sample}	0.75	0.67	0.61	0.44
r_{plot}	0.88	0.83	0.77	0.76
mean difference (sample)	3.80	1.79	-8.91	-10.12
mean difference (plot)	3.76	1.85	-8.32	-10.11

Tab. 3: Summary statistics for four remote-sensing models, ground observations and two process-based models.

		Landsat/Sentinel-2 models				**Ground**	**Mechanistic models**	
		LOG_{NDVI}	LOG_{EVI}	GAM_{NDVI}	GAM_{EVI}	**observations**	TT model	Seq. model
SOS	Convergence	0.54	0.75	1.00	1.00	-	-	-
	$Q_{0.05}$	103	114	103	117	94	105	106
	$Q_{0.5}$	119	128	121	131	107	133	133
	$Q_{0.95}$	132	136	135	142	124	143	141
	MSE	0.000868	0.000491	0.000655	0.000495	-	-	-
	Mean SOS	118	126	120	130	107	131	130
	$r_{Seq.\ model}$	0.54	0.67	0.53	0.65	0.57	1.00	-
	$r_{TT\ model}$	0.54	0.68	0.54	0.66	0.66	-	1.00
	$r_{ground\ observations}$	0.32	0.45	0.37	0.53	-	0.66	0.57
	$r_{mean\ spring\ temperature}$	-0.50	-0.63	-0.50	-0.59	-0.71	-0.94	-0.95
	$R^2_{forcing}$	0.00	0.01	0.05	0.05	0.11	0.50	0.45
	$r_{forcing}$	0.02	-0.09	0.23	0.22	-0.34	-0.71	-0.67
	$R^2_{chilling}$	0.34	0.32	0.29	0.29	0.48	0.50	0.44
	$r_{chilling}$	0.58	0.57	0.54	0.54	0.70	0.71	0.66
	$R^2_{elevation}$	0.04	0.06	0.02	0.05	0.17	0.13	0.05
	$r_{elevation}$	0.21	0.24	0.13	0.21	0.41	0.36	0.23
	$R^2_{east\text{-}west}$	0.00	0.01	0.00	0.00	0.02	0.14	0.15
	$r_{east\text{-}west}$	-0.05	0.11	-0.01	0.03	0.15	0.38	0.39
	R^2_{urban}	0.12	0.18	0.12	0.15	0.06	0.26	0.24
	r_{urban}	-0.34	-0.42	-0.34	-0.39	-0.25	-0.51	-0.49

Differences between models were smaller than differences between vegetation indices (Fig. 4). Both models tended to produce similar results when using the respective vegetation index, however, GAMs estimated slightly later SOS dates with both, EVI and NDVI, compared to the logistic model. Contrary to the two models, the two tested vegetation indices produced considerable different SOS dates (0.05^{th} - 0.95^{th} quantile): While SOS estimated from EVI ranged from DOY 114 to 136 and DOY 117 to 142, SOS estimates from NDVI spanned a larger range of DOY 103 to 132 and DOY 103 to 135, using the logistic model and GAM, respectively (Fig. 4). The median SOS was also markedly later in both models using EVI compared to NDVI (Tab. 3). We used the MSE averaged for all plots to assess model performance which differed strongly between vegetation indices. Using EVI leads to considerably lower MSE, independent from model choice, compared to models which were fitted with NDVI time series (Tab. 3).

Both models estimated consistently later SOS than observed on the ground when using EVI (Fig. 5). While SOS dates from NDVI were less delayed, their correlation to ground observations of leaf unfolding was weaker compared to SOS dates from EVI. In line with this, the regression line deviated stronger from the identity line and RMSEs were larger for NDVI model results. Considering these measures, the model combination of GAM and EVI agrees most with SOS of ground observations (r = 0.53). Consequently, mean SOS from ground observation was earliest at DOY 107 followed by SOS from NDVI models (118 and 120) and SOS from EVI models (126 and 130; logistic model and GAM, respectively).

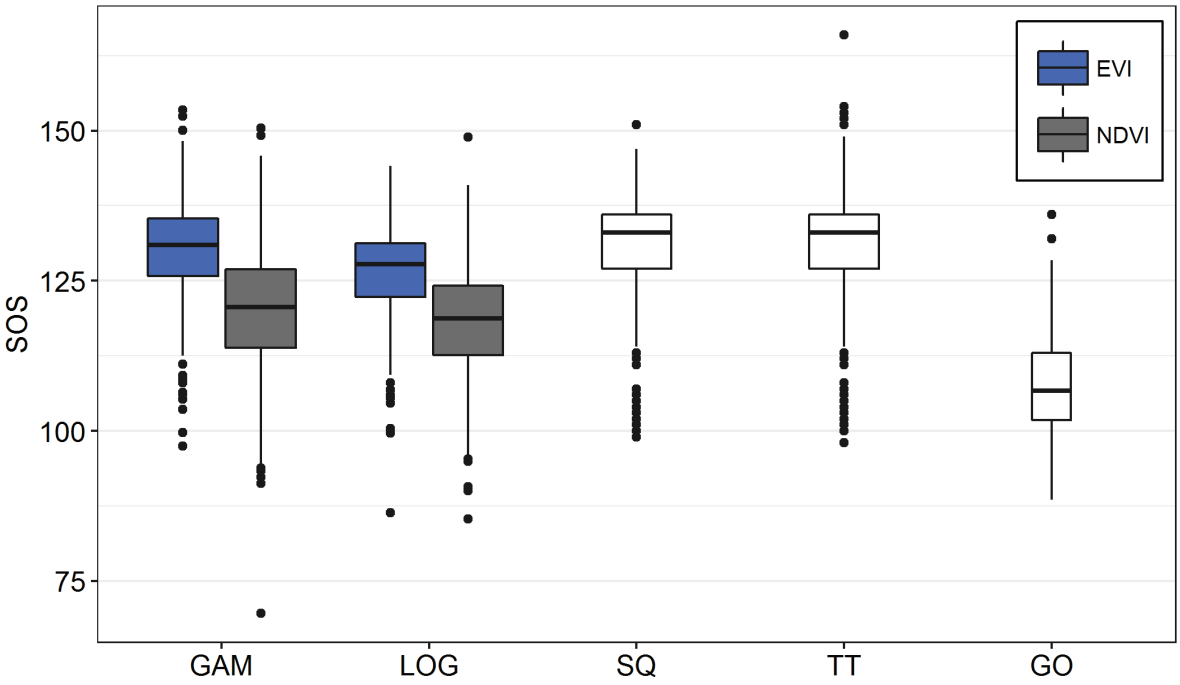


Fig. 4: Boxplots showing the median (horizontal black line), inter-quartile range (box), ±1.5-times inter-quartile range (black lines), and outliers (points) of SOS derived from the four model-index combinations (GAM, LOG), the Thermal Time model (TT), the Sequential model (SQ) and ground observations (GO).

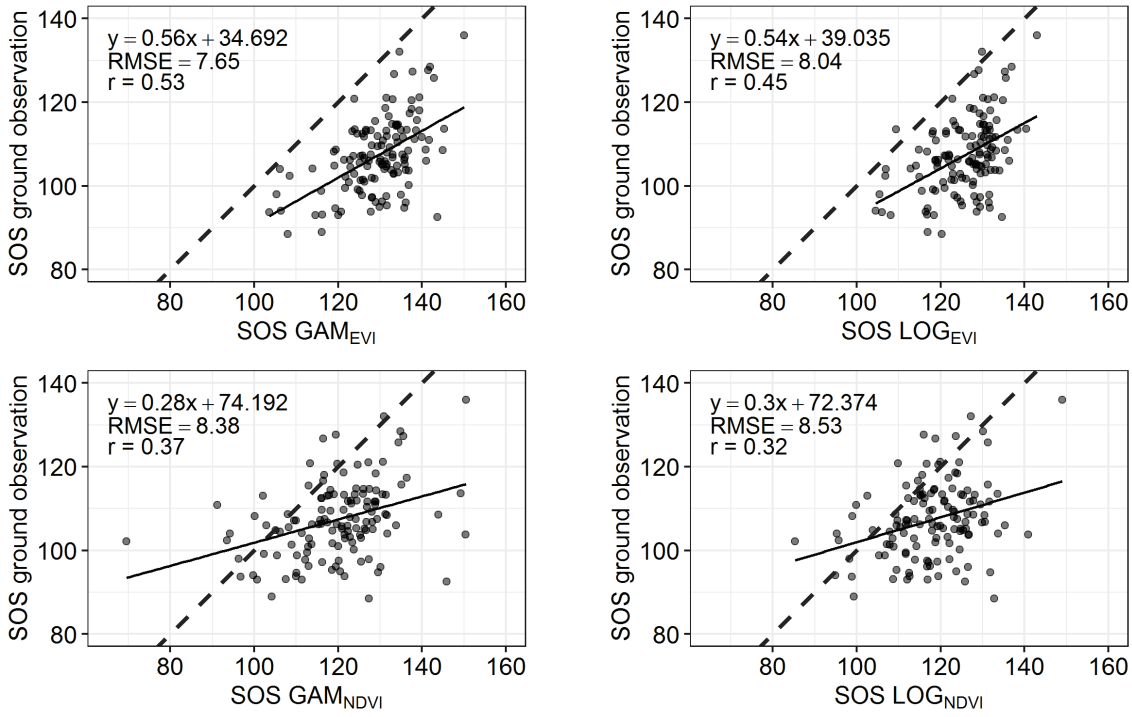


Fig. 5: SOS dates from four model-index combinations with linear regression to ground observations of leaf unfolding (solid black: regression line, dashed grey: identity line).

We assessed spatial variability in SOS dates visually for all four model-index-combinations across Germany. Moreover, we evaluated the correlation of SOS from different models and several environmental drivers. There was no clear pattern across the study area regarding SOS estimates. This also reflects in very weak correlations of opposing directions of east-west position and SOS dates from remote sensing models and ground observations (Tab. 3). However, distinct regional patterns were visible in all outputs (Fig. 6). Spring onset was markedly earlier near large urban settlements, e.g., in and around Berlin, Hamburg, Munich and the Ruhr area including the cities of Cologne, Dusseldorf, and Duisburg. This interpretation is supported by correlation coefficients ranging from r_{LOG} = -0.42 and r_{GAM} = -0.39 for EVI models to r = -0.34 for both NDVI models suggesting earlier SOS dates with an increasing amount of built-up area surrounding the plot (Tab. 3). SOS dates were also comparably earlier in the Upper Rhine Valley where temperatures are generally higher than in surrounding areas (Zöller et al., 2017).

In low mountainous areas such as the Rhenish Massif and in the vicinity of the Harz Mountains, SOS estimates were consistently later. The same holds true for plots near the lower mountain ranges along the Czech-German border including the Ore Mountains and the Bavarian forest. In southern Germany, especially in the Bavarian Alpine foreland, SOS dates tended to be later with decreasing distance to the Alps but also towards the Black Forest in the west (Fig. 6). While weak correlations with elevation were evident for all four remote sensing-based models ($0.13 \leq r \leq 0.24$), we found a moderate correlation of SOS dates from ground observations and elevation (r = 0.41). The direction of the relations indicated, as expected, that SOS occurred later with increasing elevation.

SOS estimates and thermal forcing (i.e. GDD) showed no clear correlations (Tab. 3) indicating that variation in GDD estimates does not explain spatial variation in SOS estimates. Regarding chilling though, moderate to strong positive correlations were evident for SOS estimates from all data sources. For SOS estimates from remote sensing, differences in chilling explained between 29.10% and 33.89% of the variation (Tab. 3). Overall, by evaluating the relation of SOS dates and environmental gradients, we found that chilling and the amount of built-up land influenced SOS dates most, compared to thermal forcing, elevation, and east-west gradient.

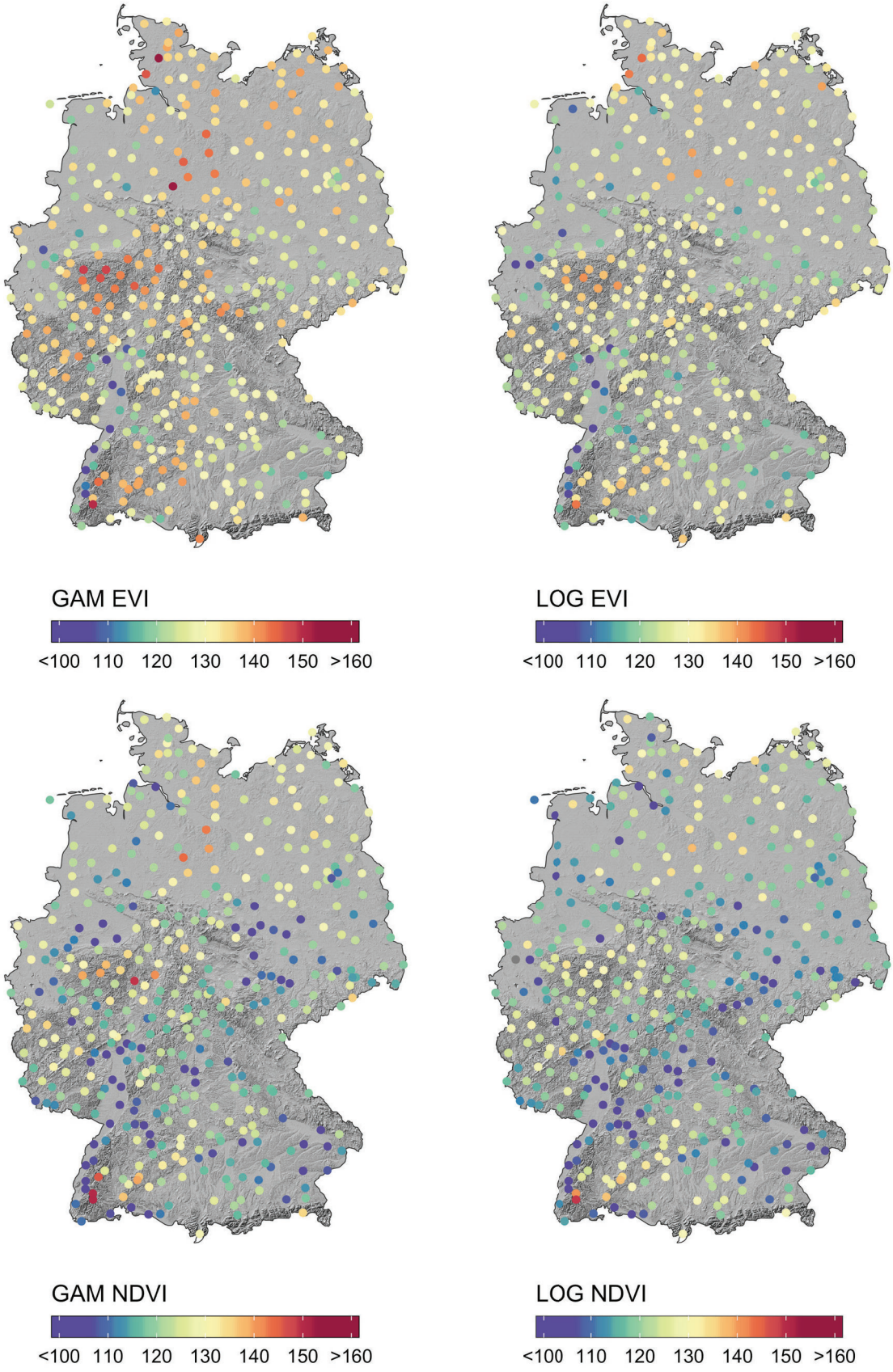


Fig. 6: SOS estimates (DOY) from Landsat/Sentinel-2 time series for all four model and index combinations (grey points: no data).

3.2 SOS estimates from mechanistic models

SOS estimates from mechanistic models (0.05^{th} - 0.95^{th} quantile) ranged from DOY 105 to DOY 143 (TT model) and from DOY 106 to DOY 141 (Sequential model) corresponding to mid-April to the end of May, generally agreeing with SOS estimates from logistic models and GAMs. When comparing the results with SOS derived from remote sensing time series, we observed moderately strong positive correlations indicating agreement between SOS resulting from both modeling approaches. However, correlations varied depending on applied vegetation index: Models fitted with EVI time series correlate stronger with SOS dates of the TT model with r_{GAM} = 0.66 and r_{LOG} = 0.68. For models fitted with NDVI, moderate correlations of r = 0.54 were obtained from GAM and the logistic model. SOS from the Sequential model correlated equally strong with SOS from time series models (Tab. 3), suggesting no substantial differences between the two process-based models. Leaf unfolding dates from ground observations also showed moderate positive correlations with the TT model (r = 0.66) and the Sequential model (r = 0.57), indicating overall good agreement. We also assessed the spatial variability of SOS estimates from mechanistic models visually. Compared to SOS estimates from Landsat/Sentinel-2 time series, SOS from the TT and Sequential model exhibit less variability especially towards the north of Germany (Fig. 7). However, distinct patterns such as earlier SOS dates in the vicinity of cities and later SOS estimates in mountainous areas are also discernible from both mechanistic models.

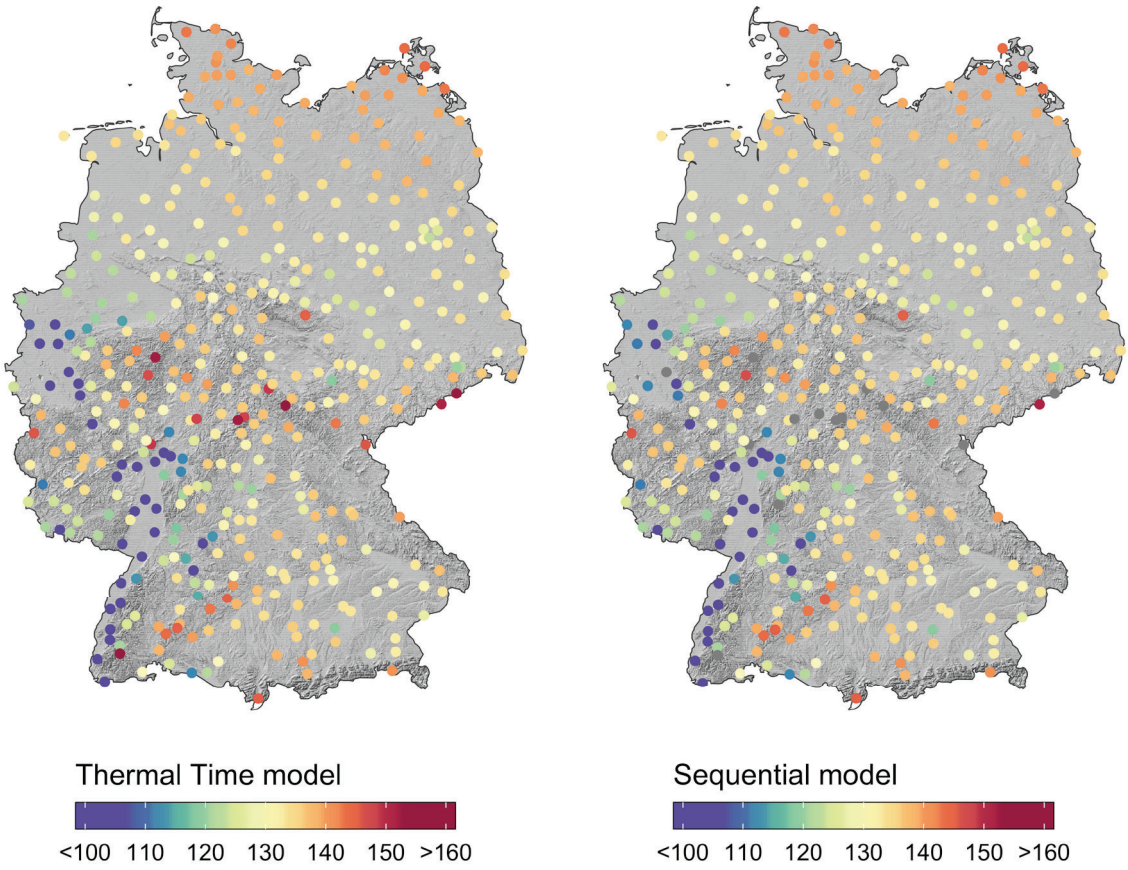


Fig. 7: SOS dates derived from the TT model and the Sequential model (grey points: no data).

Additionally, the linear regression with mean spring temperature yielded moderate correlations ($-0.63 \leq r \leq -0.50$) to SOS from remote sensing data and slightly stronger correlations for in-situ observations ($r = -0.71$) (Tab. 3). However, when only considering plots with ground observations, correlations further increased up to $r = -0.67$ for SOS from remote sensing time series. Again, correlations were stronger and RMSE were smaller for both models using EVI than for models using NDVI (Fig. 8). As expected, correlation coefficients indicate earlier spring onset with warmer temperatures and later SOS at sites with lower mean spring temperature. Regarding the mechanistic models, correlations to mean spring temperature were markedly higher with $r_{TT} = -0.94$ and $r_{Seq} = -0.95$.

These results emphasize that global mechanistic models were mainly representing differences in spring temperature and were not able to entirely reproduce spatial variability of remote sensing-based SOS dates and observations of leaf unfolding.

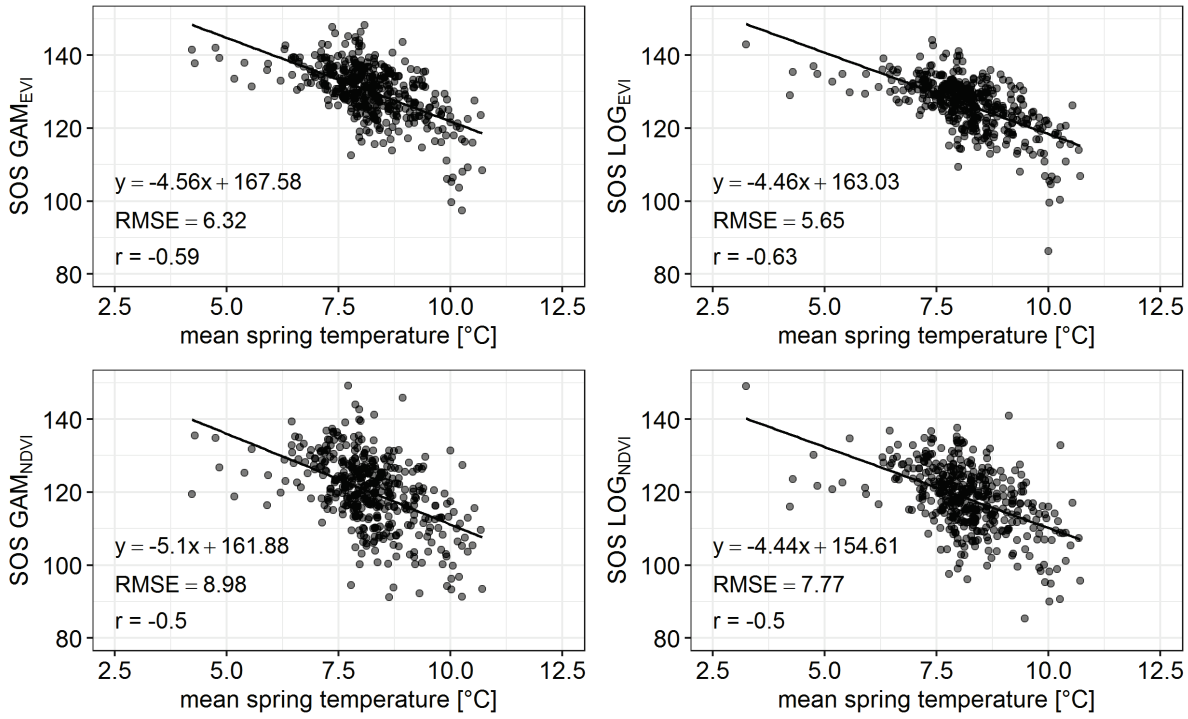


Fig. 8: Relation of SOS from four model and vegetation index combinations to mean spring temperature (averages of mean daily temperature from 1st February to 31st May) with fitted linear regression.

Regarding environmental gradients, we found no considerable correlation in east-west direction in SOS derived from remote sensing but also ground observations. However, SOS from TT and Sequential model exhibited moderately strong correlations (r_{TT} = 0.38, r_{Seq} = 0.39). Moreover, correlations of SOS from mechanistic models to thermal forcing and chilling were considerably higher than correlations of remote sensing models with r_{TT} = -0.71 and r_{Seq} = -0.67 (forcing), and r_{TT} 0.71 and r_{Seq} = 0.66 (chilling) (Tab. 4). SOS dates from both mechanistic models correlated slightly stronger with the amount of urban area (r_{TT} = -0.51, r_{Seq} = -0.49) than remote sensing models, suggesting an earlier start of season with increasing area of artificial landcover in the vicinity of the plot. Regarding elevation, SOS dates from the TT model had a higher correlation (r = 0.36) than SOS estimates from the Sequential model (r = 0.23). Overall, we found considerably higher correlations of SOS estimates from both mechanistic models to thermal forcing and east-west position compared to SOS estimates from remote sensing time series.

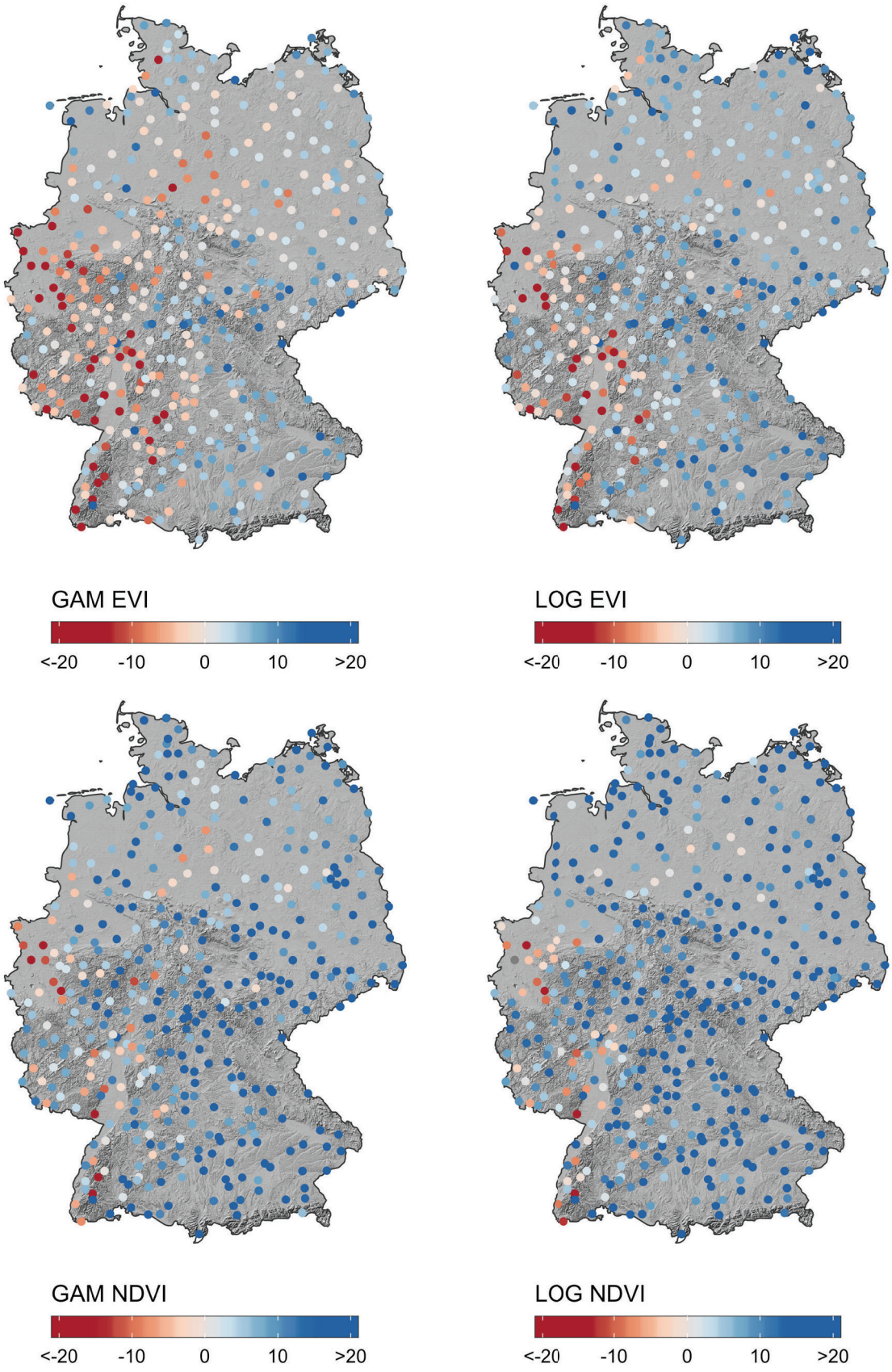


Fig. 9: Differences (in days) of SOS estimates from all four model and index combinations and SOS dates from the TT Model (grey points: no data).

3.3 Environmental drivers of spring phenology

Mean dates of season onset were nearly equal for the TT model (DOY 131, 11th of May), the Sequential model and GAM using EVI (DOY 130, 10th of May). The mean SOS from the logistic model (EVI) (DOY 126, 6th of May) is slightly earlier than SOS from process-based models, whereas mean SOS dates derived from NDVI models were more than 10 days earlier (Tab. 3). We calculated differences between the two process-based models and each model-index combination based on remote sensing data (Fig. 9, App. B). It is apparent that NDVI models consistently estimated earlier SOS dates than the process-based models. For EVI models, the differences were more equally distributed around zero, with a mean difference to the TT model of 0.39 and 4.15 days for GAM and logistic model, respectively (Tab. 4). Mean differences between SOS estimates from remote sensing and the Sequential model were minimally lower than mean residuals of the TT model, but the general direction and strength did not change.

Adding to this, spatial patterns of the residuals of both models were very consistent. We therefore describe spatial variation of differences between both process-based models and the four remote sensing models jointly. In Fig. 9 and App. B, negative values indicate an earlier SOS estimate from the TT model and the Sequential model whereas positive values indicate later SOS estimates compared to remote sensing models. The spatial distribution of residuals followed an east-west gradient among all four models with moderate correlations of TT model differences ranging from $r_{LOG} = 0.42$ to $r_{GAM} = 0.47$ for EVI and $r_{LOG} = 0.40$ to $r_{GAM} = 0.46$ for NDVI. Correlations of very similar strength were evident for differences of the Sequential model (Tab. 4). The process-based model results suggest earlier SOS dates than measured from remote sensing in western parts of the country. In central and eastern regions of Germany, spring phenology estimates from both process-based models were later than SOS dates from remote sensing-based models. For models using NDVI, residuals indicating later SOS from process-based models were prevalent and consistent throughout east, north and central Germany. Earlier SOS dates from process-based models were only evident for single plots mainly along the upper and lower Rhine valley (Fig. 9). For all remote sensing models though, differences to both mechanistic models had strong positive correlations to thermal forcing ($-0.92 \leq r \leq -0.85$). Consequently, local variability of thermal forcing, expressed as GDD, accounted for up to 84.07% ($0.73 \leq R^2 \leq 0.84$) of model differences, showing that both mechanistic models were not representing spatial variability in forcing mechanisms.

We observed weak correlations of TT model differences to elevation for all remote sensing models ($0.19 \leq r \leq 0.26$) which decreased further for differences to the Sequential model ($0.10 \leq r \leq 0.16$). In line with this, we found that model differences to the TT model exhibited weak ($r_{LOG} = 0.22$, $r_{GAM} = 0.18$ for NDVI) to moderate ($r_{LOG} = 0.44$, $r_{GAM} = 0.40$ for EVI) positive correlations to the amount of chilling present at different plots. Hence, the TT model estimated increasingly late SOS dates with increasing chilling compared to SOS estimates from EVI time series. Again, correlations decreased for differences to the Sequential model (Tab. 4) indicating that the Sequential model, by including the chilling mechanism, is able to better reproduce spatial variability in SOS caused by differences in chilling requirement. Chilling explained up to 19.08% ($0.16 \leq R^2 \leq 0.19$) of model differences for the TT model to GAM and logistic model using EVI but only up to 10.93% of model differences ($0 \leq R^2 \leq 0.11$) for the Sequential model (Tab. 4). Overall, we found that differences in SOS estimates can largely be explained by the local variation of thermal forcing connected to an east-west gradient and, to a smaller extent, by variation in underlying drivers such as chilling but also urban area and elevation.

Tab. 4: Mean, correlation coefficients and R^2 for SOS differences to the TT model and the Sequential model.

		Landsat/Sentinel-2 models				Ground
		LOG_{NDVI}	LOG_{EVI}	GAM_{NDVI}	GAM_{EVI}	**observations**
Residuals TT model	Mean	12.44	4.15	10.50	0.39	22.64
	$R^2_{forcing}$	0.74	0.84	0.76	0.83	0.54
	$r_{forcing}$	-0.86	-0.92	-0.87	-0.91	-0.74
	$R^2_{chilling}$	0.05	0.19	0.03	0.16	0.05
	$r_{chilling}$	0.22	0.44	0.18	0.40	0.21
	$R^2_{elevation}$	0.04	0.07	0.06	0.07	0.04
	$r_{elevation}$	0.19	0.26	0.24	0.26	0.20
	$R^2_{east\text{-}west}$	0.21	0.17	0.16	0.22	0.12
	$r_{east\text{-}west}$	0.46	0.42	0.40	0.47	0.35
	R^2_{urban}	0.06	0.09	0.03	0.08	0.21
	r_{urban}	-0.24	-0.30	-0.18	-0.28	-0.46
Residuals Seq. Model	Mean	12.35	4.07	10.45	0.34	22.59
	$R^2_{forcing}$	0.73	0.83	0.75	0.84	0.65
	$r_{forcing}$	-0.85	-0.91	-0.87	-0.91	-0.80
	$R^2_{chilling}$	0.01	0.11	0.01	0.09	0.04
	$r_{chilling}$	0.12	0.33	0.08	0.30	0.20
	$R^2_{elevation}$	0.01	0.02	0.03	0.02	0.03
	$r_{elevation}$	0.10	0.13	0.16	0.15	0.16
	$R^2_{east\text{-}west}$	0.20	0.16	0.15	0.22	0.12
	$r_{east\text{-}west}$	0.45	0.41	0.39	0.47	0.35
	R^2_{urban}	0.04	0.06	0.02	0.05	0.21
	r_{urban}	-0.19	-0.25	-0.13	-0.23	-0.46

Values for mean GDD varied from 188 to 244 depending on the model-index combination (Tab. 5). In accordance with SOS results, GDD estimates were larger when EVI was used compared to NDVI. GDD (0.05^{th} - 0.95^{th} quantile) from NDVI however, had a smaller range (119 to 279) than GDD from EVI (218 to 334). Both EVI models were associated with higher GDD compared to GDD from NDVI models (Fig. 10), reflecting SOS results, where EVI models estimated later SOS dates than NDVI models.

GDD based on leaf unfolding dates were smaller than GDD based on SOS estimates from remote sensing. This is in line with SOS dates from remote sensing which were later compared to ground observations of leaf unfolding. However, moderate to strong correlations ($0.48 \leq r \leq 0.74$) were evident for GDD from ground observations and SOS dates from Landsat/Sentinel-2 time series indicating a good agreement between GDD values. Here, GDD based on SOS from logistic models agreed more than GDD based on SOS from GAM indicated by lower RMSEs and higher correlation coefficients (Fig. 11).

The spatial distribution of GDD from remote sensing data followed an east-west gradient ($-0.42 \leq r \leq -0.38$) with small-scale patterns disrupting the general gradient. Overall, more GDD were observed in western Germany compared to the eastern parts of the study area. This result also corresponds well to the spatial distribution of model differences between process-based and remote sensing-based models where remote sensing models estimated earlier SOS (i.e. less GDD) in the eastern and later SOS (i.e. more GDD) in the western parts of the study area compared to mechanistic models. Moreover, lower GDD were evident in some mountainous areas as the Harz mountains and mountain ranges along the Czech-German border including the Bavarian Forest and the Ore Mountains. Consequently, a higher amount of CD was evident in these areas. Besides, we found high variations in GDD on small spatial scales (Fig. 12, App. E).

As expected, our results showed that CD and GDD estimates were negatively correlated for all models ($-0.62 \leq r \leq -0.43$), indicating less GDD with an increasing number of CD and vice versa. Correlation coefficients were higher for both logistic models compared to GAMs (Tab. 5). However, we observed local divergence from this pattern, e.g., for plots in the Rhenish Massif and single plots along the northern coast of the Baltic and the North Sea, where both, GDD and CD were comparably high (Fig. 12, App. E).

Tab. 5: Summary statistics for GDD estimates from four remote-sensing models and ground observations.

		Landsat/Sentinel-2 models				Ground
		LOG_{NDVI}	LOG_{EVI}	GAM_{NDVI}	GAM_{EVI}	observations
GDD	Mean	188	218	197	244	163
	$Q_{0.05}$	119	149	123	170	95
	$Q_{0.5}$	185	217	192	241	161
	$Q_{0.95}$	263	299	279	334	227
	$r_{GDD\ ground\ obs.}$	0.67	0.74	0.48	0.62	-
	$r_{chilling}$	-0.58	-0.62	-0.46	-0.43	-0.69
	r_{urban}	0.37	0.34	0.26	0.23	0.56
	$r_{east\text{-}west}$	-0.41	-0.38	-0.38	-0.42	-0.33
	$r_{elevation}$	-0.31	-0.35	-0.29	-0.28	-0.43

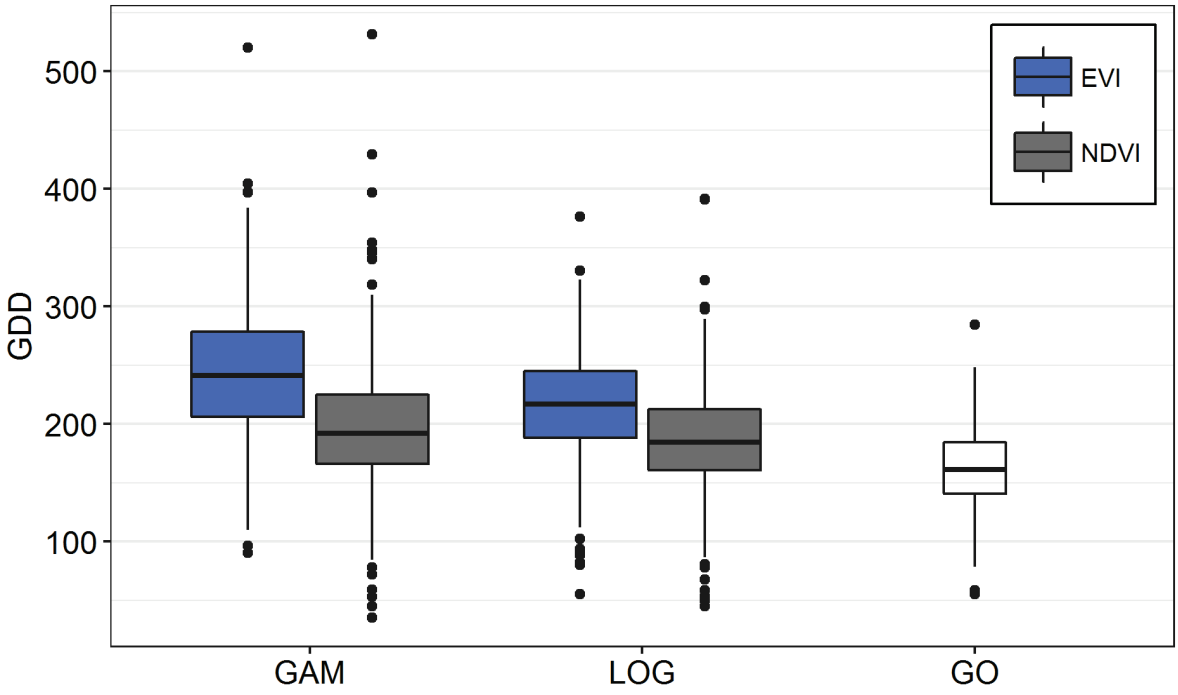


Fig. 10: Boxplots showing the median (horizontal black line), inter-quartile range (box), ±1.5-times inter-quartile range (black lines), and outliers (points) of GDD derived from the four model-index combinations (GAM, LOG) and ground observations (GO).

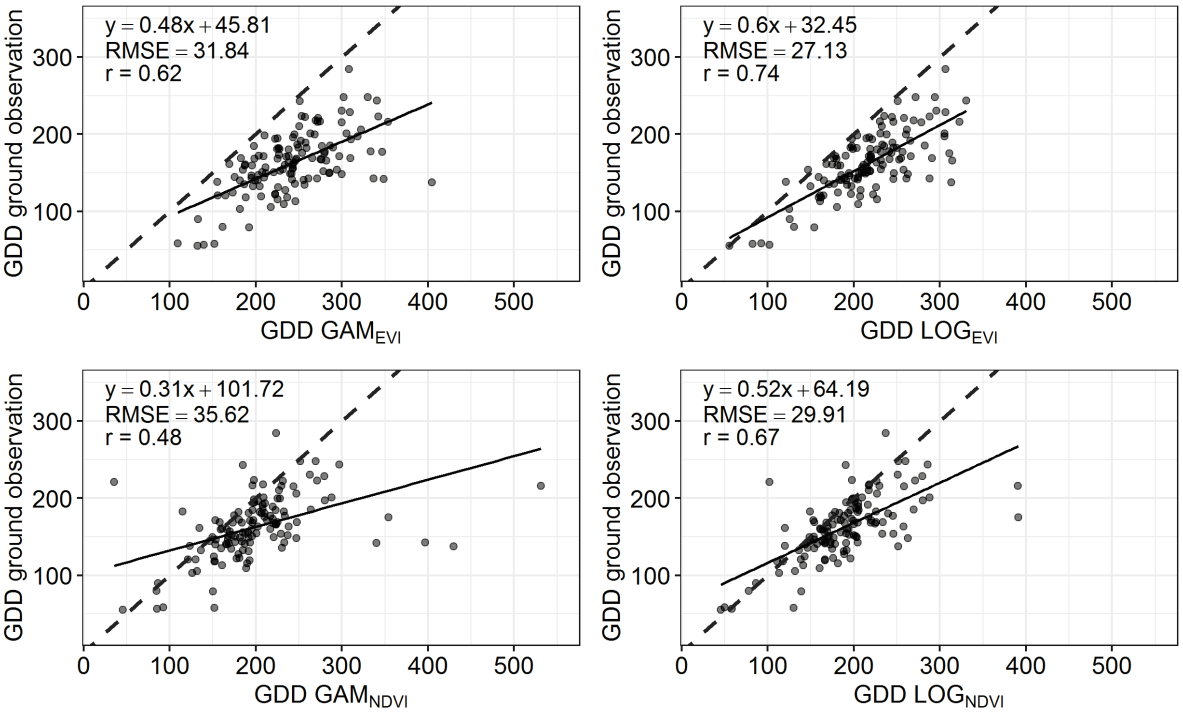


Fig. 11: GDD from four model and index combinations compared to GDD from ground observations with fitted linear regression (black: regression line, dashed grey: identity line).

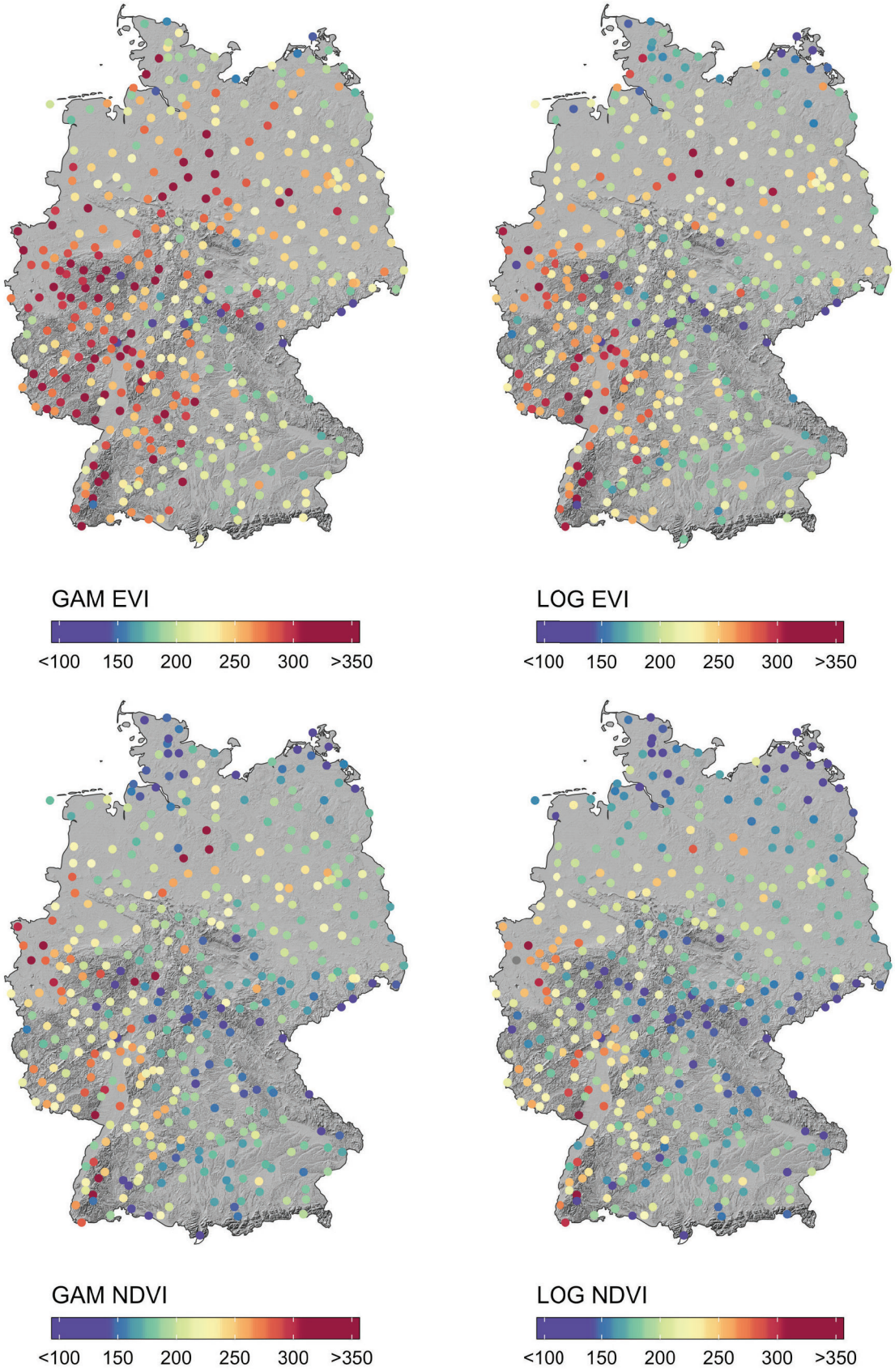


Fig. 12: GDD from TT model using SOS estimates from Landsat/Sentinel-2 time series (grey points: no data).

4 Discussion

4.1 SOS estimates from integrated Landsat/Sentinel-2 time series

In the first part of this study, we analyzed the feasibility of different model and vegetation index combinations to estimate spring phenology of temperate broadleaf forests from dense, medium resolution multi-sensor satellite time series. Overall, we tested two different models (logistic model and GAM using thin plate regression splines) with two vegetation indices (EVI and NDVI) as input. The results confirmed the applicability of both, the logistic model and the GAM to estimate SOS of temperate broadleaved forests. While logistic functions have been widely used in various studies (e.g. P. Jönsson et al., 2018; Melaas et al., 2013), considerably fewer studies have applied smoothing splines in order to derive spring phenology from medium resolution satellite time series (e.g. Melaas, Sulla-Menashe, et al., 2016). We found higher convergence rates of GAMs when fitting time series for all samples, pointing to higher flexibility compared to the logistic model. Regarding model choice though, further research is needed in order to assess the robustness of the estimates and connected to this, potential negative effects of limited observation density (P. Jönsson et al., 2018).

From our results we can infer that both, the logistic model and the GAMs, performed equally well, whereas regarding choice of vegetation index, EVI outperformed NDVI: We found higher correlations between the two EVI models and smaller MSE compared to NDVI models. Moreover, SOS estimated from EVI time series agreed better with in-situ dates of leaf unfolding despite a consistent time lag. While the majority of studies used NDVI for deriving spring phenology, higher applicability of EVI has also been shown for MODIS data (Klosterman et al., 2014). Hufkens, Friedl, Sonnentag, et al. (2012) also emphasize the impacts of different vegetation indices for retrieving phenological parameters from remote sensing time series. Moreover, NDVI tends to reproduce snowmelt dynamics rather than actual phenology in, e.g., higher latitude forests. Therefore, it has been suggested, that other vegetation indices are more suitable to derive phenological parameters (Jin, Jönsson, Bolmgren, Langvall, & Eklundh, 2017; A. M. Jönsson, Eklundh, Hellström, Bärring, & Jönsson, 2010).

4.2 Spatial variability of SOS estimates and ground observations

Our SOS estimates are in line with other studies, which found a moderate agreement of SOS from remote sensing and ground observations (Melaas, Sulla-Menashe, et al., 2016; Rodriguez-Galiano, Dash, & Atkinson, 2015). However, these studies did not find delayed SOS estimated

from remote sensing. In fact, Jin et al. (2017) and White et al. (2009) even report earlier SOS dates from MODIS and AVHRR using NDVI time series compared to ground observations.

Apart from differences between monitoring methods of ground observations, a possible explanation are differences in physiological measures as date of leaf unfolding does not directly translate to the measure which we determine as SOS date based on the spectral response of the surface. Moreover, in-situ observations target single species which are not necessarily representative of phenology as derived from remote sensing time series (Fisher & Mustard, 2007; White et al., 2009). In this regard, information on species composition, e.g., from forest inventory data, could be used to filter applicable ground observations of the respective species. Ground observations, which are part of the German Weather Service network, aim to monitor single trees outside larger forest stands (German Weather Service, 2015) and might therefore not represent conditions within larger forest stands. While we only used ground observations within 5,000m of each plot, the ground observations themselves might be up to 5,000m away from the denoted coordinate (German Weather Service, 2015), further increasing associated uncertainty. Besides, there is no possibility to evaluate potential measurement errors of in-situ observations. Adding to this, our sampling design was set up in a way that a mean SOS estimate was derived from 30 samples for an area of up to 78.54 km^2. In this area, differences in phenology are not unlikely to occur due to, e.g., differences in elevation or small-scale climate variability (Fisher et al., 2007), potentially leading to increasing inaccuracies in SOS estimates at the plot level.

Differences in elevations of samples belonging to the same plot could also be an explanation, why we observed very weak correlations of SOS and elevation as opposed to other studies. However, some of these studies are confined to smaller geographical extents in mountainous areas, where elevation is a major driver of spring phenology (e.g. Guyon et al., 2011; Senf et al., 2017). Since our study covers a much larger area, an elevation gradient might be driving phenology in distinct (e.g. mountainous) regions but might be less important across the whole study area. Regarding the amount of urban area, results indicated, as expected, an earlier SOS with increasing artificial land cover. This has also been reported by other studies (Melaas, Wang, et al., 2016; Zhang, Friedl, Schaaf, Strahler, & Schneider, 2004). Here again, we can assume that correlations are not as strong as in these studies because study areas were smaller and limited to regions with rural-urban gradients. We observed no east-west gradient in SOS estimates, suggesting that a large-scale east-west gradient is overruled by small-scale variation of other environmental drivers. In line with this, SOS estimates were not correlated with the

amount of forcing evident at each plot. Overall, weak to moderate correlation strengths of SOS to different drivers can therefore be explained by locally varying impacts of the respective environmental factors. We showed that medium resolution time series from Landsat and Sentinel-2 are able to detect spatial variation caused by different environmental drivers.

4.3 Differences in SOS estimates and environmental drivers

We evaluated the results of two mechanistic models and their differences to SOS estimates from Landsat and Sentinel-2 time series. We found a moderate agreement between SOS derived from Sentinel-2 and Landsat data and SOS from two mechanistic models. However, we used simple representations of chilling and thermal forcing in both mechanistic models. Consequently, more work is needed in order to, e.g., assess different responses to temperature such as sigmoid or bell-shaped functions (Basler, 2016). Furthermore, day- and nighttime temperatures could be used in mechanistic models since it has been suggested that daytime temperatures exert stronger effects on spring phenology than nighttime temperatures (Fu et al., 2016; Piao et al., 2015). We observed that mechanistic models using literature-based values for Germany were not able to entirely reproduce spatial patterns of spring phenology as observed from Landsat and Sentinel-2 time series.

In fact, the relations of SOS dates from remote sensing and mechanistic models to environmental gradients suggest that local factors, which are not represented in mechanistic models, influence spring phenology in our study area and drive spatial variability on smaller spatial scales. Large shares of model differences between SOS from mechanistic models and from Landsat/Sentinel-2 time series were explained by spatial differences in forcing and chilling but also other environmental drivers such as east-west gradients and amount of urban area. In line with this, our analysis indicated that GDD and CD, derived from the TT model using SOS estimates, exhibit considerable variability across our study area. Accordingly, it has been shown that forcing and chilling requirements vary spatially and also among individuals of the same species (Kramer et al., 2017). These results also hint to a strong variability of local climatic influences as described by e.g. Fisher, Mustard, & Vadeboncoeur (2006). As opposed to, e.g., MODIS data on coarser spatial resolution, it is possible to detect small-scale variation of spring phenology (Fisher et al., 2006) and therefore also forcing and chilling with medium resolution time series. As suggested in several studies (e.g. Chuine et al., 1999; Dantec et al., 2014), we also found that chilling and forcing are negatively correlated. Our results underline

the need to enhance the representation of spatial variability of thermal forcing and chilling in mechanistic models.

Regarding the agreement of GDD from ground observations and Landsat and Sentinel-2 time series, we found moderate to strong correlations. More GDD were estimated by remote sensing models, reflecting the delay of SOS estimates from remote sensing time series. GDD from in-situ observations agreed better with GDD from logistic models than with GDD from GAMs. While there is no well-founded explanation possible at this point, further evaluation of both remote sensing models might elucidate these results. In the past, ground observations have been widely used to calibrate mechanistic models (e.g. Basler, 2016; Linkosalo et al., 2008; Schaber & Badeck, 2003). However, there is evidence that resulting models are not transferable to other study areas (Richardson et al., 2006). Therefore, medium resolution remote sensing time series present a spatially continuous alternative in areas where phenological ground observations are scarce (Fitchett, Grab, & Thompson, 2015). To derive GDD and CD from the mechanistic models, we used fixed values of base temperature (T_{b1}, T_{b2}) and starting day of forcing and chilling accumulation (t_f, t_c, respectively). However, Fisher et al. (2007) showed that optimal values for both, base temperature and starting day of heat accumulation, vary spatially. It remains to be examined to what extent this applies to our study area. Moreover, uncertainties of GDD and CD values could not be quantified because we used SOS estimates for a single year. In order to draw further conclusions about the reliability of GDD and CD estimates, multiannual time series of Landsat and Sentinel-2 data could be used to calibrate mechanistic models.

5 Conclusion

In phenological studies, various data sources and approaches have been applied in the past decades to expand existing knowledge. However, a complete understanding of the mechanistic processes of forest phenology is still missing. Therefore, Tang et al. (2016) stressed the need to bridge the gap between different disciplines of phenological studies in order to assess spatial and temporal variability as well as drivers of phenological processes. This study presents a first effort to integrate mechanistic models and spring phenology estimates from medium resolution multi-sensor time series.

With the combination of current medium resolution sensors emerges the potential to derive phenological parameters across large spatial extents and with high spatial detail. In this study we used dense time series consisting of Landsat and Sentinel-2 data, to model and characterize spring phenology of temperate broadleaf forests in Germany for the year 2017. By comparing two commonly used models and vegetation indices, we showed that the choice of vegetation index has a higher impact on the resulting SOS estimates than model choice. Both, the logistic model and the generalized additive model estimated spring phenology consistently. Regarding the vegetation index however, EVI was more feasible for deriving spring phenology of temperate broadleaf forests than NDVI. The advantages of the application of medium resolution remote sensing time series include the opportunity to obtain phenological estimates for large areas and to capture regional to local phenological variation as well. Various studies emphasize the limited comparability of in-situ observations and phenological parameters estimated from remote sensing (e.g. Fisher & Mustard, 2007; Misra et al., 2016). Our results underline that differences in the meaning of both measures exist for our study area and should therefore be examined and compared carefully.

While we were able to derive SOS and GDD estimates for the year 2017, more research is needed in order to quantify uncertainties of SOS and GDD estimates by calibrating mechanistic models with spring phenology estimates from multiannual time series. Currently, process-based models are used in terrestrial biosphere models but fail to predict phenological key parameters such as season onset reliably (Richardson et al., 2012). Consequently, we showed that spring phenology estimates from Landsat and Sentinel-2 time series varied locally and were not reproduced by mechanistic models. Differences were mainly explained by spatial variability of thermal forcing and chilling. Therefore, remote sensing-based estimates of spring phenology have a high potential for calibrating mechanistic models locally but also across large spatial

extents which can ultimately enhance the representation of spring phenology in terrestrial biosphere models.

This study is a first step towards an improved representation of spring phenology in mechanistic models. Combined dense, medium resolution time series from sensors such as Sentinel-2 and the Landsat ensemble provide the opportunity to monitor and model vegetation phenology across large areas. Ultimately, this can be used to better understand the complex impacts which future climatic conditions will have on ecosystems globally.

Acknowledgements

This study was conducted as final thesis in the master program Global Change Geography at Humboldt-University of Berlin. I want to especially thank my supervisors, Dr. Dirk Pflugmacher and Dr. Cornelius Senf, who supported and helped me with frequent discussions, valuable advice and continuous encouragement.

References

Basler, D. (2016). Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe. *Agricultural and Forest Meteorology*, *217*, 10–21. http://doi.org/10.1016/j.agrformet.2015.11.007

Basler, D., & Körner, C. (2012). Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology*, *165*, 73–81. http://doi.org/10.1016/j.agrformet.2012.06.001

Basler, D., & Körner, C. (2014). Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree Physiology*, *34*(4), 377–388. http://doi.org/10.1093/treephys/tpu021

Buitenwerf, R., Rose, L., & Higgins, S. I. (2015). Three decades of multi-dimensional change in global leaf phenology. *Nature Climate Change*, *5*(4), 364–368. http://doi.org/10.1038/nclimate2533

Bundesministerium für Ernährung und Landwirtschaft. (2018). *Der Wald in Deutschland - Ausgewählte Ergebnisse der dritten Bundeswaldinventur* (3rd ed.). Retrieved from https://www.bmel.de/SharedDocs/Downloads/Broschueren/Bundeswaldinventur3.pdf;jsessionid=84E6512374113C9F13DFF6D490CBD396.1_cid358?__blob=publicationFile

Chuine, I. (2000). A Unified Model for Budburst of Trees. *Journal of Theoretical Biology*, *207*(3), 337–347. http://doi.org/10.1006/jtbi.2000.2178

Chuine, I. (2010). Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1555), 3149–3160. http://doi.org/10.1098/rstb.2010.0142

Chuine, I., & Beaubien, E. G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters*, *4*(5), 500–510. http://doi.org/10.1046/j.1461-0248.2001.00261.x

Chuine, I., Bonhomme, M., Legave, J.-M., García de Cortázar-Atauri, I., Charrier, G., Lacointe, A., & Améglio, T. (2016). Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Global Change Biology*, *22*(10), 3444–3460. http://doi.org/10.1111/gcb.13383

Chuine, I., Cour, P., & Rousseau, D. D. (1999). Selecting models to predict the timing of flowering of temperate trees: implications for tree phenology modelling. *Plant, Cell & Environment*, *22*(1), 1–13. http://doi.org/10.1046/j.1365-3040.1999.00395.x

Cook, B. I., Wolkovich, E. M., & Parmesan, C. (2012). Divergent responses to spring and winter warming drive community level flowering trends. *Proceedings of the National Academy of Sciences*, *109*(23), 9000–9005. http://doi.org/10.1073/pnas.1118364109

Crabbe, R. A., Dash, J., Rodriguez-Galiano, V. F., Janous, D., Pavelka, M., & Marek, M. V. (2016). Extreme warm temperatures alter forest phenology and productivity in Europe. *Science of The Total Environment*, *563–564*, 486–495. http://doi.org/10.1016/j.scitotenv.2016.04.124

Čufar, K., De Luis, M., Saz, M. A., Črepinšek, Z., & Kajfež-Bogataj, L. (2012). Temporal shifts in leaf phenology of beech (Fagus sylvatica) depend on elevation. *Trees*, *26*(4), 1091–1100. http://doi.org/10.1007/s00468-012-0686-7

Dallimer, M., Tang, Z., Gaston, K. J., & Davies, Z. G. (2016). The extent of shifts in vegetation phenology between rural and urban areas within a human-dominated region. *Ecology and Evolution*, *6*(7), 1942–1953. http://doi.org/10.1002/ece3.1990

Dantec, C. F., Vitasse, Y., Bonhomme, M., Louvet, J.-M., Kremer, A., & Delzon, S. (2014). Chilling and heat requirements for leaf unfolding in European beech and sessile oak populations at the southern limit of their distribution range. *International Journal of Biometeorology*, *58*(9), 1853–1864. http://doi.org/10.1007/s00484-014-0787-7

Delpierre, N., Vitasse, Y., Chuine, I., Guillemot, J., Bazot, S., Rutishauser, T., & Rathgeber, C. B. K. (2016). Temperate and boreal forest tree phenology: from organ-scale processes to terrestrial ecosystem models. *Annals of Forest Science*, *73*(1), 5–25. http://doi.org/10.1007/s13595-015-0477-6

DWD Climate Data Center (CDC). (2018). Daily station observations of mean temperature at 2 m above ground in °C for Germany, version v18.3 & recent.

Fisher, J., & Mustard, J. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment*, *109*(3), 261–273. http://doi.org/10.1016/j.rse.2007.01.004

Fisher, J., Mustard, J., & Vadeboncoeur, M. (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment*, *100*(2), 265–279. http://doi.org/10.1016/j.rse.2005.10.022

Fisher, J., Richardson, A. D., & Mustard, J. (2007). Phenology model from surface meteorology does not capture satellite-based greenup estimations. *Global Change Biology*, *13*(3), 707–

721. http://doi.org/10.1111/j.1365-2486.2006.01311.x

Fitchett, J. M., Grab, S. W., & Thompson, D. I. (2015). Plant phenology and climate change. *Progress in Physical Geography: Earth and Environment*, *39*(4), 460–482. http://doi.org/10.1177/0309133315578940

Frantz, D. (2018). *FORCE v. 2.0 - Technical User Guide*. http://doi.org/10.13140/RG.2.2.27918.41284/1

Frantz, D., Röder, A., Stellmes, M., & Hill, J. (2016). An Operational Radiometric Landsat Preprocessing Framework for Large-Area Time Series Applications. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(7), 3928–3943. http://doi.org/10.1109/TGRS.2016.2530856

Friedl, M. A., Gray, J. M., Melaas, E. K., Richardson, A. D., Hufkens, K., Keenan, T. F., … O'Keefe, J. (2014). A tale of two springs: Using recent climate anomalies to characterize the sensitivity of temperate forest phenology to climate change. *Environmental Research Letters*, *9*(5). http://doi.org/10.1088/1748-9326/9/5/054006

Fu, Y. H., Liu, Y., De Boeck, H. J., Menzel, A., Nijs, I., Peaucelle, M., … Janssens, I. A. (2016). Three times greater weight of daytime than of night-time temperature on leaf unfolding phenology in temperate trees. *New Phytologist*, *212*(3), 590–597. http://doi.org/10.1111/nph.14073

Fu, Y. H., Piao, S., Delpierre, N., Hao, F., Hänninen, H., Liu, Y., … Campioli, M. (2018). Larger temperature response of autumn leaf senescence than spring leaf-out phenology. *Global Change Biology*, *24*(5), 2159–2168. http://doi.org/10.1111/gcb.14021

Fu, Y. H., Zhao, H., Piao, S., Peaucelle, M., Peng, S., Zhou, G., … Janssens, I. A. (2015). Declining global warming effects on the phenology of spring leaf unfolding. *Nature*, *526*, 104–107. http://doi.org/10.1038/nature15402

Garonna, I., de Jong, R., Stöckli, R., Schmid, B., Schenkel, D., Schimel, D., & Schaepman, M. E. (2018). Shifting relative importance of climatic constraints on land surface phenology. *Environmental Research Letters*, *13*(2). http://doi.org/10.1088/1748-9326/aaa17b

German Weather Service. (2013). *Die Deutschen Klimabeobachtungssysteme. Inventarbericht zum Global Climate Observing System (GCOS)*. Offenbach: Selbstverlag des Deutschen Wetterdienstes.

German Weather Service. (2015). *Vorschriften und Betriebsunterlagen für die phänologischen*

Beobachter des Deutschen Wetterdienstes. Offenbach. Retrieved from https://www.dwd.de/DE/klimaumwelt/klimaueberwachung/phaenologie/daten_deutschland/beobachtersuche/beobachteranleitung_download.pdf?__blob=publicationFile&v=9

Gill, A. L., Gallinat, A. S., Sanders-DeMott, R., Rigden, A. J., Short Gianotti, D. J., Mantooth, J. A., & Templer, P. H. (2015). Changes in autumn senescence in northern hemisphere deciduous trees: a meta-analysis of autumn phenology studies. *Annals of Botany*, *116*(6), 875–888. http://doi.org/10.1093/aob/mcv055

Guyon, D., Guillot, M., Vitasse, Y., Cardot, H., Hagolle, O., Delzon, S., & Wigneron, J.-P. (2011). Monitoring elevation variations in leaf phenology of deciduous broadleaf forests from SPOT/VEGETATION time-series. *Remote Sensing of Environment*, *115*(2), 615–627. http://doi.org/10.1016/j.rse.2010.10.006

Han, Q., Wang, T., Jiang, Y., Fischer, R., & Li, C. (2018). Phenological variation decreased carbon uptake in European forests during 1999–2013. *Forest Ecology and Management*, *427*, 45–51. http://doi.org/10.1016/j.foreco.2018.05.062

Hänninen, H. (1990). Modelling bud dormancy release in trees from cool and temperate regions. *Acta Forestalia Fennica*, *213*, 1–47. http://doi.org/10.14214/aff.7660

Horvath, D. P., Anderson, J. V., Chao, W. S., & Foley, M. E. (2003). Knowing when to grow: signals regulating bud dormancy. *Trends in Plant Science*, *8*(11), 534–540. http://doi.org/10.1016/j.tplants.2003.09.013

Hufkens, K., Friedl, M. A., Keenan, T. F., Sonnentag, O., Bailey, A., O'Keefe, J., & Richardson, A. D. (2012). Ecological impacts of a widespread frost event following early spring leaf-out. *Global Change Biology*, *18*(7), 2365–2377. http://doi.org/10.1111/j.1365-2486.2012.02712.x

Hufkens, K., Friedl, M., Sonnentag, O., Braswell, B. H., Milliman, T., & Richardson, A. D. (2012). Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment*, *117*, 307–321. http://doi.org/10.1016/j.rse.2011.10.006

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. *Global Change Biology*, *17*(7), 2385–2399. http://doi.org/10.1111/j.1365-2486.2011.02397.x

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 Sensing of Environment*, *198*, 203–212. http://doi.org/10.1016/j.rse.2017.06.015

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 Sensing of Environment*, *114*(11), 2719–2730. http://doi.org/10.1016/j.rse.2010.06.005

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 Sensing*, *10*(4), 635. http://doi.org/10.3390/rs10040635

Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30*(8), 833–845. http://doi.org/10.1016/j.cageo.2004.05.006

Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., … Richardson, A. D. (2012). Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange. *Global Change Biology*, *18*(6), 1971–1987. http://doi.org/10.1111/j.1365-2486.2012.02678.x

Keenan, T. F., Gray, J., Friedl, M. A., Toomey, M., Bohrer, G., Hollinger, D. Y., … Richardson, A. D. (2014). Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change*, *4*, 598. http://doi.org/10.1038/nclimate2253

Keenan, T. F., & Richardson, A. D. (2015). The timing of autumn senescence is affected by the timing of spring phenology: implications for predictive models. *Global Change Biology*, *21*(7), 2634–2641. http://doi.org/10.1111/gcb.12890

Klosterman, S. T., Hufkens, K., Gray, J. M., Melaas, E., Sonnentag, O., Lavine, I., … Richardson, A. D. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, *11*(16), 4305–4320. http://doi.org/10.5194/bg-11-4305-2014

Kramer, K., Ducousso, A., Gömöry, D., Hansen, J. K., Ionita, L., Liesebach, M., … von Wühlisch, G. (2017). Chilling and forcing requirements for foliage bud burst of European

beech (Fagus sylvatica L.) differ between provenances and are phenotypically plastic. *Agricultural and Forest Meteorology*, *234–235*, 172–181. http://doi.org/10.1016/j.agrformet.2016.12.002

Lang, G. A., Early, J. D., Martin, G. C., & Darnell, R. L. (1987). Endo-, para-, and ecodormancy: Physiological terminology and classification for dormancy research. *HortScience*, *22*(3), 371–377.

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 Sensing*, *9*(9), 902. http://doi.org/10.3390/rs9090902

Li, X., Zhou, Y., Asrar, G. R., Mao, J., Li, X., & Li, W. (2017). Response of vegetation phenology to urbanization in the conterminous United States. *Global Change Biology*, *23*(7), 2818–2830. http://doi.org/10.1111/gcb.13562

Liang, L. (2016). Beyond the Bioclimatic Law: Geographic adaptation patterns of temperate plant phenology. *Progress in Physical Geography: Earth and Environment*, *40*(6), 811–834. http://doi.org/10.1177/0309133316656558

Liang, L., & Schwartz, M. D. (2009). Landscape phenology: an integrative approach to seasonal vegetation dynamics. *Landscape Ecology*, *24*(4), 465–472. http://doi.org/10.1007/s10980-009-9328-x

Linkosalo, T., Carter, T. R., Häkkinen, R., & Hari, P. (2000). Predicting spring phenology and frost damage risk of Betula spp. under climatic warming: a comparison of two models. *Tree Physiology*, *20*(17), 1175–1182. http://doi.org/10.1093/treephys/20.17.1175

Linkosalo, T., Lappalainen, H. K., & Hari, P. (2008). A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations. *Tree Physiology*, *28*(12), 1873–1882. http://doi.org/10.1093/treephys/28.12.1873

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 457–465. http://doi.org/10.1109/36.377946

Liu, L., Liang, L., Schwartz, M. D., Donnelly, A., Wang, Z., Schaaf, C. B., & Liu, L. (2015). Evaluating the potential of MODIS satellite data to track temporal dynamics of autumn phenology in a temperate mixed forest. *Remote Sensing of Environment*, *160*, 156–165. http://doi.org/10.1016/j.rse.2015.01.011

Luyssaert, S., Schulze, E.-D., Börner, A., Knohl, A., Hessenmöller, D., Law, B. E., … Grace, J. (2008). Old-growth forests as global carbon sinks. *Nature*, *455*, 213–215. http://doi.org/10.1038/nature07276

Markham, B. L., Storey, J. C., Williams, D. L., & Irons, J. R. (2004). Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(12), 2691–2694. http://doi.org/10.1109/TGRS.2004.840720

Meier, U. (2018). *Growth stages of mono- and dicotyledonous plants: BBCH Monograph*. Open Agrar Repositorum. http://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 Sensing of Environment*, *132*, 176–185. http://doi.org/10.1016/j.rse.2013.01.011

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. *Geophysical Research Letters*, *45*(6), 2679–2687. http://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 Sensing of Environment*, *186*, 452–464. http://doi.org/10.1016/j.rse.2016.09.014

Melaas, E. K., Wang, J. A., Miller, D. L., & Friedl, M. A. (2016). Interactions between urban vegetation and surface urban heat islands: a case study in the Boston metropolitan region. *Environmental Research Letters*, *11*(5), 54020. http://doi.org/10.1088/1748-9326/11/5/054020

Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., … Zust, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology*, *12*(10), 1969–1976. http://doi.org/10.1111/j.1365-2486.2006.01193.x

Misra, G., Buras, A., & Menzel, A. (2016). Effects of Different Methods on the Comparison between Land Surface and Ground Phenology—A Methodological Case Study from South-Western Germany. *Remote Sensing*, *8*(9), 753. http://doi.org/10.3390/rs8090753

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*, 37. http://doi.org/10.1038/nature01286

Perry, T. O. (1971). Dormancy of Trees in Winter. *Science*, *171*(3966), 29–36. http://doi.org/10.1126/science.171.3966.29

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 Sensing of Environment*, *221*, 583–595. http://doi.org/10.1016/j.rse.2018.12.001

Pflugmacher, D., Senf, C., Yang, Z., Seidl, R., & Hostert, P. (2019). Characterizing disturbance regimes of Central European forests using remote sensing. *Manuscript in Preparation*.

Piao, S., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luyssaert, S., … Vesala, T. (2008). Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature*, *451*, 49–52. http://doi.org/10.1038/nature06444

Piao, S., Tan, J., Chen, A., Fu, Y. H., Ciais, P., Liu, Q., … Peñuelas, J. (2015). Leaf onset in the northern hemisphere triggered by daytime temperature. *Nature Communications*, *6*(1), 6911. http://doi.org/10.1038/ncomms7911

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Comuting.

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., … Xue, Y. (2012). Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis. *Global Change Biology*, *18*(2), 566–584. http://doi.org/10.1111/j.1365-2486.2011.02562.x

Richardson, A. D., Andy Black, T., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., … Varlagin, A. (2010). Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1555), 3227–3246. http://doi.org/10.1098/rstb.2010.0102

Richardson, A. D., Bailey, A. S., Denny, E. G., Martin, C. W., & O'Keefe, J. (2006). Phenology of a northern hardwood forest canopy. *Global Change Biology*, *12*(7), 1174–1188. http://doi.org/10.1111/j.1365-2486.2006.01164.x

Rodriguez-Galiano, V. F., Dash, J., & Atkinson, P. M. (2015). Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. *Geophysical Research Letters*, *42*(7), 2253–2260. http://doi.org/10.1002/2015GL063586

Rohde, A., & Bhalerao, R. P. (2007). Plant dormancy in the perennial context. *Trends in Plant Science*, *12*(5), 217–223. http://doi.org/10.1016/j.tplants.2007.03.012

Rosenzweig, C., Casassa, G., Karoly, D. J., Imeson, A., Liu, C., Menzel, A., … Tryjanowski, P. (2007). Assessment of observed changes and responses in natural and managed systems. In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, & C. E. Hanson (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 79–131). Cambridge, UK: Cambridge University Press.

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., … Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature*, *453*(7193), 353–357. http://doi.org/10.1038/nature06937

Schaber, J., & Badeck, F.-W. (2003). Physiology-based phenology models for forest tree species in Germany. *International Journal of Biometeorology*, *47*(4), 193–201. http://doi.org/10.1007/s00484-003-0171-5

Semmel, A. (1996). *Geomorphologie der Bundesrepublik Deutschland* (5th ed.). Stuttgart: Franz Steiner Verlag.

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 Sensing of Environment*, *194*, 155–160. http://doi.org/10.1016/j.rse.2017.03.020

Simpson, G. L. (2018). Modelling Palaeoecological Time Series Using Generalised Additive Models. *Frontiers in Ecology and Evolution*, *6*, 149. http://doi.org/10.3389/fevo.2018.00149

Stephenson, N. L., Das, A. J., Condit, R., Russo, S. E., Baker, P. J., Beckman, N. G., … Zavala, M. A. (2014). Rate of tree carbon accumulation increases continuously with tree size. *Nature*, *507*, 90. http://doi.org/10.1038/nature12914

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*(8). http://doi.org/10.1002/ecs2.1436

Templ, B., Koch, E., Bolmgren, K., Ungersböck, M., Paul, A., Scheifinger, H., … Zust, A. (2018). Pan European Phenological database (PEP725): a single point of access for European data. *International Journal of Biometeorology*, *62*(6), 1109–1113. http://doi.org/10.1007/s00484-018-1512-8

Tucker, C. T. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, *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. *Ecological Modelling*, *323*, 123–132. http://doi.org/10.1016/j.ecolmodel.2015.12.005

Vitasse, Y., Signarbieux, C., & Fu, Y. H. (2018). Global warming leads to more uniform spring phenology across elevations. *Proceedings of the National Academy of Sciences*, *115*(5), 1004–1008. http://doi.org/10.1073/pnas.1717342115

Wang, J., Wu, C., Zhang, C., Ju, W., Wang, X., Chen, Z., & Fang, B. (2018). Improved modeling of gross primary productivity (GPP) by better representation of plant phenological indicators from remote sensing using a process model. *Ecological Indicators*, *88*, 332–340. http://doi.org/10.1016/j.ecolind.2018.01.042

Wesołowski, T., & Rowiński, P. (2006). Timing of bud burst and tree-leaf development in a multispecies temperate forest. *Forest Ecology and Management*, *237*(1–3), 387–393. http://doi.org/10.1016/j.foreco.2006.09.061

White, M. A., de Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., … Lauenroth, W. K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, *15*(10), 2335–2359. http://doi.org/10.1111/j.1365-2486.2009.01910.x

White, M. A., & Nemani, R. R. (2006). Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment*, *104*(1), 43–49. http://doi.org/10.1016/j.rse.2006.04.014

Winter, S., Höfler, J., Michel, A. K., Böck, A., & Ankerst, D. P. (2015). Association of tree and plot characteristics with microhabitat formation in European beech and Douglas-fir forests. *European Journal of Forest Research*, *134*(2), 335–347. http://doi.org/10.1007/s10342-014-0855-x

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114. http://doi.org/10.1111/1467-9868.00374

Wu, C., Wang, X., Wang, H., Ciais, P., Peñuelas, J., Myneni, R. B., … Ge, Q. (2018). Contrasting responses of autumn-leaf senescence to daytime and night-time warming.

Nature Climate Change, *8*(12), 1092–1096. http://doi.org/10.1038/s41558-018-0346-z

Yun, J., Jeong, S.-J., Ho, C.-H., Park, C.-E., Park, H., & Kim, J. (2018). Influence of winter precipitation on spring phenology in boreal forests. *Global Change Biology*, *24*(11), 5176–5187. http://doi.org/10.1111/gcb.14414

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., … Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*(3), 471–475. http://doi.org/10.1016/S0034-4257(02)00135-9

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., & Schneider, A. (2004). The footprint of urban climates on vegetation phenology. *Geophysical Research Letters*, *31*(12). http://doi.org/10.1029/2004GL020137

Zhou, D., Zhao, S., Zhang, L., & Liu, S. (2016). Remotely sensed assessment of urbanization effects on vegetation phenology in China's 32 major cities. *Remote Sensing of Environment*, *176*, 272–281. http://doi.org/10.1016/j.rse.2016.02.010

Zipper, S. C., Schatz, J., Singh, A., Kucharik, C. J., Townsend, P. A., & Loheide, S. P. (2016). Urban heat island impacts on plant phenology: intra-urban variability and response to land cover. *Environmental Research Letters*, *11*(5), 54023. http://doi.org/10.1088/1748-9326/11/5/054023

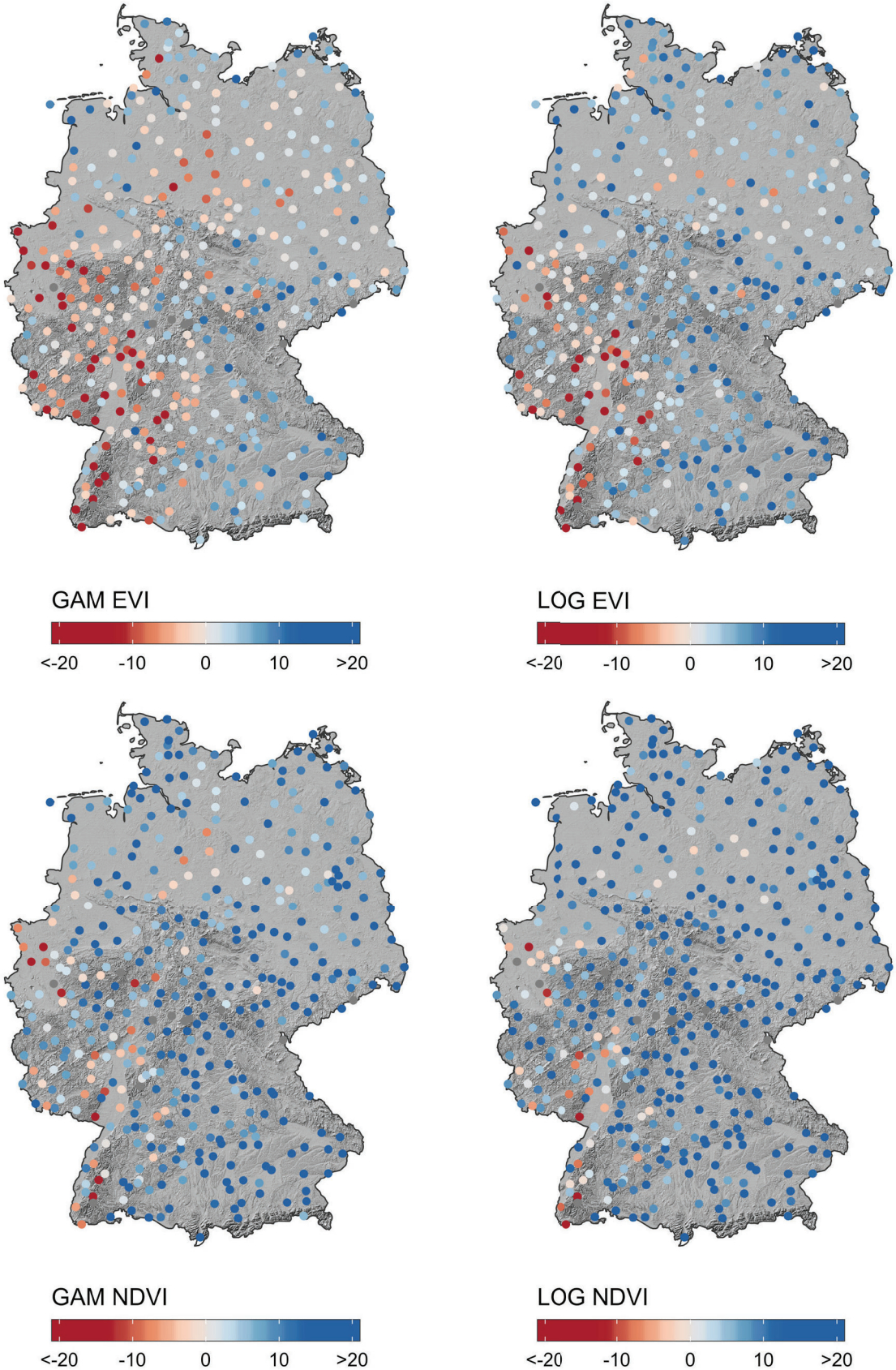
Zöller, L., Beierkuhnlein, C., Samimi, C., & Faust, D. (2017). *Die Physische Geographie Deutschlands*. (L. Zöller, Ed.). Darmstadt: Wissenschaftliche Buchgesellschaft (WBG).

Appendix

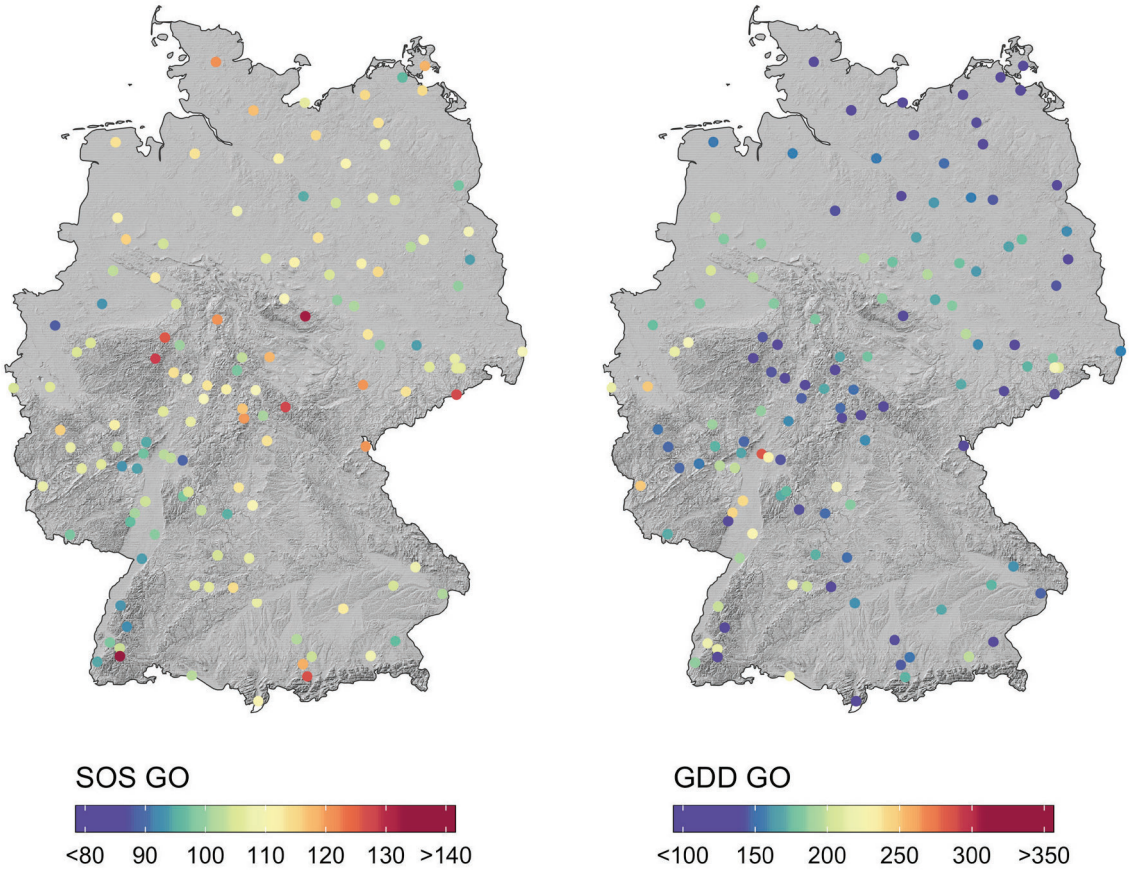
App. A: List of species and corresponding phenological ground observations.

Species	**No. of observations (BBCH = 11)**
Aesculus hippocastanum	1,028
Alnus glutinosa	854
Betula pendula	1,033
Fagus sylvatica	943
Fraxinus excelsior	836
Quercus robur	974
Sum of observations	5,668

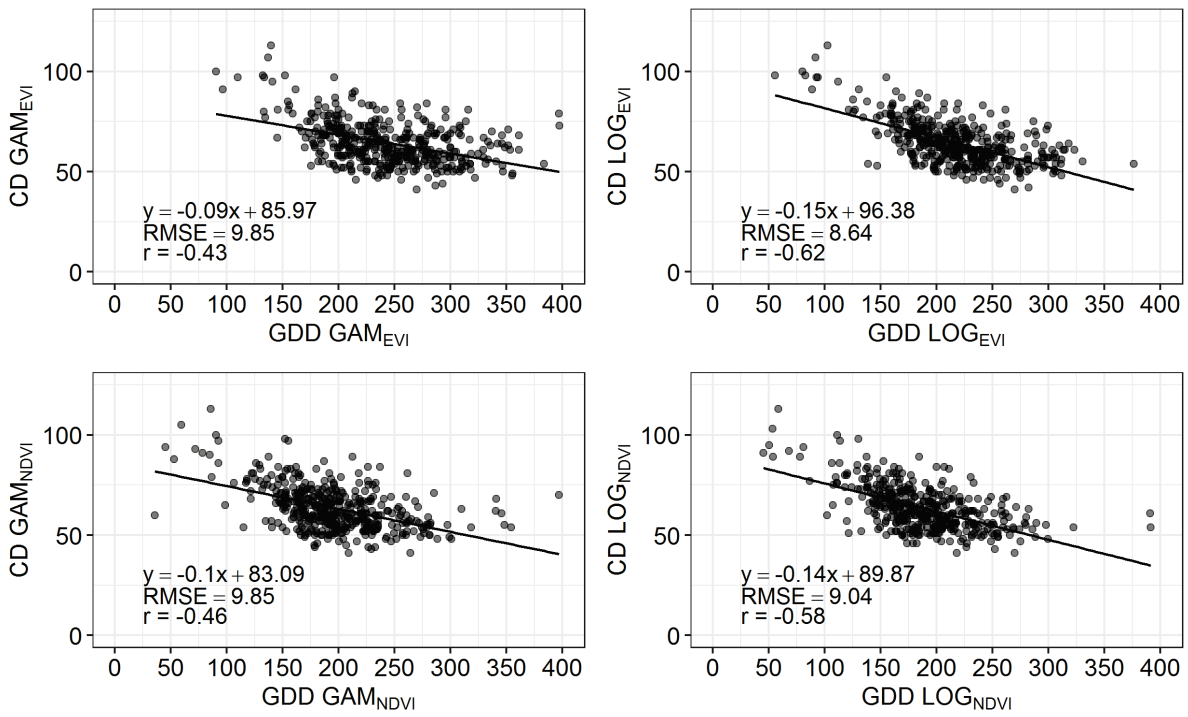
App. B: Differences of SOS estimates (in days) from all four model and index combinations and SOS estimates from the Sequential Model (grey points: no data).



App. C: SOS (left) and GDD estimates (right) using ground observations of leaf unfolding.



App. D: Relation of GDD and CD across all plots.



App. E: Average chilling days (CD) for every plot (grey points: no data).