Phenological Metrics Derived over the European Continent from NDVI3g Data and MODIS Time Series

Clement Atzberger *, Anja Klisch, Matteo Mattiuzzi and Francesco Vuolo

Institute for Surveying, Remote Sensing and Land Information, University of Natural Resources and Life Sciences (BOKU), Peter Jordan Strasse 82, A-1190 Vienna, Austria;
E-Mails: anja.klisch@boku.ac.at (A.K.); matteo.mattiuzzi@boku.ac.at (M.M.);
francesco.vuolo@boku.ac.at (F.V.)

* Author to whom correspondence should be addressed; E-Mail: clement.atzberger@boku.ac.at;
Tel.: +43-1-4576-5101; Fax: +43-1-4576-3902.

Received: 22 July 2013; in revised form: 21 November 2013 / Accepted: 25 November 2013 / Published: 27 December 2013

Abstract: Time series of normalized difference vegetation index (NDVI) are important data sources for environmental monitoring. Continuous efforts are put into their production and updating. The recently released Global Inventory Modeling and Mapping Studies (GIMMS) NDVI3g data set is a consistent time series with 1/12° spatial and bi-monthly temporal resolution. It covers the time period from 1981 to 2011. However, it is unclear if vegetation density and phenology derived from GIMMS are comparable to those obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI with 250 m ground resolution. To check the consistency between GIMMS and MODIS data sets, a comparative analysis was performed. For a large European window (40 × 40°), data distribution, spatial and temporal agreement were analyzed, as well as the timing of important phenological events. Overall, only a moderately good agreement of NDVI values was found. Large differences occurred during winter. Large discrepancies were also observed for phenological metrics, in particular the start of season. Information regarding the maximum of season was more consistent. Hence, both data sets should be well inter-calibrated before being used concurrently.

Keywords: MODIS; GIMMS; AVHRR; consistency analysis; Whittaker smoother; phenology; start of season; maximum of season; vegetation density; NDVI
1. Introduction

Time series of remotely sensed normalized difference vegetation index (NDVI) are valuable data sets in various Earth science fields. For example, NDVI time series have been successfully used for environmental monitoring [1–3], for providing agricultural outlooks and yield predictions [4–6], and for modeling wildlife occurrence/movement and biodiversity [7]. Such time series are also excellent and unique proxies for assessing possible climate change impacts across the globe [8,9] and for large scale drought monitoring [10]. A number of prominent applications are listed in Table 1 ([11–43]). Additional examples are presented within this special issue. Most applications build on the fact that NDVI is a (coarse) measure of vegetation density and more or less linearly related to the fraction of absorbed photosynthetically active radiation (fAPAR).

Table 1. Prominent applications building on time series of vegetation indices.

<table>
<thead>
<tr>
<th>Type of Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonality extraction and vegetation dynamics</td>
<td>[1–3,11–13]</td>
</tr>
<tr>
<td>Environmental monitoring and climate change</td>
<td>[8,9,14–17]</td>
</tr>
<tr>
<td>Land use and land cover mapping</td>
<td>[18–23]</td>
</tr>
<tr>
<td>Vegetation and landscape characterization</td>
<td>[24–29]</td>
</tr>
<tr>
<td>Biodiversity and wildlife distribution</td>
<td>[4,7,30–33]</td>
</tr>
<tr>
<td>Primary production and yield/production estimates</td>
<td>[4,6,34,35]</td>
</tr>
<tr>
<td>Drought monitoring and food security</td>
<td>[10,36]</td>
</tr>
<tr>
<td>Change detection (incl. natural disasters, forest fires)</td>
<td>[37–39]</td>
</tr>
<tr>
<td>Characterizing vegetation’s biophysical variables</td>
<td>[40–43]</td>
</tr>
</tbody>
</table>

The value of remotely sensed time series naturally increases with the accuracy of the measurement (i.e., radiometric and geometric accuracy). Equally important are, however, the length of the time series, its consistency, and temporal and spatial resolution. With the recently released Global Inventory Modeling and Mapping Studies NDVI3g (GIMMS) data set, a major breakthrough in harmonizing Advanced Very High Resolution Radiometer (AVHRR) data from various NOAA satellites (National Oceanic and Atmospheric Administration) was made [44]. The GIMMS data set spans the time period from 1981 to 2011, in 1/12° spatial resolution, and with bi-monthly temporal resolution. This offers interesting perspectives for a joint use with other NDVI data sets with higher spatial and temporal resolution but shorter length of the series. For example MODIS products at 250 m are only available from 2000 onwards. Before combining different NDVI time series, however, their consistency has to be checked [45,46]. The current paper addresses this issue and provides a comparative analysis of the GIMMS NDVI3g dataset and MODIS NDVI time series over a 40 × 40° European window for the time frame of 2002–2011. The comparative analysis covers several aspects of such time series and explicitly distinguishes between spatial and temporal agreement:

- data (frequency) distribution of vegetation density information (NDVI);
- agreement/disagreement and correlation of NDVI;
- timing of major phenological events (here start and maximum of season).
The study complements ongoing direct validation exercises, but is not intended to replace such efforts. A comparative analysis between existing Earth Observation (EO) products yields for example important information regarding the feasibility of creating historical records of NDVI and/or satellite-derived pheno-indicators. Phenological events were included in the study as pheno-indicators provide unique information for assessing possible climatic impacts on vegetation growth and the bio-sphere [1,3,13,47–49].

2. Data and Methods

The paper addresses the consistency of GIMMS and MODIS NDVI time series over Europe with a special focus on the timing of phenological events (start/maximum of season). The comparative analysis covers the years 2002–2011. Before analysis, we performed a careful harmonization of the two data sets in terms of temporal and spatial resolution. This also implied a smoothing/filtering of the two time series to remove undesired artifacts due to poor atmospheric conditions and undetected clouds, etc. Major pre-processing steps for spatial and temporal homogenization of the two time series are outlined in the following sub-sections and summarized in Table 2.

Table 2. Specification of the spatial and temporal pre-processing steps for homogenizing GIMMS and MODIS NDVI time series. The smoothing parameter (λ) of the Whittaker smoother determines the roughness of the smoothed curve. See text for further details.

<table>
<thead>
<tr>
<th>Parameterization of the Whittaker Smoother for Providing Smoothed Weekly Data</th>
<th>Spatial Resampling</th>
<th>Land/Water Masking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ x10^3</td>
<td>Degree</td>
</tr>
<tr>
<td>GIMMS</td>
<td>30</td>
<td>2nd</td>
</tr>
<tr>
<td>MODIS</td>
<td>30</td>
<td>2nd</td>
</tr>
</tbody>
</table>

The method used for the comparison capitalizes on previous studies involving either single image acquisitions or multi-temporal data [46,50]. In addition to analyzing the ‘overall’ agreement between paired data sets, we explicitly addressed—following the work of [46]—the seasonality of their spatial coherence and the spatial variability of their temporal agreement. Thus, both spatial and temporal patterns were analyzed within the present study. On the other hand, no attempts were made to determine which of the two series is of ‘better’ quality [46]. Seriously answering this question would only be possible with an extended network of NDVI reference measurements [51,52]. Such an extended reference network is currently not available. Existing prototype reference networks have only a reduced spatial and temporal coverage and are mainly focused on biophysical variables of vegetation such as leaf area index (LAI) and fraction of absorbed photosynthetically active radiation (fAPAR) [53,54]. Moreover, direct validation of coarse-resolution satellite data against point-like reference measurements is a nontrivial task [55]. Instead, the plausibility of seasonal and spatial pattern was checked by comparing with published results of other research teams.
2.1. Study Area

The study area is depicted in Figure 1 ([56–58]) together with some environmental variables. The selected variables demonstrate a very large spatial variability of the landscape as well as climatic conditions. The study region covers an area of approximately 18 million km\(^2\) (land surface: 12 million km\(^2\)). Only land pixels (see Section 2.3) are included in this study without any additional efforts to mask out inland water bodies (e.g., lakes and rivers). According to the Köppen climate classification [59], the climate in the area ranges from polar to arid. Precipitation and temperatures show a large gradient from the fully humid to very dry Sahara desert and from polar tundra to cold/hot arid. Land cover comprises a large number of different classes (mainly cropland, forest, shrubs, grassland and bare areas), and is generally fine structured [23,60]. This highly variable environmental setting permits comparing GIMMS and MODIS NDVI time series across various eco-zones and environments. Nonetheless, the findings of this study are necessarily restricted to the selected study area.

**Figure 1.** Study area and general environmental conditions. **(top left)** typical NDVI values in the study area in July, **(top right)** mean elevation [56], **(center left)** recoded land cover type 1 product MCD12Q1 v005 from MODIS [23,60], **(center right)** terrestrial ecoregions of World Wildlife Fund (WWF) [57], **(bottom left)** annual mean temperature [58], **(bottom right)** annual mean precipitation [58].
2.2. MODIS Dataset and Temporal Smoothing

We used the MOD13Q1 and MYD13Q1 NDVI collection 5 products of the MODIS Terra and Aqua satellites from 2000 onwards from the Land Processes Distributed Active Archive Center (LP DAAC). The data of both satellites are processed in a phased production providing a global NDVI with improved temporal frequency of eight days (Terra 16-day period starting Day 001, Aqua 16-day period starting Day 009). The gridded Level-3 product offers data in approximately 250-m spatial resolution in Sinusoidal projection. The Level-3 product is calculated from the Level-2G daily surface reflectance gridded data (MOD09 and MYD09 series) using the Constrained View angle-Maximum Value Composite (CV-MVC) compositing method [61].

For processing the data, we set-up in a first step an operational processing chain that downloads, mosaics, resamples, reprojects, and performs an image cropping to a dedicated tile system on a daily basis based on the “MODIS” R package [62]. The data are reprojected to geographic coordinates (datum WGS84) using nearest neighbor resampling. One file of the dedicated tile system covers the extent of 1 × 1°. The 448 × 448 pixels per tile have thus a spatial resolution of 0.002232°. The reprojected data are archived and further processed in two different ways: (1) an on-demand simplified filtering is applied upon user’s request via a web application [63], and (2) operational filtering procedures are applied for European areas with two processing lines: (i) an enhanced near real-time (NRT) weekly processing of new/updated NDVI data, and (ii) a processing of the entire NDVI time-series at once (launched only once a year). For this study, only the last data (ii) were considered. All procedures were implemented in MATLAB and are based on the Whittaker smoother [28,64,65]. Compared to our web application [63], the operational filtering procedures have the advantage to take the MODIS quality flag during smoothing into account, as well as the exact composite day. The Whittaker smoother was shown to perform well against a number of other filters based either on curve fitting or Fast Fourier transform (FFT) [66]. In [64], it was shown that this filter permits a significant increase in the signal-to-noise ratio (SNR) of coarse resolution NDVI time series.

Besides VI usefulness information, the smoother also takes into account land/water mask layers of the VI Quality Assessment Science Data Set [61]. Pixels having a VI usefulness value lower than three were considered to be acceptable and assigned a weight of one (very good to good quality), while a VI
usefulness larger than seven was excluded from further processing with a weight of zero (not acceptable). VI usefulness values between three and seven were linearly scaled between one and zero.

For the back-processing, the smoothing parameter $\lambda$ of the Whittaker filter was fixed to a constant value for the entire study region. After some trial-and-error tests, a value of 30.000 was found acceptable, balancing fidelity to the MODIS data with the roughness of the resulting curve. To further reduce the possible impact of undetected clouds and poor atmospheric conditions, three filtering iterations were performed to fit the upper envelope of the NDVI. Multiple filter runs were for example recommended by [2,12]. The second parameter of the Whittaker smoother (i.e., order difference) was set after some tests to two. Based on the order difference, the smoother calculates the roughness of the smoothed curve [65]. From the resulting daily NDVI values (combining inputs from MOD13Q1 and MYD13Q1), all images corresponding to “Mondays” were extracted and stored as a 7-daily data set. This NDVI is further referred as filtered MODIS data set. The weekly data were considered sufficient for most of the analyses, reducing the computational load of the exercise. For the analysis of timing of (phenological) events, however, the (smoothed) daily NDVI series were used allowing a better retrieval of start/maximum of season.

2.3. Spatial Degradation of MODIS Data

After smoothing, the MODIS data set was spatially aggregated to fit the resolution of the GIMMS data set (1/12°). This was done by averaging the smoothed MODIS data to the GIMMS grid. Pixels flagged as water were excluded from the averaging procedure. With respect to coastal pixels, a mask was applied for the subsequent analysis. It comprises all pixels that represent water in at least one time step of either of both time series. This mask was expanded by one pixel to further reduce possible artifacts. The remaining land pixels should be largely unaffected by coastal water. However, no attempts were made to exclude inland water bodies from the two series.

As the NDVI does not scale linearly there will be a certain impact compared to aggregating the reflectances and then calculating NDVI e.g., [67]. Nevertheless, we decided to aggregate NDVI rather than the reflectances. The decision was done after having analyzed the impact of the different aggregation approaches for a test dataset (2002 to 2004) covering the Iberian Peninsula. For the exercise, MOD09Q1 reflectance data were used. From the reflectances, the two spatially aggregated NDVI were calculated. It turned out that the impact of different aggregation schemes was significantly smaller than the one caused by using different sensors (not shown).

2.4. GIMMS Dataset

The NDVI3g data set is the latest version of the Global Inventory Modeling and Mapping Studies (GIMMS) project [44]. The data set is currently under investigation by a number of research teams. Important contributions can be found in this special issue. The original GIMMS data set was produced from the AVHRR sensor data taking into account various deleterious effects, such as calibration loss, orbital drift, volcanic eruptions, etc. [68]. The third generation GIMMS NDVI is a 15-day Maximum Value Composite derived from AVHRR sensor data from NOAA 7 to 18 satellites. It covers the period 1981–2011. The data set has an improved calibration and cloud masking compared to older versions of the GIMMS data set [44]. For the purpose of our study, the extent of the GIMMS NDVI3g data set was
cut to the extent of the study area (Figure 1), while preserving the original GIMMS NDVI3g grid with a spatial resolution of 1/12°.

2.5. Temporal Smoothing of GIMMS

To allow comparability between the two time series, the GIMMS data set was smoothed in a similar way as the MODIS data. As the exact composite DOY of the GIMMS data set is not known, the 8th and 23th day of the month were assigned as DOY for the 1st and 2nd compositing period of the month, respectively (Table 2). Next, the Whittaker filter was applied to the GIMMS data set to produce daily NDVI values between 2002 and 2011 (ten years). Similar to the MODIS data, the filtered GIMMS data were saved as 7-daily values (corresponding to Mondays). For detecting the phenological indicators, we made again use of the daily GIMMS values.

Regarding a suitable filter setting for the GIMMS data, different tests were performed. The tests aimed at identifying a parameter setting providing GIMMS data similar to the MODIS data in terms of curve shape, amplitude and timing of feature of the NDVI profiles. The decision was made based on the root mean square difference (RMSD) between the data sets. The selected setting was a $\lambda$ of 30.000, one iteration and a second order difference. Table 2 summarizes the main choices regarding the spatio-temporal homogenization of the two data sets.

Note that filtering the GIMMS time series ending December 2011 implies that the last observations were possibly less accurately filtered compared to observations situated in the middle of the series (e.g., temporal edge effects). This was considered overall acceptable but might influence some of the results. This problem did not occur when filtering the MODIS data because the MODIS time series extends into 2013 (and starts before 2002).

2.6. Extraction of Start (SOS) and Maximum of Season (MOS)

Phenological metrics were extracted for all pixels having a maximum NDVI of at least 0.3. An approach was implemented for detecting uni- and bi-modal growing seasons. In case of bi-modal growing seasons, only the metrics of the first season were extracted. The selected approach can also handle seasons spanning over different years. The extraction was done in four steps:

1. Detecting the number of cycles by means of auto-correlation information (e.g., [69]): The number of lags was chosen to be little less than two years. Next the algorithm was searching for maxima and minima on the auto-correlation function (ACF). Two cycles were detected, if a second distinct maximum in the ACF before the end of the year appeared. The maximum was accepted, if the increase before the maximum and decrease after the maximum in the ACF was more than 0.1.

2. Focus on the first season in case of multiple cycles: If in either of the time series two growing seasons were identified, the algorithm searched for two starts and two maxima each year, while only the first start and maximum was included in the subsequent comparison of the time series.

3. Defining a temporal search window for each growing season: In a first step the algorithm was searching for the maximum NDVI of each season. The growing season was considered to be centered around the maximum, while the size of the range depends on the number of detected
cycles. The start (SOS) will be found between the moment of minimum and the moment of maximum. Even if filtered NDVI curves were used, there might be still several preceding minima. To identify a reasonable and reproducible minimum, the steepest preceding increase was chosen as a criterion. This criterion accounts for the number of increasing time steps and the NDVI difference.

(4) Detecting start of season (SOS) and maximum of season (MOS): The start of season was detected using a relative threshold [3,70,71], namely when the NDVI increased 20% of the seasonal amplitude from seasonal minimum. The maximum was defined at the time of the largest NDVI in the growing season.

The extraction procedure used the entire filtered time series from 2002 to 2011 at daily time steps. However, the actual start and maximum were detected for the growing seasons from 2003 to 2011. This restriction was necessary as growing seasons might span different years. For example, valid starts of the growing season of 2003 can range from beginning of July 2002 (−183) to end of June 2003 (181) counted in days of year (DOY). The timing of the seasonal maximum is defined in a similar way, e.g., ranging from the beginning of November 2002 (DOY −60) to end of October 2003 (DOY 304). If values occurred outside the defined start and maximum ranges, they were considered misdetections and flagged as “no data”. A number of alternative algorithms for extracting phenological signals from remotely sensed time series have been described in the literature (for a review see [3,72]). However, it was out of scope of this study to explore more than one algorithm.

3. Results

3.1. Data Distribution and Correlation

Generally, GIMMS and MODIS NDVI data sets show a reasonable agreement regarding their spatial distribution and temporal behavior. Exemplary annual profiles (averaged over the 2002–2011 time period) of the two time series are reported in Figure 2. Likewise, the two cumulated distribution functions derived from the entire time series display only little divergences (Figure 3). This also holds for the multi-annual mean and standard deviation (Figure 4). Together these findings demonstrate a generally relative good correspondence between the GIMMS and MODIS derived vegetation densities. It also demonstrates the success of the applied pre-processing and smoothing, yielding identical spatial and temporal resolutions of the originally very different data sets.

However, despite the general agreement, some persistent differences can be observed. In particular, in winter time, significant differences in the averaged NDVI of the two time series are visible in Northern and Eastern Europe (e.g., Sweden, Finland) (Figure 5). On the contrary, the agreement between the two data sets is better for these areas during summer months (June, July, and August) and (to a lesser extent) in spring and autumn. The same can be observed in Figure 2 where some of the displayed profiles reveal a stronger disagreement during winter time. However, particularly in Spain, France and Italy, one can observe higher differences in the summer and spring months than in winter.
Figure 2. NDVI profiles (top) and first derivatives (bottom) for individual pixels from GIMMS (left) and MODIS (right). Example pixels were selected to represent different land cover types. Curves were obtained by averaging the respective NDVI of sample locations over the period of 2002–2011.

Figure 3. Data distribution of GIMMS and MODIS time series (weekly data) extracted from the full image extent and pooled across all years (2002–2011).
Figure 4. Spatial distribution of NDVI from GIMMS (left) and MODIS (right) time series. Displayed are the average NDVI (top) and standard deviation (bottom) calculated from weekly data over the full time period (2002–2011).

Figure 5. Spatial distribution of weekly NDVI from GIMMS (left) and MODIS (center) time series averaged over 2002–2011 for each of the four climatological seasons. (right) Differences between the mean NDVI of GIMMS and MODIS.
The differences between the two time series become more evident, when plotting (pixel-by-pixel) the NDVI values against each other (Figure 6). The results indicate relatively large scatter around the 1-to-1 line with sometimes extremely large differences between individual observations. In addition, a small bias between GIMMS and MODIS is observed. The GIMMS data reveal on average a somewhat higher NDVI compared to MODIS. However, most of the observations fall well within a range of ± 0.1 NDVI units (Figure 6, right). Calculated over all pixels and all time steps (i.e., all weekly observations between 2002 and 2011), we find a correlation between GIMMS and MODIS time series of about \( R^2 \) 0.88 (Figure 6). This correlation shows a clear seasonal component (Figure 7, left). The intra-annual coefficient of determination \( (R^2) \) varies between 0.78 (winter) and 0.92 (summer).

A value of \( R^2 \geq 0.9 \) is found between week 13 (end of March) and 37 (mid-September). The root mean square difference (RMSD) exceeds >0.05 throughout the year (blue line), with values approaching or exceeding 0.10 between week 40 (Beginning of October) and 13 (End of March). This confirms again that the largest differences regarding the observed vegetation densities occur during winter time. The inter-annual agreement/disagreement between the two data sets is depicted in Figure 7 (right). Results show only little variability from year to year, except for the last year in the time series (2011). Indeed, this year is the last year in the original GIMMS time series. Hence, the filtering/smoothing might be affected by the temporal edge effects mentioned before (2.5). This will in particular affect the November/December observations of GIMMS (see also weeks 49 to 52 in Figure 7 left). We tested the effect of excluding the weeks 49–52 of 2011 on the calculations of Figure 7. The
biggest absolute differences for the intra-annual variation of RMSD and $R^2$ appear in week 52 with 0.006 and 0.02, respectively. The absolute differences for the inter-annual variation of RMSD and $R^2$ calculated for 2011 are slightly lower. As the described differences in these particular weeks are irrelevant for calculating SOS and MOS, we kept all data in our analysis.

**Figure 6.** Co-distribution of NDVI values from GIMMS and MODIS derived from the full image extent and weekly observations across all weeks between 2002 and 2011. (left) scatterplot with 1-to-1 line (red) and regression line (black); (right) frequency distribution of the differences between GIMMS and MODIS NDVI.

![Comparison of filtered results for 2002-2011](image)

**Figure 7.** Agreement/disagreement between weekly GIMMS and MODIS NDVI values. (left) Intra-annual analysis; (right) inter-annual analysis. (solid green) coefficient of determination ($R^2$), (blue) root mean square difference (RMSD), (dashed green) time course of the average NDVI. Per week, or year, one value is shown per indicator. Lines are only shown for reader’s convenience.

![Comparison of filtered results for 2002-2011](image)

The spatial pattern of the pixel-wise correlation between the two time series across all weekly observations is shown in Figure 8 (left). The map reveals a clear spatial pattern with high correlations (>0.8) over much of the central and eastern part of Europe as well as the northern parts of the Maghreb countries (Algeria and Morocco). Correlations are markedly lower in the coastal parts of Spain, Italy, and Greece/Turkey and to a lesser extent in France, Ireland and the UK. Particularly in Portugal,
Galicia, and to a lesser extent in Brittany, very low correlations are accompanied with almost continuously high NDVI levels and a low variability (see Figures 4 and 5). In these areas, clear differences between the NDVI profiles occur (not shown). Very low correlations between the time series are also found over the arid parts of North Africa. In these areas, however, the low correlations are a direct result of the constantly low vegetation density (Figure 5). To further explore the impact of intra-annual vegetation dynamics on the derived (pixel-wise) agreement/disagreement between the two time series, Figure 8 (right) shows the RMSD between the two data sets.

**Figure 8.** ([left]) Coefficient of determination ($R^2$) and ([right]) root mean square difference (RMSD) between the temporal GIMMS and MODIS series calculated across all weekly observations between 2002 and 2011.

![Coefficient of determination (R²) and root mean square difference (RMSD) between the temporal GIMMS and MODIS series calculated across all weekly observations between 2002 and 2011.](image)

The RMSD pattern displays a relatively high disagreement over Scandinavia and Eastern Europe, but also in the alpine region and Spain. The best agreement is observed over the constantly dry parts of North Africa (e.g., Sahara) and the non-coastal areas of the UK and Ireland. Some larger reddish spots (RMSD > 0.2) in Sweden and Estonia/Russia are due to inland lakes not masked out for this study.

### 3.2. Phenology: Start of Season (SOS)

The start of the season (SOS) was determined for each pixel and each year in the time series (2003–2011) using the relative threshold approach described in Section 2.6. The algorithm yields up to nine SOS estimates per pixel. The cumulated distribution functions of SOS derived from the two time series are shown in Figure 9. Both data sets reveal a similar shape of the SOS distribution. However, GIMMS (in red) depicts SOS frequently more than one week earlier, compared to MODIS (in blue). A similar trend is already visible in Figure 2 where the NDVI profiles of the GIMMS data start rising earlier than MODIS.

The inter-annual variability (“range”: difference (in days) between maximum and minimum) of SOS exhibits a relative similar spatial distribution in both data sets (Figure 10). However, the SOS extracted from the MODIS time series shows a larger inter-annual variability compared to GIMMS in many coastal areas of the North Sea (Belgium, Netherlands, Germany, Poland, and Sweden). The opposite can be observed over the north part of Spain.
Figure 9. Cumulated distribution function of the pixel-wise estimated start of season (SOS) for GIMMS (red) and MODIS (blue) for all years (2003–2011) using the relative threshold approach. Smoothed products at daily temporal resolution were used for the calculations. SOS of zero corresponds to a turn of the year.

Figure 10. Range of estimated start of season (SOS) across nine years (2003–2011) for two time series (GIMMS and MODIS). The ranges are expressed in days.

A more detailed analysis at pixel level shows that the SOS extracted from the two series often reveal high differences. Figure 11 (left) presents the frequency distribution of the SOS differences (SOS_{GIMMS}−SOS_{MODIS}) over the entire time series (2003–2011). It can be seen that many observations fall outside the range of ±30 days. Only 32% of the observations have (acceptable) differences ≤1 week. Generally, GIMMS depicts the SOS earlier than MODIS confirming earlier findings (Figures 2 and 9). This can be consistently observed for individual years and is exemplarily shown in Figure 11 (right) for the year 2007. In this example, MODIS data evidences a SOS on average 14 days later than GIMMS. The differences between the two time series can also be clearly seen when mapping the spatial distribution of the SOS for a particular year. For example, Figure 12 (right) displays the SOS
for 2007 estimated by GIMMS (top) and MODIS (bottom). The median SOS across all years (Figure 12, left) confirms again the large differences between the two data sets regarding the SOS.

**Figure 11.** Frequency distribution of differences in the start of season (SOS) from 2003 to 2011 (left) and for 2007 (right). Differences in SOS are expressed in days.

**Figure 12.** Spatial distribution of start of season (SOS) for GIMMS and MODIS time series. (top) averaged SOS (median) over all years (2003–2011) and (bottom) SOS of 2007.

The disagreement between the two time series is further sustained when plotting the SOS of both time series against each other (Figure 13). The regression analysis between the two SOS data sets reveals only
a $R^2$ of $\sim0.66$, if data from all years are pooled together. A larger number of strong outliers can also be detected. They mostly correspond to cases where the detection algorithm (see Section 2.6) depicts different parts of the growth cycles of an individual pixel. This often happens if the profiles show multiple—but not well distinguished—peaks in one but not the other data set (not shown).

Analyzed separately for each individual year, some inter-annual variability in the correlation between SOS from GIMMS and MODIS are found (Figure 14). Somehow higher (spatial) correlations can be observed in 2005, 2006, and 2009, which also correspond to years with delayed onset of vegetation growth (negative North Atlantic Oscillation index) and more homogeneous growth patterns. Despite the $R^2$ occasionally exceeding 0.7, the RMSD between the two series of SOS always remains in the range 30–40 days (blue).

**Figure 13.** Scatterplots between the start of season from GIMMS vs. MODIS for full image extent and all observations. (left) data from 2003 to 2011 (right) only data from 2007. The 1-to-1 line (red) and the regression line (black) are also shown. Densities increase from blue to red.

**Figure 14.** (top) annual analysis of the pixel-wise correlation between SOS from GIMMS and MODIS with coefficient of determination ($R^2$) and root mean square difference (RMSD). (bottom) inter-annual distribution of SOS for GIMMS (left) and MODIS (right). The box-and-whisker plots show the median in red. Box edges represent the 25th and 75th percentiles. The whiskers extend to data points still within the 1.5 interquartile range.
At the pixel level, we find both high and low rank correlation (Spearman’s Rho) between SOS from GIMMS and MODIS (Figure 15). P-values were computed for testing the hypothesis of no correlation against correlation is greater than zero (Figure 15, upper right). In most cases, the strong positive correlations (>0.6) were significant at the 0.05 level. Low/negative correlations are often present in areas with reduced inter-annual variability of SOS and where extracting SOS is influenced by snow cover (Figure 10).

Figure 15. Correlation (at pixel level) between the start of season (SOS) estimated from GIMMS and MODIS time series. (top left) rank correlation coefficient, (top right) p-value, (bottom) the frequency distribution of Spearman’s Rho. The correlation was only calculated if the relative threshold algorithm could at least detect five valid SOS observations (out of nine). In the maps (top) areas not fulfilling these conditions are shown as ‘no data’.

Figure 14. Cont.
3.3. Phenology: Maximum of Season (MOS)

Similar to the SOS, the maximum of the season (MOS) was determined within the growing season for each pixel and each year in the time series (2003–2011), yielding up to nine MOS estimates per pixel. The cumulated frequency distributions of the MOS extracted from GIMMS and MODIS are shown in Figure 16 (top) demonstrating a good correspondence between the two data sets. Likewise, the co-distribution shown in the scatterplot of Figure 16 (bottom) reveals, that the majority of the pixels (in green-yellow-red) well distribute around the 1-to-1 line.

**Figure 16.** Maximum of season (MOS) derived from GIMMS and MODIS (full image extent and all data from 2003–2011. (top) cumulated frequency distribution, (bottom) scatterplot. In addition, also shown are the 1-to-1 line (red) and the regression line (black).

The spatial pattern in MOS is similar for the two data sets when averaged over all years (Figure 17, top). The same holds to a lesser extent for the observed inter-annual range of MOS (bottom); compared to the MODIS extracted MOS, the inter-annual variability derived from GIMMS data is often much higher. In some areas, the range between the earliest and the latest MOS exceeds 100 days where MODIS yields an inter-annual range of MOS of about 30 days (e.g., southern part of Sweden). Secondary maxima in the GIMMS data, not present in MODIS, are the main reason (not shown).
The correlation between the two MOS products at pixel level is displayed in Figure 18 (left) along with the median of differences (right). In most regions, the GIMMS and MODIS MOS values had a Spearman’s Rho exceeding 0.4. Regions with low rank correlation (left) and high p-value (center) often correspond to regions with low SOS correlation (Figure 15, top left).

**Figure 17.** Spatial distribution of the maximum of season (MOS) from GIMMS (left) and MODIS (right). (top) median across nine years, (bottom) inter-annual range.

**Figure 18.** Correlation (at pixel level) between the maximum of season (MOS) from GIMMS, respectively, MODIS time series. The rank correlation coefficient is shown on the left whereas the corresponding p-value is shown in the center. The median of differences (in days) is shown on the right.
4. Discussions

Before starting the comparative analysis, efforts were made to harmonize the two (originally very distinct) data sets to a common spatial (1/12°) and temporal (weekly/daily) resolution while eliminating at the same time noise and other artifacts (Table 2). Despite these efforts, we have to acknowledge that overall the two NDVI time series show only a moderately good agreement. The agreement is better between the NDVI values compared to the phenological indicators.

The overall relationship between the two weekly NDVI products (ca. 125,000 pixels × 10 years × 52 weeks) has a R² of 0.88 and RMSE of 0.09 (Figure 6). The correlation is relatively high if one considers that the original data came with 1/12° (GIMMS) vs. 250 m resolution (MODIS) and bi-weekly vs. 8-daily temporal resolution. Comparatively analyzing fAPAR time series (fraction of absorbed photosynthetically active radiation) obtained from one single sensor (SPOT-VEGETATION), [46] found R² between 0.60 and 0.94 for three test windows with each 10° × 10° (RMSE 0.06–0.24).

A marked seasonal component of the correlation/agreement between the two NDVI time series is observed (Figure 7, left). Disagreement between the two series is generally higher during winter (RMSE ≥ 0.10) and reduces considerably during summer (RMSE ≤ 0.08). This seems a reasonable finding as observation conditions degrade strongly during winter time leading to fewer observations with generally lower quality. This is also confirmed by analyzing the occurrence of good quality measurements in MODIS data such as VI usefulness ≤ 3 (see [61]), where fewer favorable cases are counted from November to February than in summer (not shown). Surprisingly, however, the agreement (e.g., R²) does not always go parallel with vegetation density and/or season. Instead, we observed a relatively strong asymmetry around summer. For example, the agreement in spring is relatively close to summer conditions, whereas autumn is closer to winter. Considering solely LAI, sun angle and temperatures, etc., one would expect a relatively symmetric behavior of the seasonal agreement profiles. This neglects, however, the fact that not only LAI changes during the year (with a more or less good symmetry) but also the leaf pigmentation and structure. Possibly during autumn (leaf senescence), these changes at the leaf level accentuate the (otherwise insignificant) differences in the band setting of the two sensors (AVHRR vs. MODIS).

The correlation between the two data sets shows on the other hand only a low inter-annual variability. For all years, except 2011, the R² is close to 0.9 with RMSD of ~0.09 showing a good correspondence in the spatial patterns of vegetation density (NDVI) of the two data sets (Figure 7, right). The significantly higher disagreement observed for 2011 is partly related to the fact that the original GIMMS series ends in 2011. Hence, when applying the smoother for noise removal, the very last observations of this data set were smoothed with a lower accuracy compared to the MODIS data for which we had observations beyond 2011 (thus, avoiding this well-known border effect). All data of year 2011 was nevertheless kept in the comparative analysis as the overall effect on NDVI differences and in particular on extracting start and maximum of season should be small.

A good agreement between the two series can also be observed regarding the spatial maps of average NDVI and corresponding standard deviation (Figure 4). The former indicator reflects the average vegetation density of a pixel, whereas the standard deviation is a proxy for vegetation dynamics/variability at a given location (combining seasonal and inter-annual dynamics). In our study, the vegetation dynamics directly influences the (pixel-wise) correlations of the temporal signatures...
across the continent (Figure 8). As expected, in areas with low dynamics, the \( R^2 \) between the two series was considerably reduced. They fall even down to zero in the absence of any dynamics such as in the Sahara. However, we find interesting differences between the Sahara where (as expected) both \( R^2 \) and RMSD approach zero, and other parts of Europe with equally low vegetation dynamics (thus, \( R^2 \sim 0 \)) but with higher RMSD. This is for example observed in many coastal areas of the Mediterranean Sea and the NW coast of Spain/Portugal. It is currently not clear why the two data sets behave so differently in these areas.

Using the relative threshold approach similar to e.g., [3,12,49,70] two major phenological metrics were extracted for each year of the time series: start (SOS) and maximum of season (MOS). The observed spatial pattern from both data sets (Figures 12 and 17) reflects well existing gradients within Europe resulting from the combined impact of land cover, latitude, altitude, temperature, and rainfall regimes within the study area [13].

Regarding MOS, only few differences can be observed regarding the multi-annual averages (Figure 17) as well as its general frequency distribution (Figure 16). For major parts of the European continent, the Spearman’s Rho (rank) correlation coefficient between the two MOS products is \( \geq 0.5 \) (Figure 18). The average (median) difference in MOS is often in the (acceptable) range of ±10 days. However, some striking differences can be observed. For example, lower correlations appear over the (boreomemoral) southern part of Sweden. In this area the two products differ most strongly. Whereas both products show moderate differences in terms of average MOS (10–20 days) for this region, the corresponding inter-annual variability (range) differ significantly (Figure 17); GIMMS data yielded an inter-annual range of MOS of more than 50 (sometimes more than 75) days. On the contrary, the MODIS data exhibit smaller (and possibly more reasonable) ranges of ±20 days.

The same region is again noticeable in the SOS data of GIMMS (Figure 12). Indeed, the GIMMS data show a start of season as early as for example (oceanic) Ireland (e.g., in February). This contrasts for example with maps published by [73] in which the average onset of spring for this region was reported to be April/May with similar values in a large region ranging from Denmark to Southern Finland. A very similar spatial pattern of SOS is observed in the MODIS data in the mentioned region. At the same time, the MODIS data also depict the earlier onset of vegetation growth in Ireland. The MODIS data also well reflect the spatial pattern (and absolute values) of SOS published in [74] for southern Norway. In their work the range of Birch flowering was between April and June with an earlier onset towards coastal (low level) areas.

Regarding the inter-annual variability (range) of SOS for the most northern parts of the test window (e.g., Norway, Sweden, and Finland) we could compare our maps against data published by [75]. In the mentioned publication it was shown that the SOS of the entire region could vary at least ±10 days. This is not well reflected in the GIMMS data where the range of SOS is often only 0–10 or at maximum 10–20 days (Figure 10). On the contrary, the MODIS derived SOS range is at least 10–20 days or more.

The most striking difference in SOS concerns however the (large) area of central Europe (from France/UK to Germany/Poland) (Figure 12). For the mentioned region, GIMMS maps the average SOS as early as February, whereas MODIS depicts the SOS between March and April. Although direct evidence is missing we believe that the GIMMS derived vegetation onset is probably too early. For example, maps produced by the German Weather Service (DWD) involving key tree species such as
birch and apple, as well as common agricultural (winter) crops such as winter wheat and oats, defined the onset of growing season in Germany only around March/April. A more detailed study will have to correlate the remotely sensed maps with the mentioned reference maps, respectively, field observations. This is however beyond the scope of the present study.

More research is also needed to verify that the observed differences in SOS persist in case of using different detection algorithms. For MOS, there might be differences in using for example a curve fitting approach, e.g., [12]. Special attention is needed in case of flat plateaus or when secondary maxima appear near the plateau. Several studies have pointed out that huge differences can be obtained for individual phenological metrics, depending on the employed phenological algorithm [3,72]. Hence, it cannot be excluded that the observed differences in SOS and MOS patterns are solely the result of the selected detection algorithm.

5. Conclusions

The research is a first attempt to analyze in a comparative way NDVI time series from GIMMS and MODIS (2002–2011) over a large and heterogeneous study area in Europe. The intention is not to qualify one of the two data sets of being ‘better’ but to highlight existing differences and agreements. The comparison between existing products yields important information regarding the feasibility of creating “combined” historical archives of NDVI (and derived phenological indicators). With the launch of Sentinel and Proba-V satellites, such cross-sensorial studies will become more and more important as they permit to effectively address issues related to the ongoing climatic change.

The study reveals several points worth being mentioned:

- Overall, the NDVI data sets from GIMMS and MODIS show only a moderately good agreement. Differences in observed vegetation density (NDVI) are largest in winter, followed by autumn. The best agreement is observed during summer. This highlights the need to carefully inter-calibrate the two data sets, giving special attention to autumn and winter. Without such an inter-calibration, the two data sets cannot be analyzed concurrently.
- The maximum of season (MOS) extracted from the two time series shows a good agreement with respect to the observed spatial pattern and its inter-annual variability. However, more research is necessary to explain the observed differences in the timing of vegetation onset (start of season—SOS). GIMMS consistently shows higher vegetation densities during spring. Its temporal NDVI profiles therefore start rising very early in the year and earlier than those of MODIS. This leads to, sometimes, very large differences in the mapped SOS of both data sets. Without removing these differences, the two data sets cannot be combined for phenological studies.
- Regarding the start of season (SOS), differences were observed regarding spatial pattern and inter-annual variation. Only a part of these differences are of systematic nature and could therefore be corrected a posteriori. This again clearly highlights the need of a well-done cross-sensor inter-calibration.
- Future studies should focus on validating phenological indicators. At least for some countries of Europe, long lasting and relative dense phenological networks exist (e.g., Germany and Austria). This permits comparing the remotely sensed indicators against field observations as already shown by e.g., [76–78].
Acknowledgments

We thank Ranga Myneni, Jorge Pinzón and Zaichun Zhu for providing the NDVI3g of Europe for this paper, and the three reviewers for their valuable contributions. The MODIS MOD13Q1, MYD13Q1, MCD12Q1 and MOD09Q1 data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/get_data). The computations have been partly achieved using the Vienna Scientific Cluster (VSC).

Conflicts of Interest

The authors declare no conflict of interest.

References


© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).