**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 stems from the global reanalysis of vegetation phenology by Stöckli et al. (2011) and will be labelled LAIre, the measured LAI is produced from the GIMMS NDVI3g and will be labelled LAI3g.

*LAIre – Modelled LAI*

Stöckli et al. (2011) produced a daily global modelled LAI dataset for the past 50 years with 0.5 degree spatial resolution.

In addition to daily modelled LAI values the dataset also includes climatic control values for temperature, moisture and radiation based on the parameters used by the GSI as proposed by Jolly et al. (2005). The GSI uses thresholds for fully limiting and non-limiting values of each control and assumes a linear gradient between the two thresholds (see figure X). The controls are calculated using ECMWF ERA 40 and ERA Interim data.

Stöckli extended the GSI model into a prognostic phenology model to predict FPAR and LAI, using MODIS FPAR and LAI data from 2000 to 2009 for data assimilation.

*LAI3g – Measured LAI*

The remotely sensed dataset is an LAI dataset derived from the third generation GIMMS NDVI dataset (NDVI3g) by Zhu et al. (2013). They used 10 years of MODIS LAI data to assimilate the NDVI3g dataset into a global 15-day dataset with 1/12 degree spatial resolution for the years 1981 to 2011.

***2.2 Methods***

*Data Preprocessing*

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 to 0.5 degrees using a bilinear resampling method within a 6\*6 window. If more than half the pixels within this 6\*6 window had no value (mainly due to water bodies), the pixel was excluded.

***2.2.1 Extraction of LSP 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 pixels and by then defining a Start of Seasons (SOS) and an End of Season (EOS) for each growing season (references, fig X)

*Smoothing*

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 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 timeseries, it can be used for LAI datasets too (Jiang et al. 2010).

*LSP Indices*

In the scientific literature there are a variety of different methods used to define the SOS (fig X, Reed et al. 2003). However, the Midpoint and Maxincrease methods are generally considered the most suitable to describe measured and modelled phenology (White et al. 2009; Garonna et al. 2014). Therefore, these two methods are used in this thesis for the comparison of the two datasets.

*Implementation of a LSP parameter extraction*

The extraction of LSP parameters is the basis of most research in remote sensing of LSP and as such warrants a closer look in this thesis. To understand the mechanisms of LSP parameter extractions, an attempt was made at implementing the Midpoint and Maxincrease extraction methods, rather than using available tools.

For both methods, a mask was first applied to exclude water pixels. Then, pixels which show an inter-annual LAI variability below a threshold of 0.5 were filtered out.

The implementation of the Midpoint method was rather straight-forward by calculating LAIMP=(LAImin + LAImax)/2 for each year and pixel, and defining the point in time where the LAI-curve first goes above LAIMP as the SOS. A special case had to be implemented for the southern hemisphere, where the growing season is shifted by half a year.

The implementation of the Maxincrease method was 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. The approach used is based on the very simplistic assumption that the maximum increase must occur between the two adjacent datapoints which have the highest positive difference in LAI between them (fig X). The LAIMI was then defined as (LAIi+LAIi+1)/2. A more precise definition does not make sense with the given temporal resolution of the data.

When comparing the self-made implementation with more refined algorithms as used by Garonna et al. (2014), the results for the Midpoint method were comparable, the Maxinflection method however produced very different results. This is due to the more realistic extraction method, which uses a spline fit to the datapoints and extracts the point of maximum inflection from the spline rather than the bimonthly datapoints. The algorithm also uses a more thorough filtering method for pixels where an extraction is not possible.

For these reasons, as well as for making 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 minimum and maximum values for each pixel was performed to check for any systematic over or underestimations in one of the datasets.

*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 and split by northern and southern hemisphere to take into account the shift of half a year in growing seasons.

Due to the latitudinal and regional influences 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 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 equally discarded. For pixels with less than 10 years of data missing, the missing values were linearly interpolated.

***2.2.3 Analyzing Climatic Controls***

To investigate changes in dominant controls, the yearly dominating climatic control for each pixel were extracted. For every pixel, each climatic control 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. This method is a simplified version of the approach described in Jolly et al. (2005) but was found to produce comparable results.

The yearly maps were then analysed on number of changes per pixel over the 30 years. Pixels were grouped into those changing only between two climatic factors and those showing changes between all 3 climatic factors.

To better show changing trends for climatic controls over 30 years, the decadal change method used in the LSP trend analysis was used again. The climatic data were aggregated to 15-day means to fit the temporal resolution of the LAI3g. The resulting rate of change is then given as percent/decade.

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

To assess the effect climatic controls have on SOS and EOS, 30-day mean values for each climatic control were extracted for the 30 days prior to SOS and EOS. This 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 differences produced by the MI and MP method are generally much smaller than the 30 day window used for the climatic controls (citation needed) so a distinction between the two will not produce significantly different results.

With the extracted 30-day means, the dominating factors for each year were computed and compared, analogous to the methods described in 2.2.3.

A trend analysis using the same change-rate method as in 2.2.3 was also computed. The trend analysis from the 30-day window was then correlated to the change-rate trend analysis from corresponding the LSP parameters, SOS and EOS and tested on significance on the 5% level.