**2. Data & Methods**

***2. 1 Data:***

The two LAI datasets used for the analysis, a modelled and a measured one, will be described in this section. The modelled LAI comes from the global reanalysis of vegetation phenology by Stöckli et al. (2011) and will be labelled LAIre, the measured LAI is based on the longest available NDVI time-series, the AVHRR GIMMS NDVI3g (Tucker et al. 2005; Pinzon & Tucker 2014), and will be labelled LAI3g. The data was analysed over a period of 30 years from the years 1982 to 2011.

*LAIre – Modelled LAI*

Stöckli et al. (2011) produced a global LAI dataset with a model based on Jolly’s (2005) climatic controls used in the Growing Season Index (GSI). Jolly’s GSI uses two thresholds for each control: a minimum threshold where growth is fully limited, and maximum threshold where plant growth is not affected by the control anymore. In between the minimum and maximum threshold, a linear gradient is assumed. The controls are calculated using European Centre for Medium-Range Weather Forecast (ECMWF) ERA-Interim data.

Stöckli extended the GSI model into a prognostic phenology model to predict LAI and fractional photosynthetically active radiation (FPAR). To more accurately represent different vegetation types and other land surface covers such as urban areas or water bodies, he included a set of 35 natural and non-natural plant functional types (PFT) and set different thresholds for each type, based on prior research. The percentage of each PFT for each pixel is set based on a combination of several MODIS land cover, vegetation and crop products and is kept static over the years. Finally, MODIS LAI and FPAR data from 2000 to 2009 was used to assimilate the model.

The final LAIre dataset provides daily LAI and FPAR estimates for the past 50 years with a ½ degree spatial resolution. In addition to the modelled LAI values the dataset also includes daily climatic control factors for temperature, VPD and radiation, on which the model is based.

*LAI3g – Measured LAI*

The remotely sensed LAI dataset was derived from the third generation GIMMS AVHRR NDVI dataset (NDVI3g) by Zhu et al. (2013). They used temporally overlapping MODIS LAI data from 2000 to 2009 to train a neural network with the NDVI3g. The resulting LAI3g is a global dataset with 1/12-degree spatial resolution and a 15-day (bimonthly) temporal resolution for the years 1982 to 2011.

It is important to note that the 15-day temporal resolution of the NDVI3g, and therefore also the LAI3g, is based on an unknown number of actual acquisitions within each 15-day period. Due to the extensive processing of the NDVI to create the NDVI3g, it is then not possible anymore to extract the exact date of each acquisition in order to assess cloud cover or other influencing factors (Pinzon & Tucker 2014).

***2.2 Methods***

*Data Pre-processing*

Both datasets had to be resampled to the same temporal and spatial resolution in order to be comparable. The temporal resolution of the LAIre was resized from daily to bimonthly using 15-day means to fit the LAI3g dataset. The LAI3g dataset was resized spatially from 1/12-degree to ½ degree using bilinear resampling in a 6 x 6 window. A special case had to be made if a 6 x 6 pixel-window for the LAI3g included pixels with no values. A no-value pixel can either be due to it representing a water body or desert, or if no acquisition could be made during the 15-day window. Excluding all 6x6-windows which include at least one no-value pixel would have resulted in a very large number of excluded pixels. On the other hand, ignoring those no-value pixels would introduce bigger uncertainty in the data. Therefore, a middle way had to be found to deal with no-value pixels. Finally it was decided that if more than half the pixels within the 6 x 6 window had no value, the pixel is excluded. If less than half the pixels had no value, the no-value pixels were set to a LAI of 0 and included in the bilinear resampling. This was done to mimick the LAIre production, which similarly calculates the effect of water or desert-bodies based on the PFT percentage in the pixel.

Due to the half-year shift in growing seasons between the northern and southern hemisphere, both datasets were split by hemisphere. The southern hemisphere was redefined to start in the middle of each year (scene 13 out of 24 bimonthly scenes) and last until the middle of the following year (scene 12 of the following year). Since there was no data for the year 2012 which could be added for the year 2011 in the southern hemisphere, the first 12 scenes of 2011 were added as a substitute to get a complete growing season.

***2.2.1 Extraction of Land Surface Phenological parameters***

In order to do LSP comparisons between the two datasets, LSP parameters have to be extracted. This is usually done by first smoothing the dataset to eliminate artefacts from cloudy or dead pixels and by then defining a Start-of-Season (SOS) and an End-of-Season (EOS) for each growing season (Garonna et al. 2014; de Jong et al. 2011).

*Smoothing the dataset*

In order to extract LSP indices from the remotely sensed LAI3g dataset, the data had to be smoothed to eliminate outliers due to cloud contamination and sensor errors. This was done using the Harmonic Analysis of NDVI Time-series (HANTS) algorithm (Roerink et al. 2000; Roerink et al. 2003). The algorithm works by applying a fast fourier transform to the measured values and extracting first, second and third order sine-functions to produce a smoothed version of the original signal. Even though the algorithm was originally developed for NDVI time-series, it can be used for LAI datasets in the same way (Jiang et al. 2010).

|  |
| --- |
| Macintosh HD:Users:davidschenkel:Dropbox:Masterarbeit:geo511:writing:figures:methods:example_stöcklidata.png |
| **Fig. X**: Extreme example of a non-continuous LAI profile near the equator in the LAIre dataset. |

The LAIre does not suffer from cloud-contamination or dead pixels, and therefore produces generally much smoother profiles than the raw LAI3g data. The data was still processed with HANTS, to ensure that no differences due to differing processing chains are introduced. Additionally, the LAIre suffers from non-continuous LAI profiles for some pixels particularly in the tropics (fig X), which are conveniently eliminated by applying HANTS. Otherwise, those pixels would lead to problems in extracting LSP parameters. The exact parameterization used for the HANTS algorithm can be found in Appendix A.

*LSP Parameters*

In the scientific literature there are a variety of different methods used to define the SOS (Reed et al. 2003). However, the Midpoint (MP) and Maxincrease (MI) methods (fig X) are generally considered the most suitable to describe measured and modelled phenology (White et al. 2009; Garonna et al. 2014) and are therefore the two methods used in this thesis. The Midpoint method defines SOS as the value half-way between lowest and highest LAI value within one year. The Maxincrease method defines SOS at the point where the LAI curve shows the steepest positive slope – or in other words, where the first derivative of the curve is maximal. The EOS is usually defined as the date after SOS on the LAI curve where the LAI drops below the SOS value again.

*Implementation of LSP parameter extraction*

Since the extraction of LSP parameters is a fundamental concept of most research in remote sensing of LSP, it warrants a closer look in this thesis.

A mask is first applied to exclude water pixels. Then, pixels which show a weak intra-annual LAI variability below a threshold of 0.5 were filtered out. This is done mainly in desert areas where no vegetation is present and in tropical areas where vegetation shows no clear seasonality.

|  |
| --- |
| Macintosh HD:Users:davidschenkel:Dropbox:Masterarbeit:geo511:writing:figures:methods:MI-MP-illustration-01.png |
| **Fig. X**: Illustration of the extraction methods for LSP parameters. The Midpoint (MP) method sets the date of SOSMP at half the amplitude. The Maxincrease (MI) method sets the date of SOSMI at point of steepest increase in the LAI profile. EOS is set at the point where the curve goes under the SOS value again. |

The Midpoint method uses minimum and maximum LAI for each pixel and year and is calculated with LAIMP=(LAImin + LAImax)/2. Then, the point in time where the LAI-curve first goes above LAIMP is defined as the SOS.

The implementation of the Maxincrease method is more challenging in its conception since the theoretical definition is based on continuous data rather than the discrete datapoints that are being processed in practice. In this thesis, an advanced algorithm from Garonna et al. (2014) was used. They use a spline fit to the datapoints and extract the point of maximum inflection from the spline. The algorithm also uses a more thorough filtering method for pixels where an extraction is not possible.

To understand the mechanisms of LSP parameter extractions, an attempt was made at implementing the Midpoint and Maxincrease extraction methods, rather than relying on already available algorithms. While the Midpoint method was implemented as described here, the implementation of the Maxincrease method followed a simplified approach. The approach is based on the assumption that the maximum increase must occur between the two adjacent datapoints, which have the highest positive difference in LAI between them. The LAIMI can then be defined as LAIMI = (LAIi+LAIi+1)/2. Then the extraction follows as for the Midpoint method.

Even though the implemented Midpoint method showed very similar results as the algorithm used by Garonna et al., the Maxincrease methods showed a bigger variability of the data due to the two different approaches used. To make the results more comparable to other similar studies with NDVI3g datasets (Garonna et al. 2014, more?), it was decided to use the more advanced algorithm for further analysis.

***2.2.2 Comparing LAIre to LAI3g***

*Comparing raw LAI values*

To make a first assessment on the comparability of the two datasets, the raw values for each 15-day period were correlated. Since a linear 1:1 relationship was expected, Pearson’s r was used for the correlation statistics. A linear regression analysis for yearly mean, minimum and maximum values for each pixel was performed to check for any systematic over or underestimations in the modelled dataset.

*Comparing Phenological Parameters*

A linear regression was performed to compare the extracted LSP parameters SOS, EOS and GSL from the two datasets. The regression was done for both extraction methods, MI and MP. A separate linear regression was performed for the region between 45 degrees and 90 degrees northern latitude to exclude effects from multiple annual growing seasons in the tropics and the Mediterranean zone (Garonna et al. 2014).

Due to the latitudinal and regional effects on phenology, maps showing the absolute difference in days between the parameters extracted from the two datasets were also computed. This allows for a more intuitive comparison on regional agreements and disagreements between the two datasets.

*Decadal Change*

To complete the comparison of the two datasets, a trend analysis was conducted. The rate of change of the three LSP parameters was expressed in days/decade for each pixel as described in Garonna et al. (2014). To get this rate of change, a linear regression model was applied and the slope of the resulting linear fit taken. Pixels which did not show a significant change according to the analysis of variance (ANOVA) with a significance level of alpha = 0.05 were discarded. If more than 1/3 of observations (more than 10 years of data) or either the first or last year of observation were missing, the pixel was also discarded. For pixels with less than 10 years of data missing, the missing values were linearly interpolated. This was done to minimize effects of short-term changes in trends, which might otherwise skew the results if a significant part of observation was linearly interpolated rather than actually observed.

***2.2.3 Analyzing Climatic Controls***

To investigate changes in dominant controls, the yearly dominating climatic control for each pixel was extracted. For every pixel, the magnitude of each daily climatic control factor was integrated over the whole year. The climatic control with the lowest total sum was then chosen as the most limiting factor for each year. It is a simplified version of the approach described in Jolly et al. (2005) but was found to produce comparable results. This method does not adequately represent areas where two climatic controls are almost equally dominating in their limitation, however, when analysed in a time series, those areas are easily identified since the two dominating controls both tend to show up alternating between years. To identify those areas more easily, the maps were analysed on changes per pixel over the 30 years. Pixels were grouped into those changing only between two specific climatic factors and those showing changes between all 3 climatic factors over the observed time period.

To show inter-annual changes in climatic controls for different times of the year, the decadal change method used in the LSP trend-analysis was used on bimonthly scenes for each climatic control. Instead of days/decade for SOS and EOS, the climatic controls change is then expressed in percentage points per decade. Finally, 6 bimonthly scenes each were aggregated to quarterly scenes to allow for a more intuitive analysis.

***2.2.4 Influence of Climatic Controls on Phenology***

To assess the effect climatic controls might have on SOS and EOS, 30-day mean values for each climatic control were extracted for the 30 days prior to SOS and EOS (see fig X). 30 days was chosen as an interval since it corresponds to the temporal uncertainty caused by the bimonthly resolution of the LAI3g. The extraction of climatic controls was done for LSP parameters extracted from both the LAI3g and LAIre dataset. Only the LSP parameters extracted with the MP method were used however, because the shift of LSP parameters extracted from MI and MP methods are very similar so a distinction between the two will not produce significantly different results. Additionally, the MP method has proven to give more stable results and is therefore more useful for this trend analysis of climatic controls by being less sensitive to outliers.

|  |
| --- |
| Macintosh HD:Users:davidschenkel:Dropbox:Masterarbeit:geo511:writing:figures:methods:30_day_extraction.png |
| **Fig. X**: SOS and EOS are extracted from LAI profile (black) and then climatic controls (red, blue, green, taken from Jolly et al., 2005) in the 30 days prior to SOS/EOS can be extracted. |

With the extracted 30-day climatic control means, the dominating factors for each year were computed for both the 30-day period before SOS and before EOS, analogous to the methods described in 2.2.3. This can then be used to study dominant factors during SOS and EOS as well as to study the effects of changes in climatic controls extracted in chapter 2.2.3 on the changes in LSP extracted in chapter 2.2.1.

A trend analysis for the individual changes in climatic controls in the 30-day window prior to SOS and EOS was also computed tested on significance on the 10% level. A significance level of 10% was chosen in this analysis since the dependence on trends in both LSP parameters and climatic controls for the observed periods and for such a coarse resolution results in a greater year-to-year variability. The trend analysis from the 30-day window was then correlated to the change-rate trend analysis from the corresponding LSP parameters, SOS and EOS.