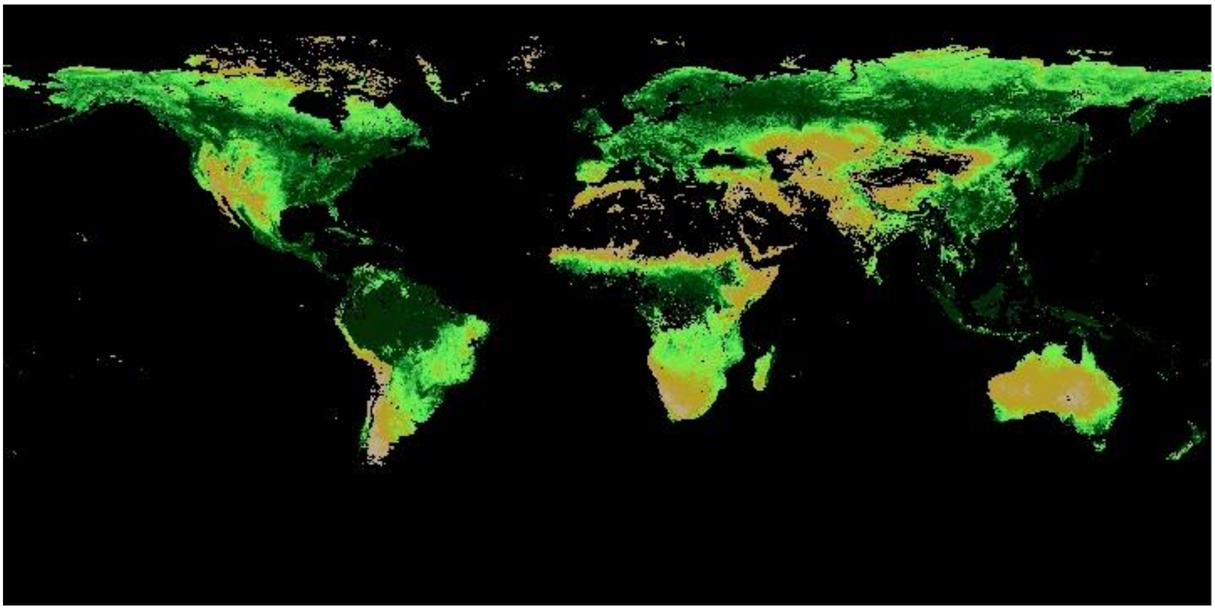
# Derivation of global empirical relationship between climate variables and leaf area index for the parametrization of large-scale hydrological models



Anna Teuerle
August 2018

Supervisors
Dr. ir. Ruud van der Ent
Dr. Rens L.P.H. Van Beek

MSc thesis for the Master's Program Earth Surface and Water
30 ECTS
Faculty of Physical Geography
Utrecht University

**Content**

1. Introduction

   1.1. Research questions

2. Methods

   2.1. Data sets characteristics

   2.2. Data processing

   2.3. Model development for five locations (Local analysis)

   2.4. Global analysis

   2.5. Predictive model

   2.6. PCR-GLOBWB

3. Results

   3.1. Model development for five locations (Local analysis)

   3.2. Global analysis

   3.3. Predictive model

4. Discussion

5. Conclusion

6. Outlook and Recommendations

References

Appendix A: Parameters for climate variables in multivariate analysis for five locations

Appendix B: Correlation coefficient between climate variables for five locations

Appendix C: Time series of climate variables in multivariate analysis for five locations

Appendix D: Multivariate analysis in a python script

Appendix E: Global maps of the mean LAI of the plant functional type per grid cell

Appendix F: Global maps of the maximum LAI of the plant functional type per grid cell

Appendix G: Global maps of the minimum LAI of the plant functional type per grid cell

Appendix H: Global maps of the standard deviation of LAI of the plant function type per grid cell

Appendix I: Time series of climatic variables in ANPI formula for five locations

Appendix J: Time series of climatic variables in Jolly formula for five locations

Appendix K: Time series of climatic variables in S Formula for five locations

**Abstract**

A novel model was developed to find the dynamic relationship between the climate variables and regional leaf area index (LAI). The NASA MODIS datasets were used for the LAI and Land Use Data. The Climate Research Unit (CRU) provided meteorological data. The model was developed in Python programming language with the NumPy and Gdal packages. The examination of the climatic variables reveals that temperature, precipitation, vapour pressure, and potential evapotranspiration influence LAI the most on a regional and global scale. The local analysis was performed for five locations at the Temperate and Tropical Zone of the Köppen's Classification. The results revel relatively good prediction. By taking additional parameters in the model (for instance radiation) the model could improve its performance at the local scale. The global analysis of the most influential climate variable shows strong connection between potential evapotranspiration and precipitation at the equator. The temperature and vapour pressure increase in importance with northern and southern direction. The climate variables, that influence the regional LAI, depend more on the climate zone then plant functional type. The satellite data is very sensitive and LAI measurements are dependent on the correct interpretation of the data and on accurate filtering measurements errors (clouds). To improve the predictive equation for the LAI estimation, climate variables should have a memory of the previous months.

**Key words:** predictive model, LAI, satellite dataset (MODIS), Köppen's Climate Zone, python

## Preface

This Master Research has been written to fulfil the graduation requirements of the Master's programme Earth Surface and Water at Utrecht University in the Netherlands. I was engaged in researching and writing from November 2017 to August 2018.

This research allows me not only to understand the complex hydrological – vegetation relationship, but also to develop some additional programming skills. Since, I developed this skill by myself the datasets pre-processing caused a lot of time and effort in the limited time for this research. The countless hours, days, nights of programming and frustration, cannot be expressed here, neither the satisfaction of the achievement.

I would like to thank my supervisors, Dr. Rens L.P.H. Van Beek and Dr. ir. Ruud van der Ent, for their guidance and support during this process.

My parents deserve a particular note of thanks: thank you for your love, unfailing encouragement and support.

Finally, I would like to dedicate this thesis to my daughter Amelia and her father Stephan, without whom I would have never (or earlier) completed it.

I hope you enjoy your reading.

Anna Teuerle

Amsterdam, August 30, 2018

## 1. Introduction

Seasons and foliage duration of plant vegetative cycles, determine exchange of carbon dioxide and water between the Earth's surface and the atmosphere (Jolly et al., 2005). Over the last 50 years climate change has altered the vegetation (Betts et al., 1990), thus this exchange differs in time and its potential capacity. The long-term change in environmental conditions affect the vegetation, as a result of which nutrient cycles, microbial activities, and physiological activities of plants differs (Mohammad, 2013). The vegetation determines regional and continental hydrological cycle by factors like: the water holding capacity of the vegetation system, the transport of water to deeper levels, and transport of moisture within the ground (Rind, 2013). Thus, to improve understanding of the hydrological response to climate change, the variability in seasonal canopy characteristics must be implemented in the hydrological model.

To follow the rapid climate change, and its effect on interannually variability, the hydrological model requires dynamic calculations. Environmental conditions are caused by climate variables, that increasingly differ annually with climate change. Therefore, the demand of climate variables to determine response of the hydrological models increases (Nielson et al., 1994). However, most of them neglect the link of driving variables, such as climatic conditions, and use fixed vegetation phenology parameters (Sellers et al., 1996; Jolly et al. 2005). The hydrological model to describe vegetation use parameters, such as leaf area index (LAI), derived by the satellite-observations (Baret and Guyot, 1991). LAI is a dimensionless quantity that characterizes plant canopies, defined as the one-sided green leaf area per unit ground surface area (LAI = leaf area / ground area, m2 / m2) in broadleaf canopy (Watson, 1947). In most of the hydrological models the LAI have a prescribed seasonal function. In another words, LAI estimations are prescribed for an ''average year''. This means, that observed extreme events, such as lengthy dry and wet periods, are not implemented in the predictions.

High-accuracy spatial distribution of LAI is an essential contributor to model atmosphere-biosphere interactions (Tang et al., 2014). The attempt to derive LAI based on climatic conditions has been a subject of very few studies. For instance, Tesemma et al. (2015) developed a regional nonlinear model for estimating changes in LAI due to climatic fluctuations. The results demonstrated the importance of including the effects of changes in LAI in projections of discharge. While, Tesemma et al. (2015) developed a regional model, Jolly et al. (2005) proposed a globally generalized index that describes regional vegetation from climatic conditions. A good agreement (r>0.8) between the developed Growing Season Index (GSI) and Normalized Difference Vegetation Index (NDVI) was demonstrated. However, Jolly et al. (2005) did not propose a direct and dynamic calculation of LAI. Thus, a general formula (or series of formulas) to describe the link between climate variables and global LAI has still to be developed.

The main objective of this research is to examine the effects of climate variables (precipitation, temperature, vapour pressure deficit, and evapotranspiration) on the seasonal changes of LAI. The aim is to develop a predictive model to simulate the dynamics of local LAI for different land cover types. The sought formula can be equally applied under future or past conditions. The constituent regression analysis, of the time series of climate data, highlights the contribution of the individual meteorological variables to temporal and spatial variations in LAI. Moreover, the model provided an important opportunity to advance the understanding of using dynamic LAI calculations with 'meteorological memory', thus by using the meteorological dataset over the previous period of time to predict current LAI. This study therefore sets out to assess the effectiveness of dynamic LAI predictions and the influence of short-term climate conditions on current LAI predictions. This regression analysis is developed and illustrated on the basis of five characteristic locations that represent various climatic zones and land cover types. Future studies are expected to implement the formula in the hydrological model PCR-GLOBWB, and to demonstrate whether modelling dynamic LAI, in the way that responds to climatic conditions, is a key factor for modelling discharge of the study area.

### 1.1. Research questions

A novel method must be developed to answer the following research questions:

1) What are the main variables (e.g., precipitation and temperature) that influence LAI?
2) What is the relationship between those variables and LAI? How does it vary regionally, globally, and in different climatic zones?
3) What can be improved to obtain a predictive equation for LAI estimation, without using remote sensing, for any given time and place?

Therefore, the objectives of this studies are (i) examination of the climate variables that influence LAI on a regional and global scale; (ii) development of relationship(s) between these variables and LAI as a numerical predictive model; (iii) spatial distribution of the climatic variables that influence LAI in the global scale; and (iv) evaluation of the results.

## 2. Methods

This section provides details about the data set, the characteristics of the selected study areas and the modelling approach. The data characteristics used in this research are briefly described in sect. 2.1. The data processing is explained in sect. 2.2. The model development for five locations (local analysis) is presented in sect. 2.3, and the global analysis is described in 2.4 The various dynamic formulas (predictive models) at widely diverse locations are judge against satellite-derived LAI (MODIS15A) in sect. 2.5. The possible application in the hydrological model is described in sect. 2.6.

### 2.1. Data sets characteristics

The LAI dataset (MODIS15AH2) and Land Use Type Cover dataset (MODIS12Q1) are obtained from the MODIS (Moderate Resolution Imaging Spectroradiometer) platform run by the NASA Earth Observing System. The data sets are derived from the satellite imagery and describe features of the Earth's surface and the atmosphere, that are used worldwide for studies of processes and trends on local to global scale (www.lpdaac.usgs.gov). The meteorological dataset is provided by the CRU (Climate Research Unit) from various datasets of climatic variables that they offer (www.cru.uea.ac.uk).

#### 2.1.1. The LAI dataset

The leaf area index (LAI) data obtained from the MODIS LAI/FPAR product is in a Sinusoidal grid at 500 m resolution. The algorithms choose the "most accurate" pixel available from all the acquisitions of both MODIS sensors located on NASA's Terra and Aqua satellites from within an 8-day period (LPDAAC, 2018). However, to obtain the monthly LAI, the 8-day value is recalculated to the same eight values for each of the eight days. Then an average value for a month (30 or 31 days) is calculated. The dataset provides direct LAI value with a valid range from 0-100, over a 10-year period (2001-2010).

#### 2.1.2. The Land Use Type Cover dataset

The MODIS Land Cover Type product (MCD12Q1) dataset provides data characterizing five global land cover classification system with a resolution of 500 m (LPDAAC, 2018). In the research the Type 5 Classification is used with 11 land cover types (Table 1). However, LAI represents natural vegetation, thus for the calculations were taken land cover types of numbers from 1-8. The classification is based on the Plant Functional Type (PFT) scheme that is measured annually at a global scale. This classification has been chosen for its diversity in natural vegetation types, and for excluding detailed characteristics of anthropogenic land covers. This paper used the dataset from the year

2010 with an assumption that it represents the natural vegetation in the year period 2001-2010.

*Table 1. Land Cover Types Description: Type 5 classification (lpdaac.usgs.gov)*

| Class | PFT (Type 5) |
|---|---|
| 0 | Water |
| 1 | Evergreen Needleleaf trees |
| 2 | Evergreen Broadleaf trees |
| 3 | Deciduous Needleleaf trees |
| 4 | Deciduous Broadleaf trees |
| 5 | Shrub |
| 6 | Grass |
| 7 | Cereal crops |
| 8 | Broad-leaf crops |
| 9 | Urban and built-up |
| 10 | Snow and ice |
| 11 | Barren or sparse vegetation |
| 254 | Unclassified |
| 255 | Fill Value |

### 2.1.3. Meteorological data

Climatic Research Unit represents the meteorological data for four climate variables: precipitation (mm/month), temperature (°C), vapour pressure (hPa), and potential evapotranspiration (mm/month) (CRU TS v. 3.24) from 2000 to 2010. These climate variables are chosen based on the preceding literature review (Jolly 2005, Tesemma 2015) as the most influential regional LAI climatic conditions. The dataset is at a 0.5° by 0.5° resolution with the monthly average value for each location over a 10-year period (2001-2010).

## 2.2. Data processing

The data processing was made in Python with a NumPy library. Transformations where done using gdal extracted metadata. Visualization was done with Matplotlib extension. The data was stored in Hierarchical Data Format 5 (HDF5) to build a multidimensional array that connects the datasets and allows manipulation within the NumPy environment. For each granule of MODIS grid were created an hdf5 file which contained all datasets.

### 2.2.1. Data preparation

To be able to do computations with the three different dataset sources, they need to correspond in dimensions and time. The CRU data is at the global scale at 720 longitude degrees and 360 latitude degrees. This covers the entire globe like a sphere. It is important to realize that each position in the dataset is not a square, although it is represented this way. For the local analysis the meteorological dataset is selected based on the latitude and longitude of the place. However, for the spatial distribution of the LAI driven variables, meteorological datasets are selected based on the LAI grid borders that are represented by latitude and longitude. The MODIS datasets have a much higher resolution then CRU data. Each MODIS tile is 1200 by 1200 km and represents 500 m by 500 m cell. Thus, the resolution is 2400 by 2400 pixel. The climatic dataset uses Latitude Longitude of wg84 of ESPI:4326 in 0.5 degrees. MODIS grid is at the equator 10 by 10 degrees, and the cover area decreases in both directions from the equator.

### Data correction

Remote sensing datasets, as MODIS, contain errors that occur during the measurements. The mistakes are due to the natural obstacles (ex. clouds, pollution) between the satellite sensors and Earth's surface. In the measurements the errors from MODIS15A were smoothed using the Savitzky Filter. The filter was described by Savitzky and Golay (1964) and the methodology described by Chen et al. (2004) offers easy implementation. The filter itself is based on the polynomial fit within a width of the window. The data points are replaced by the polynomial function, with specific degree, in a dedicated width of the window that surrounds the point. In this paper, the degree of 4, and width of 9, is used (Figure 1).

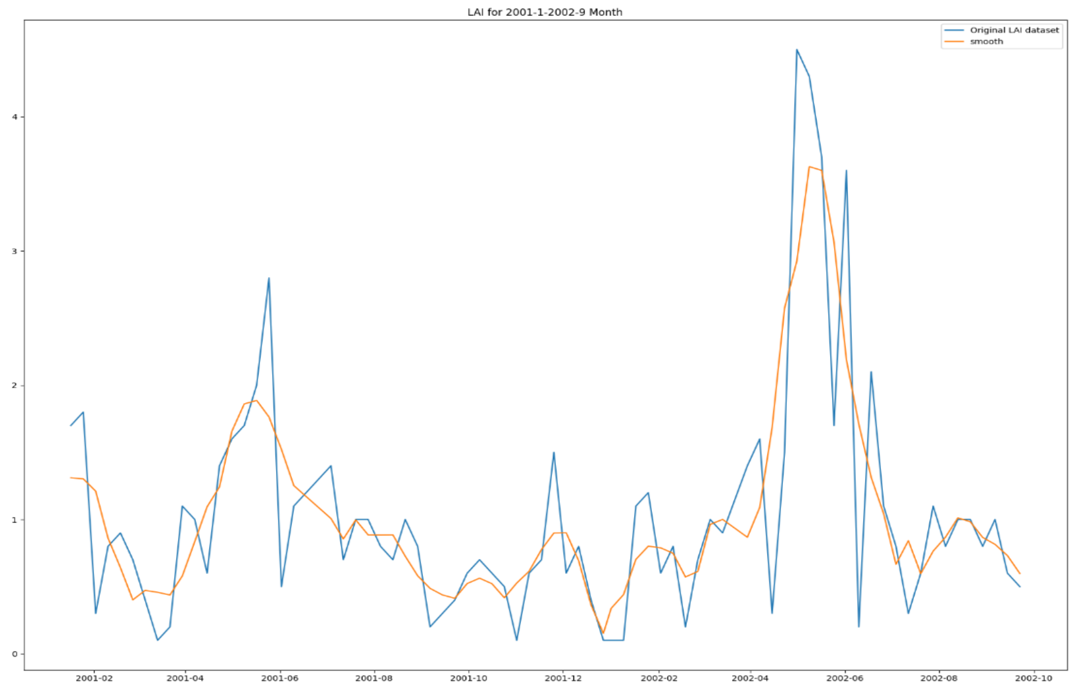



*Figure 1. LAI time series correction in Germany. The blue line indicates the original dataset. The orange line indicates corrected (smoother) time series of the LAI.*

**Standardization**

Prior to the regression analysis, the meteorological data and LAI dataset were standardized (Equation 1).

$$Z\big(x(i)\big) = \frac{x(i) - \text{average}(x)}{\text{standard deviation }(x)} \quad (1)$$

### 2.2.2. Multivariate regression analysis

Multivariate regression analysis is used as a statistical technique which uses more than one variable to generate the output. In climate modelling, the key to the use of regression analysis is the assumption that a linear fit of the meteorological variable in three dimensions is relatively correlated in the chosen area (Splitt, 2002).

### 2.2.3. Akaike information criterion (AIC)

The quality of the LAI predictive model is measured by Akaike information criterion (AIC), that was proposed by the statistician Hirotugu Akaike (Akaike, 1973). The AIC is a world-wide known estimator of quality for statistical models. It offers an estimation of the information lost by using a given model for the calculations of the results, in relative to the other models. It can be used to find compromise between correctness of the fit and simplicity of the model.

The AIC formula is based on the number of estimated parameters *k*, and the maximum value of the likelihood function *L* (Equation 2).

$$\text{AIC} = 2\text{k} - \ln(\text{L}) \quad (2)$$

A predictive model, with a low root-mean-square-error and a low AIC, is one to be considered to be applied globally. There is an infinite amount of possible combinations of predictors for LAI. Therefore, the AIC criterion is used as a tool to limit model options.

## 2.3. Model development for five locations (local analysis)

Analysis of the historical LAI datasets for five characteristic study areas in terms of their climatic conditions (temperature, precipitation, vapour pressure, potential evapotranspiration) was conducted. All five study-locations were chosen to represent the natural vegetation of different climatic zones with contrasting meteorological conditions (Figure 2).

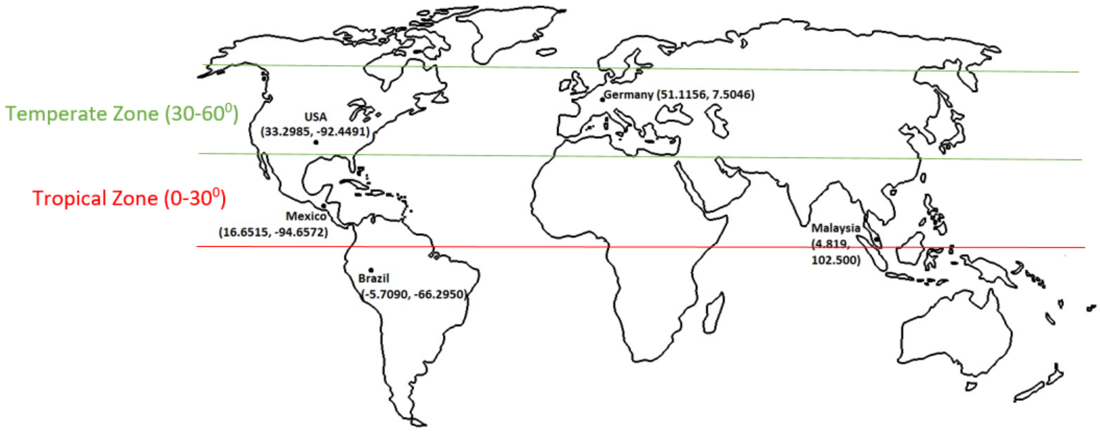



*Figure 2. Locations of study areas, with their latitude and longitude, respectively (coast line from www.didu.me).*

*Table 2. Land Cover Types Description: Type 5 classification for five study locations.*

| Location | Class | Plant Functional Type (PFT) |
|---|---|---|
| Germany | 4 | Deciduous Broadleaf trees |
| Brazil | 2 | Evergreen Broadleaf trees |
| Mexico | 2 | Evergreen Broadleaf trees |
| USA | 4 | Deciduous Broadleaf trees |
| Malaysia | 2 | Evergreen Broadleaf trees |

The meteorological dataset was represented by 120 data points (monthly values over a 10-year period) for each climate variable (temperature (*T*), precipitation (*P*), vapour pressure deficit *(V)*, and potential evapotranspiration *(PET)*). The dataset is optimized with the historical LAI dataset by using the lstsq SciPy package to predict the LAI over the same period. After the calculations were done for the local analysis, the further computations were made for the spatial distribution most influencing climate variable on the regional LAI. The calculations were done for every 0,5$^0$ by 0,5$^0$ area. In each area were collected all pixels of the LAI dataset, that represents natural vegetation. They were combined with the meteorological dataset.

The analysis required developing a few simple formula(s), that based on the time series of climate and historical LAI dataset of the location, predicts the LAI. The formulas share the shape, however the variances of the climate variables were optimized for each of the five locations (Equation 3).

$$LAI\ (predited) = A * temperature + B * precipitation + C * vapour\ pressure + D * potential\ evapotranspiration \quad (3)$$

Where A, B, C, D are the variances of climate variables for each location that optimize the prediction.

To determine the contribution in the LAI by each of the climate variables, the unexplained variance (UV) was calculated. UV is the fraction of variance of dependent variable (predicted LAI) which cannot be explained by independent variables (climate variables).

## 2.4. Global analysis

The global analysis required obtaining all the MODIS15A (LAI) grids for the period 2001-2010 (Figure 3) and all the MODIS12Q1 (Land Type Cover) grids for the year 2010

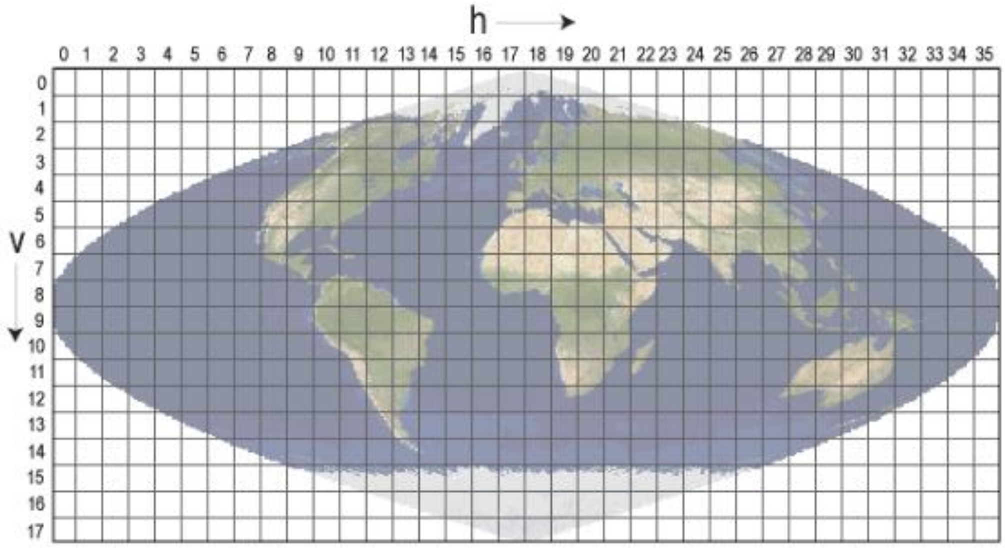



*Figure 3. Modis grid distribution in horizontal and vertical direction (www.modis-land.gsfc.nasa.gov)*

There are 460 non-fill tiles. The size of tiles is 10 degrees by 10 degrees at the equator. The size changes according to the latitude and longitude of the grid. The tile coordinate system starts at (0,0) (horizontal tile number, vertical tile number) in the upper left corner and proceeds right (horizontal) and downward (vertical). The tile in the bottom right corner is (35,17) (www.modis-land.gsfc.nasa.gov).

Global analysis was made by using SciPy and NumPy packages and resulted in eight global maps at the resolution of 0,5 degrees. Maps represented the dominant climate variable of the eight plant functional type. To support the results of the dominant climate variable of the plant functional type per, global LAI data were visualized as mean LAI over a 10-year period, maximum and minimum LAI, and standard deviation of the calculations for each dominant plant functional type. The maps are in the 720 by 360 pixels resolution. Each pixel corresponds to $0.5^0$ by $0.5^0$ and contains three types of datasets: meteorological dataset, LAI and Land Use. Land Use data was used to determine the type of the vegetation. The meteorological dataset and LAI dataset covers 10-year period.

### 2.5. Predictive model

Three formulas were used to calculate predictions of LAI. The formulas were compared at the local scale and global scale by using RMSE.

#### Antecedent Net Precipitation Index (ANPI)

The Antecedent Net Precipitation Index (ANPI) is used as an extended formula for the multivariate regression analysis of the 5 study locations. The function includes a conditioning factor as growing degrees days (GDD), that are days when the temperature persistently is below 5°C.

$$LAI\ (predicted) = a * precipitation + b * (precipitation\ from\ previous\ month) + c * (precipitation\ from\ two\ months\ before) + d * temperature\ above\ 5\ degrees \quad (4)$$

#### Jolly Formula

The Jolly Formula is a transformed formula from Jolly et al. (2005) inspired by globally generalized index that describes regional vegetation from climatic conditions.

$$LAI\ (predicted) = a * \left(\frac{temperature + 20^0C}{70^0C}\right) + b * (1 - \frac{vapour\ pressure + 0.9kPa}{5kPa}) \quad (5)$$

#### SFormula

SFormula is the developed formula in this paper, based on literature and results from the previous prediction models. Once the software was completed a large amount of models were experimented with. Experience was gained over which models or part of models were the best predictors. With the AIC criteria as guidance, a selection of model functions and functions with time memory was used to create a new model which outperformed previous models on the entire globe.

$$LAI\ (predicted) = a * potential\ evapo + b * (precipitation\ from\ two\ months\ before) + c * temperature + d * temperature\ above\ 5\ degrees \quad (6)$$

### 2.6. PCR-GLOBWB

The proposed LAI estimation improvements are dedicated to being built in future into the existing global hydrological model: The PCRaster Global Water Balance model (PCR-GLOBWB) developed by Van Beek & Bierkens (2008). The model provides a grid-based representation of terrestrial hydrology. The spatial resolution is 30′ (~50x50 km) or 5′ (~10x10 km), thus the current resolution of the dynamic LAI model has to be adjust. The developed predictive formulas can be used in the seasonal prediction, to measure the hydrological effects of climate variability and to observe changes in climate.

## 3. Results

The results are divided into three sections: first is a local analysis for five study locations (sec. 3.1), second spatial distribution of climatic variables at the global scale for different land covers (3.2) and predictive model (3.3).

### 3.1. Local analysis for five study locations

The multivariate linear regression was obtained for the five study locations. The predictive model was run using Python to estimate the LAI from the four climatic variables (temperature, precipitation, vapour pressure, and potential evapotranspiration).

### Deciduous Broadleaf trees

Deciduous Broadleaf Trees land cover is represented by two locations: Germany and USA (Arkansas). According to the Köppen's climate classification, they are both located in the Temperate Zone (30-60°). The climate is characterized by a smaller angle of solar radiation, more regular distribution of precipitation and differing daylength over the course of the year. In the time series of climate variables (Appendix B) are clear annual patterns for the temperature, vapour pressure, and potential evapotranspiration. The effect of this annual seasonality can be seen on the local LAI time series. The predictive model estimate LAI over the same period of time, 2001-2010. The comparison of the time series shows smoother values of LAI for the predictive model (Figure 4, 5), except for the low temperature in the USA, where predictive model estimates lower LAI then in reality.

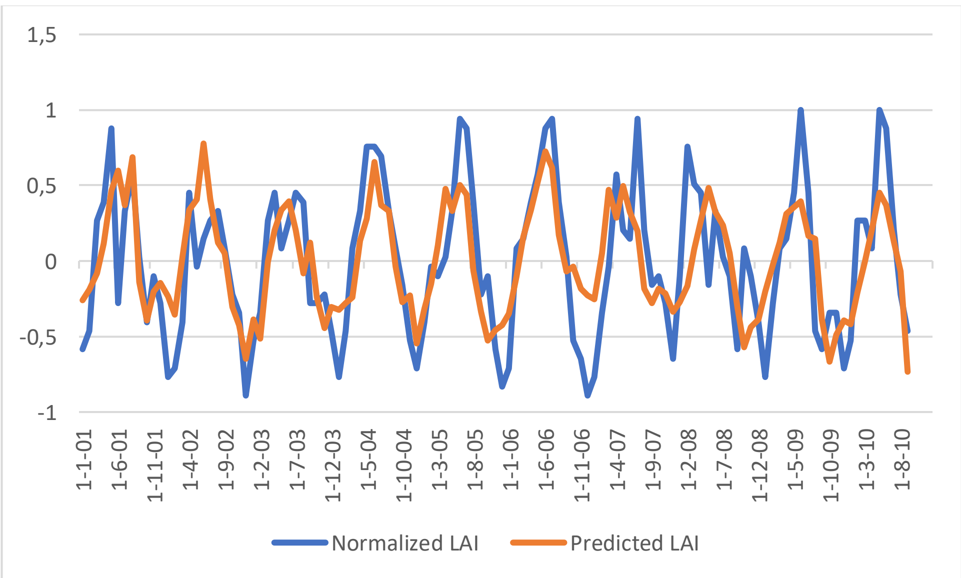



*Figure 4. Time series of Original Normalized LAI (blue line) and Predicted LAI (orange line) from the climate variables over a 10-year period (2001-2010) in Germany.*

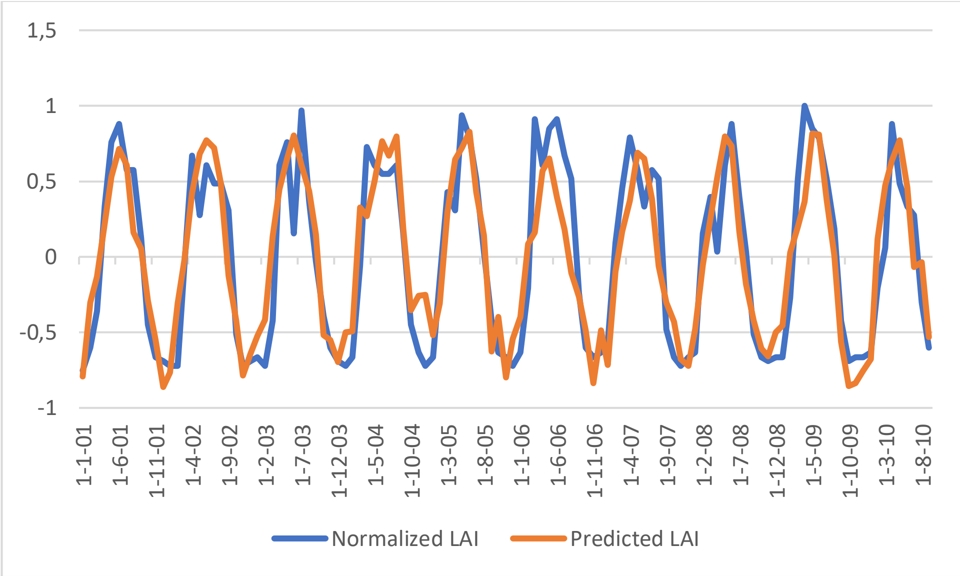



*Figure 5. Time series of Original Normalized LAI (blue line) and Predicted LAI (orange line) from the climate variables over a 10-year period (2001-2010) in USA.*

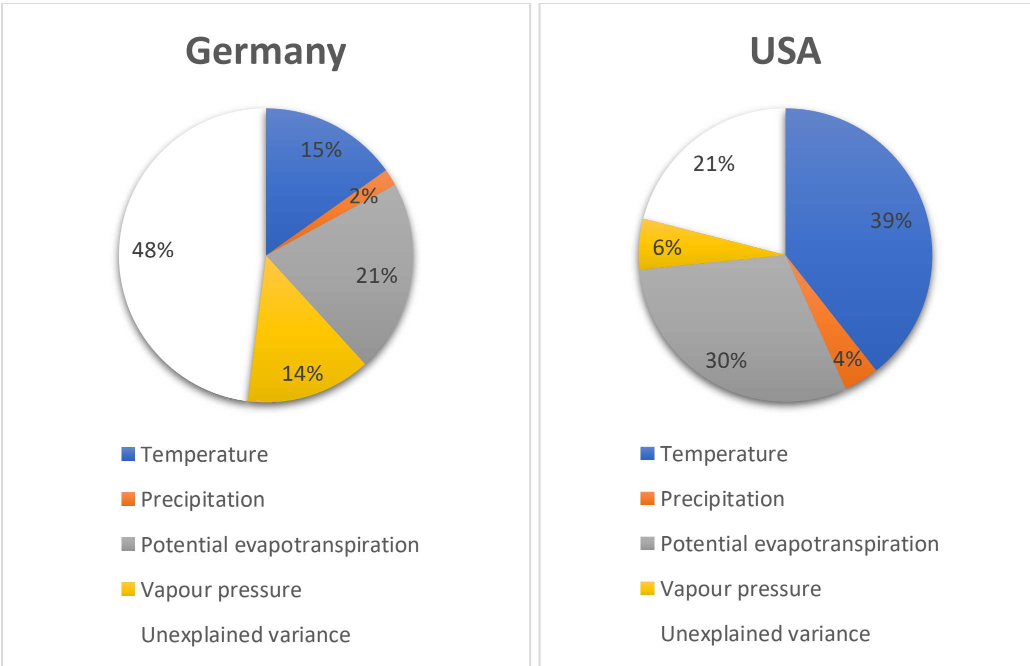



*Figure 6. Pie charts with the percentage distribution of the climate variables that influence the local LAI.*

The unexplained fraction (UV) differs for both of the locations: Germany 48%, USA 21%. LAI and temperature varies more extremely over the year in Germany than in the USA. The temperature is the most dominant climate variable for both locations. However, it has more influence on the local LAI in the USA than in the Germany. The potential evapotranspiration is important in both of the location with higher percentage value in the USA. This might be due to a warmer climate in the USA (with increasing temperature, potential evapotranspiration is higher).

**Evergreen Broadleaf Trees**

Evergreen Broadleaf Trees land cover is represented by three locations: Brazil, Mexico, and Malaysia. According to Köppen's climate classification, they are all situated in the Tropical Zone (0-30°). The characteristics for this zone are very warm conditions caused by nearly vertical solar radiation at noon through almost the entire year. However, the elevated temperature causes high evaporation and high moisture in the air. This results in dense cloud cover that reduces the effect of solar radiation.

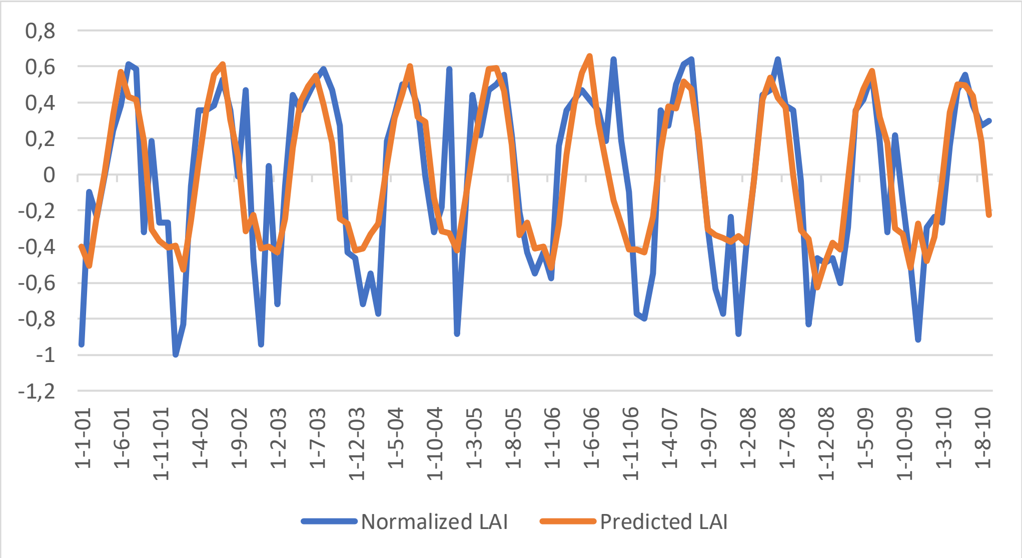



*Figure 7. Time series of Original Normalized LAI (blue line) and Predicted LAI (orange line) from the climate variables over a 10-year period (2001-2010) in Brazil.*

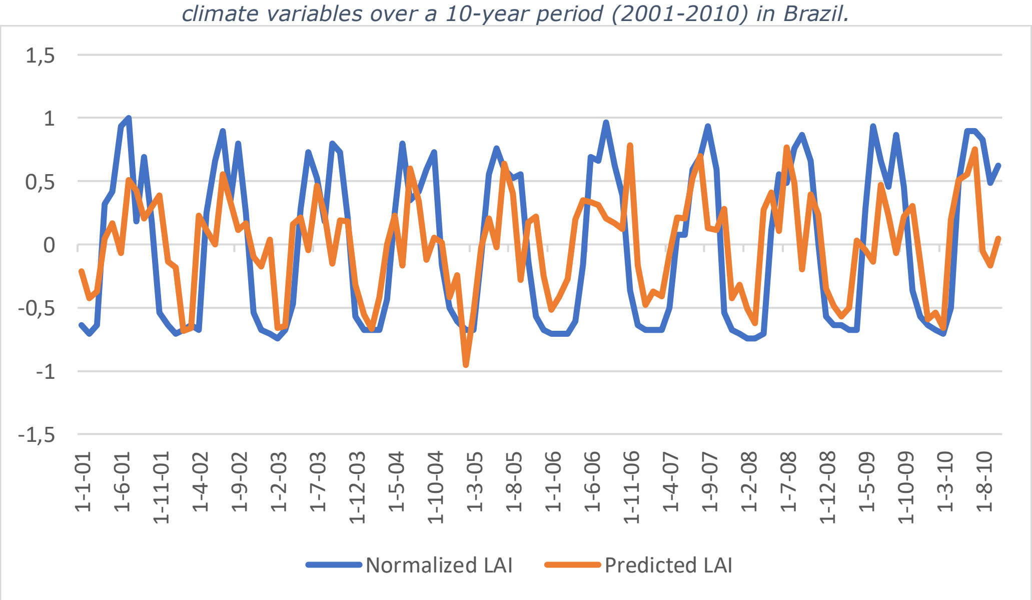



*Figure 8. Time series of Original Normalized LAI (blue line) and Predicted LAI (orange line) from the climate variables over a 10-year period (2001-2010) in Mexico.*

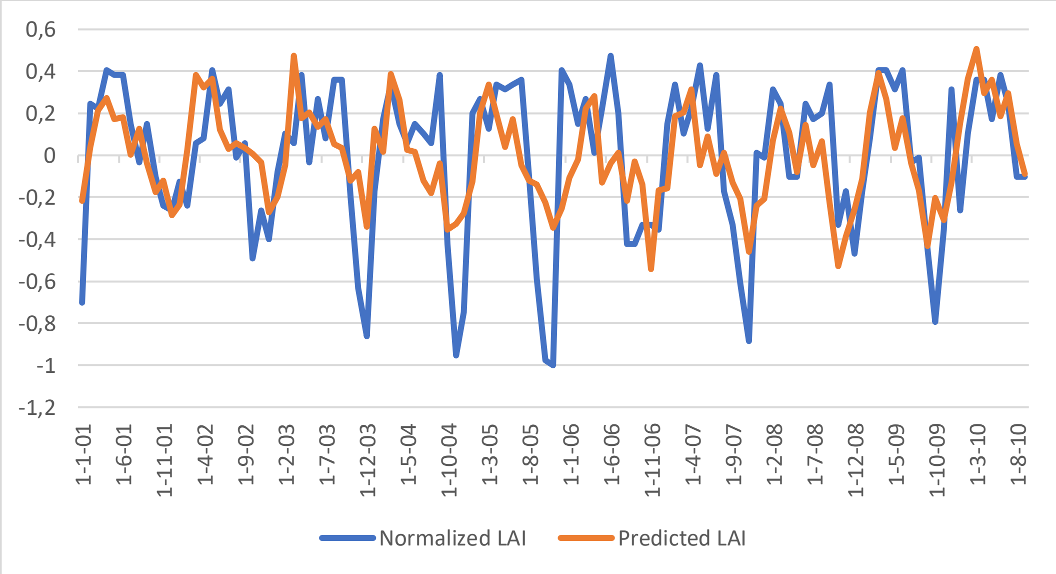



*Figure 9. Time series of Original Normalized LAI (blue line) and Predicted LAI (orange line) from the climate variables over a 10-year period (2001-2010) in Malaysia.*

In comparison to the local LAI of the Deciduous Broadleaf Trees (DBT), the estimations of the predictive model for Evergreen Broadleaf Trees (EBT) are less accurate. The variability of the annual LAI is also much lower in the EBT. The extremely low values are presented in the Malaysia and Brazil, which are underestimated by the prediction.

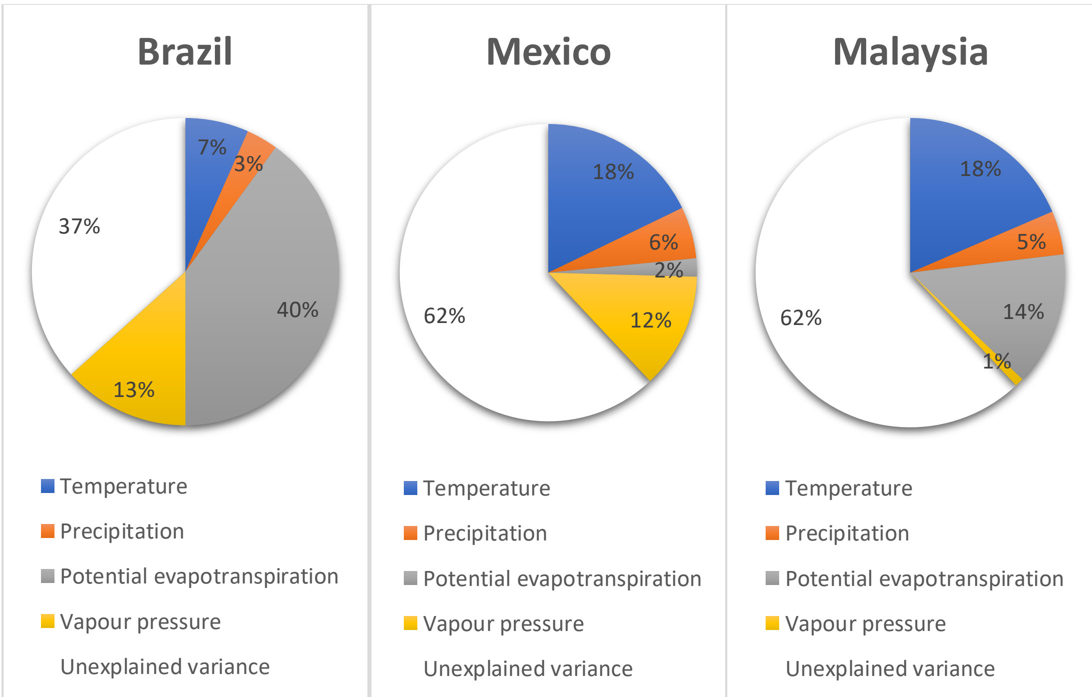



*Figure 10. Pie charts with the percentage distribution of the climatic variables that influence the local LAI.*

The unexplained variance is much higher for Evergreen Broadleaf Trees Land Type cover (37% - 62%). Locations have different dominant climate variables. Brazil's most important climatic variable that influences local LAI is precipitation.

The time series shows clear seasonality of the precipitation, that influences vegetation, thus LAI. Moreover, potential evapotranspiration is also high in importance and negatively correlated to precipitation (Appendix B). The reason is that with high precipitation there is low potential evapotranspiration, however, after precipitation the elevated temperature on the equator allows for high potential evapotranspiration, that also stimulate the growth of the plants. Mexico has two dominant variables: vapour pressure and temperature. The dominant climate variable for Malaysia is temperature, that shows the strongest seasonality in comparison to other climate variables. Vapour pressure might also influence local LAI because the air moisture is lower than in the Brazil. Malaysia's dominant climate variables is temperature, that has the strongest seasonality from all the climate variables in this typical tropical zone.

*Table 3. Statistical properties over time series for multivariate analysis of five locations.*

| Location | Average Original LAI | Standard deviation of Original LAI | RMSE | $R^2$ |
|---|---|---|---|---|
| Germany | 1.52 | 0.82 | 0.34 | 0.52 |
| USA | 2.62 | 1.95 | 0.26 | 0.79 |
| Brazil | 4.03 | 1.77 | 0.29 | 0.63 |
| Mexico | 2.83 | 1.82 | 0.49 | 0.38 |
| Malaysia | 4.42 | 1.69 | 0.28 | 0.38 |

The average mean value of LAI for the Deciduous Broadleaf Trees Land Type Cover is relatively low 1.52 and 2.62, for Germany and the USA respectively (Table 3). Standard deviation for Germany is unexpectedly low (0.82), that means that LAI values does not differ much from the mean value over the year. However, the USA shows high standard deviation with very different values of LAI over a year. The cause might be smaller deviation in temperature for Germany, sometimes with values below 0°C. The root-mean-square error (RMSE) is in the similar range for both types of land covers.

The average mean value of LAI for the Evergreen Broadleaf Trees Land Type Cover is high 4.03, 2.83 and 4.42, for Brazil, Mexico, and Malaysia, respectively (Table 3). The standard deviation is similar for all locations. The mean higher values of LAI are more difficult to predict then relatively low mean values.

### 3.2. Spatial distribution of climate variables at the global scale for different land type covers (global analysis)

The first set of analyses examined the impact of climate variables on the vegetation of each land type cover.

### 3.2.1. Global maps of the highest mean LAI and its representative plant functional type.

The combined datasets of climate data, LAI data and Land Use data were analysed at the global scale. The outputs produced by the model consists of time series of global maps comprised of $0.5^0$ by $0.5^0$ cells containing the values of the parameters. However, there are also seen errors at the regional scale that are caused by inferior quality data and limited factors that were used to force model with the prediction.

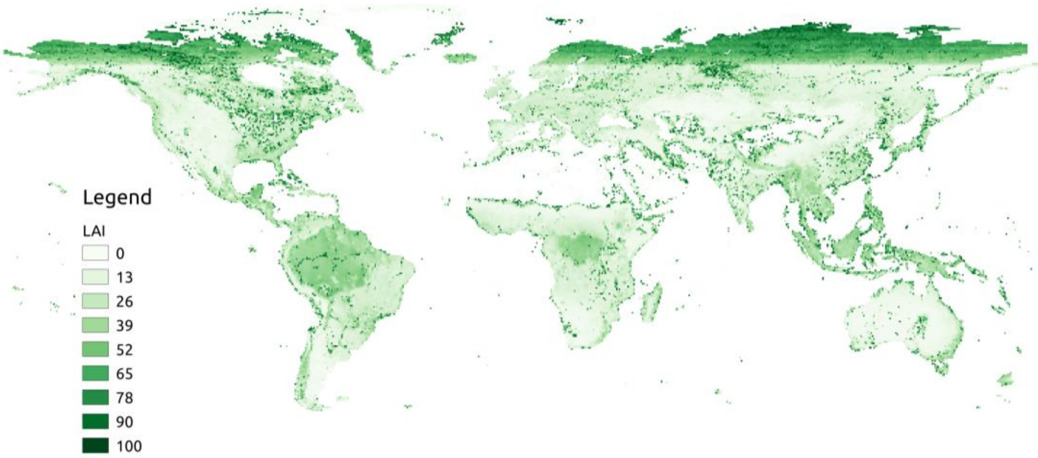



*Figure 11. Global map of the maximum mean LAI of the plant functional type per grid cell.*

Global maps were created in order to analyse climate influence on the LAI pattern. The Land Use dataset was used to select natural vegetation within each $0.5^0$ by $0.5^0$ area. The natural vegetation represented eight plant functional types. Further, for each plant functional type were determined the maximum, minimum, mean LAI, and additionally standard deviation of LAI over a 10-year period. Based on those results, the global maps were created for each plant functional type (Appendix E, F, G, H).

Because of much higher resolution of MODIS data, each $0.5^0$ by $0.5^0$ area contained several plant functional types, thus several mean, maximum and minimum LAI outputs. To compare all the results, additional maps were made. The highest mean LAI were taken for each $0.5^0$ by $0.5^0$ (Figure 11) and the plant functional type to which it belongs (Figure 12) to determine the most dominant plant functional type that represents LAI over $0.5^0$ by $0.5^0$ area.

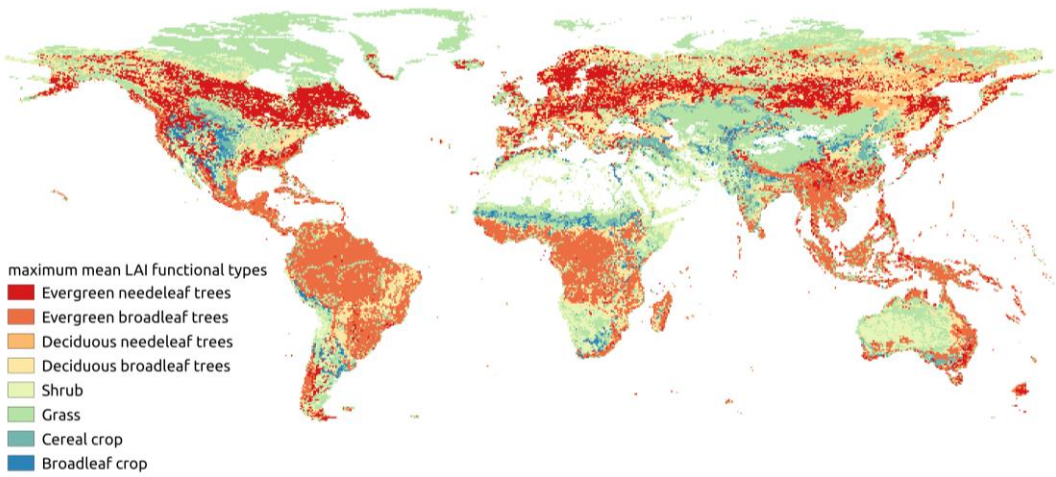



*Figure 12. Global map of the dominant plant functional type that represent the maximum mean LAI.*

The global well-known vegetation pattern is clearly seen on the global map of the dominant plant functional type that represents the highest mean LAI over the area. Moreover, the highest mean regional LAI is driven by the same plant functional type at the same climatic zone (the horizontal pattern). The tropical zone (0-30$^0$) is mostly dominated by the Evergreen Broadleaf Trees that response to the warm climate conditions of the equator. This is seen in South America, Africa, as well as Indonesia. The highest precipitation also occurs at the equator as well as high evapotranspiration. The climate conditions allow for the optimal growth of the plants of the year, thus the high mean LAI. As further to the northern and southern direction, the more dominant becomes Evergreen Needleleaf Trees which is characteristic for the plant type that can survive more extreme and various climate conditions. The grass is a dominant land cover type for the LAI at the 40 degrees in America and in Europe which is temperate zone. In the Asia grass land cover corresponds to the higher altitude of the region. Deciduous Needleleaf trees represents LAI mostly in the harsh conditions of the North-Western Asia. The lack of the optimal climate conditions (low precipitation, low temperature) exclude other types of the plant functional type. Shrubs are the most dominant plant functional type of the LAI in Australia. The regional LAI, that is determined by plant functional type, is strongly correlated to the annual variations of the climate variables.

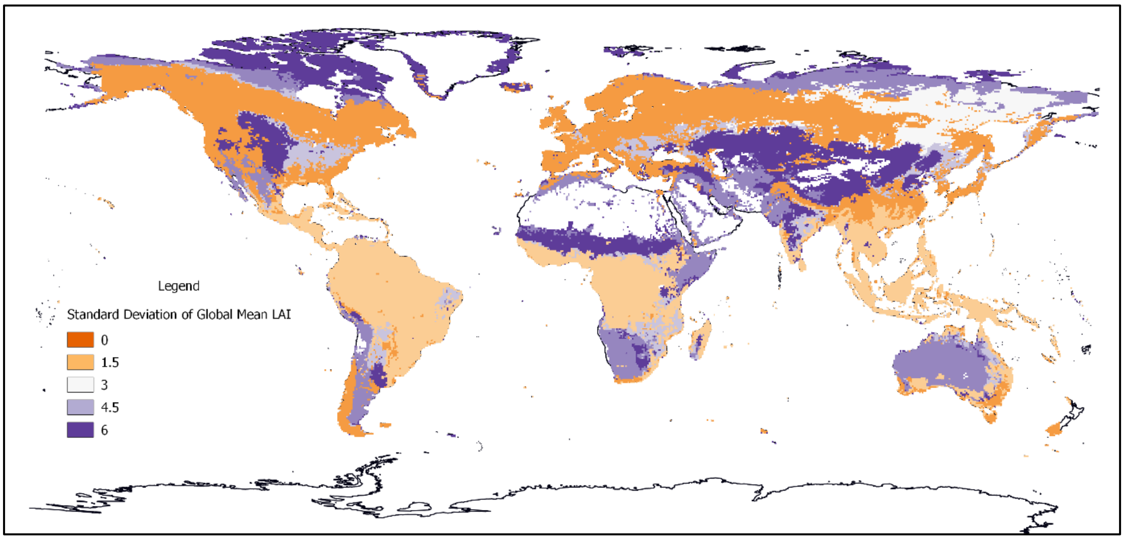



*Figure 13. Standard deviation of the global mean LAI type per grid.*

The map reveals the difference in the standard deviation of the global mean LAI type per grid cell (Figure 13). The difference is correlated to the climate conditions and plant functional type of the region. The optimal climate conditions of the equator allow for the sustainable annually growth of the plants, thus the standard deviation is relatively low. In the northern hemisphere (North America, Europe, and North-Eastern Asia) the standard deviation of the mean LAI over the year is very low. Comparing the dominant plant functional type that represent the Max Mean LAI map (Figure 12), it can be seen that this is strongly correlated to the plant functional typ. The areas that are mostly represented by the Evergreen Needleleaf Trees, are less effected by climate conditions and are more dependent on the groundwater resources, as well they can survive harder climate conditions. In opposition, Grass land type cover has high standard deviation and responds to the seasonal changes faster.

### 3.2.2. The dominant climate variables that determines the regional LAI for plant functional type.

To determine the direct connection between climate variables and regional LAI at the global scale, eight maps were generated (for each plant functional type). They reveal the most dominant of the four climate variables (precipitation, temperature, vapour pressure, or potential evapotranspiration) that determines the regional LAI. It has to be noticed that those variables are strongly correlated to each other, and where in the local scale it was possible to measure, this was the limitation for the global scale.

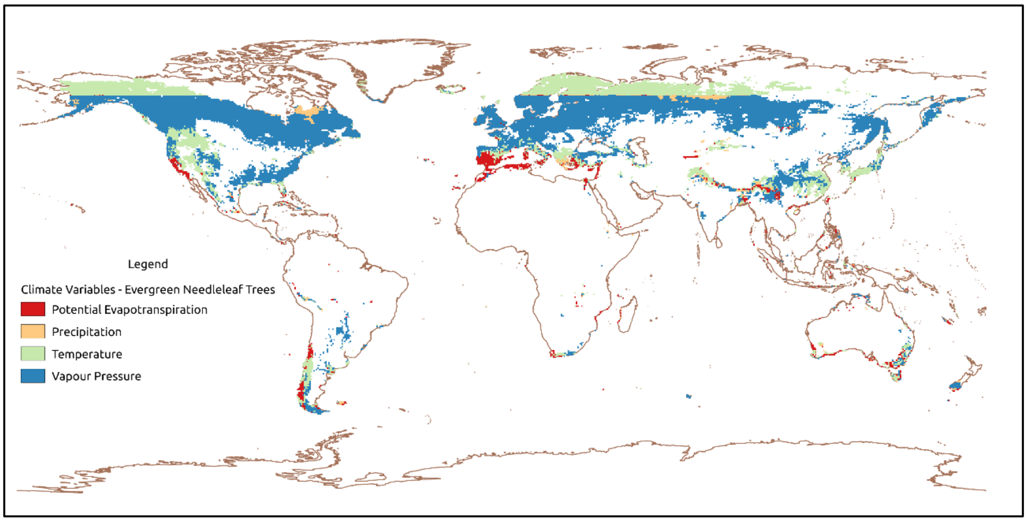



*Figure 14. Spatial distribution of the climate variables with the dominant influence on regional LAI for Evergreen Needleleaf Trees.*

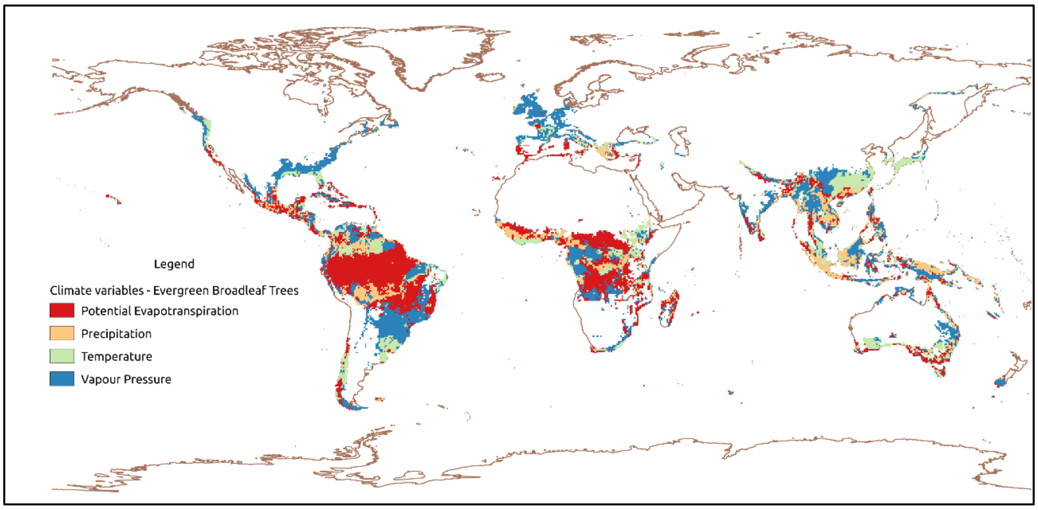



*Figure 15. Spatial distribution of the climate variables with the dominant influence on regional LAI for Evergreen Broadleaf Trees.*

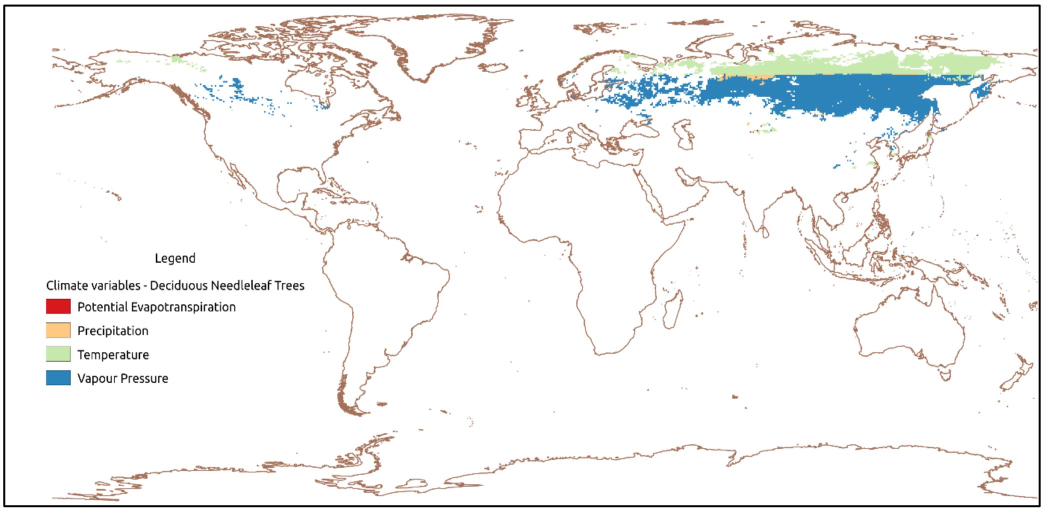



*Figure 16. Spatial distribution of the climate variables with the dominant influence on regional LAI for Deciduous Needleleaf Trees.*

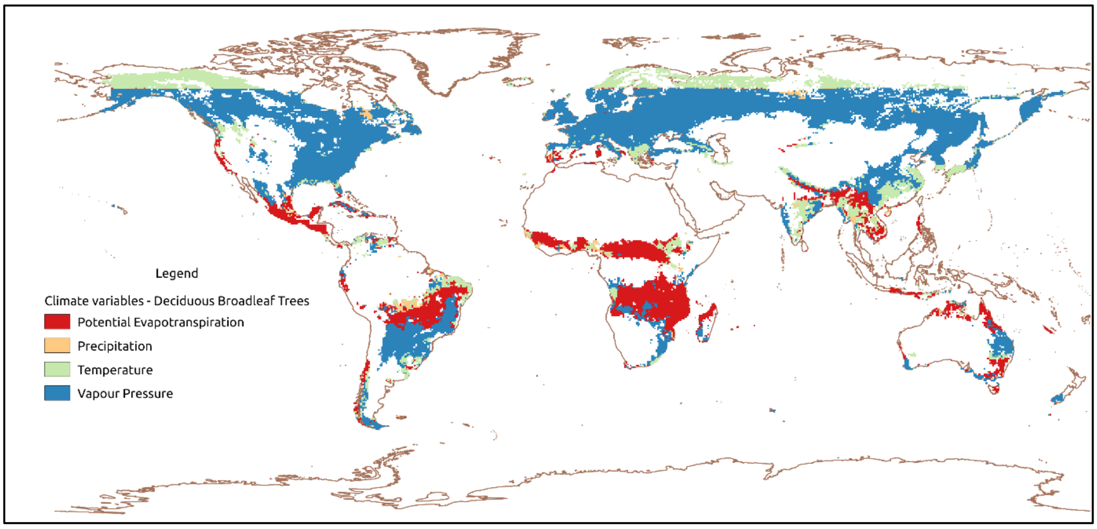



*Figure 17. Spatial distribution of the climate variables with the dominant influence on regional LAI for Deciduous Broadleaf Trees.*

The Trees land type covers occurs mostly at the northern hemisphere except for Evergreen Broadleaf Trees that appear at the equator. Similarity can be found between both types of deciduous trees and Evergreen Needleleaf Trees: the temperature determines LAI in further north and vapour pressure in the degrees below. The Broadleaf Trees that are at the equator have in common that LAI is mostly driven by potential evapotranspiration. Precipitation does not seem to influence LAI at the large scale.

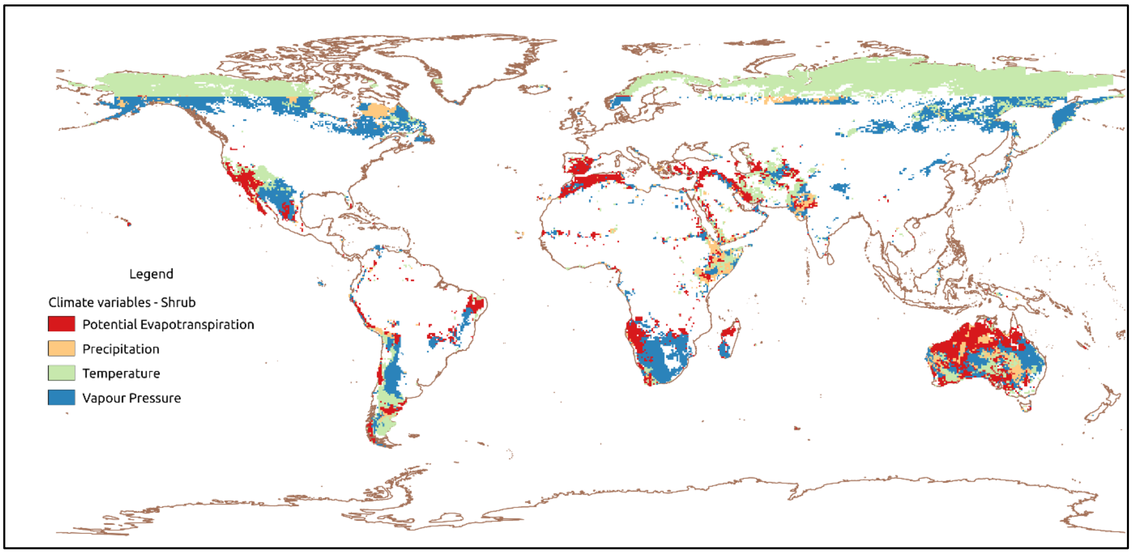



*Figure 18. Spatial distribution of the climate variables with the dominant influence on regional LAI for Shrub.*

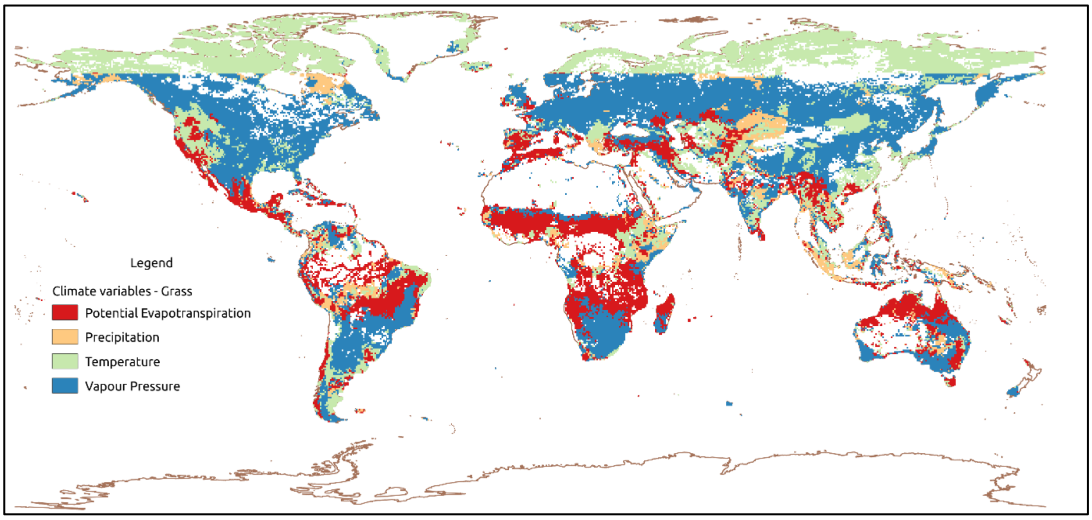



*Figure 19. Spatial distribution of the climate variables with the dominant influence on regional LAI for Grass.*

Where Grass appears all over the world, Shrubs are mostly in the northern hemisphere, Australia, south of Africa, and south of South America. Temperature and vapour pressure are dominant climate variables in the north. Interestingly, in Australia potential evapotranspiration and vapour pressure determines LAI. Precipitation is not important at the large scale, except for the Grass in the middle Asia, where this pattern might be connected with the higher altitudes of the region.

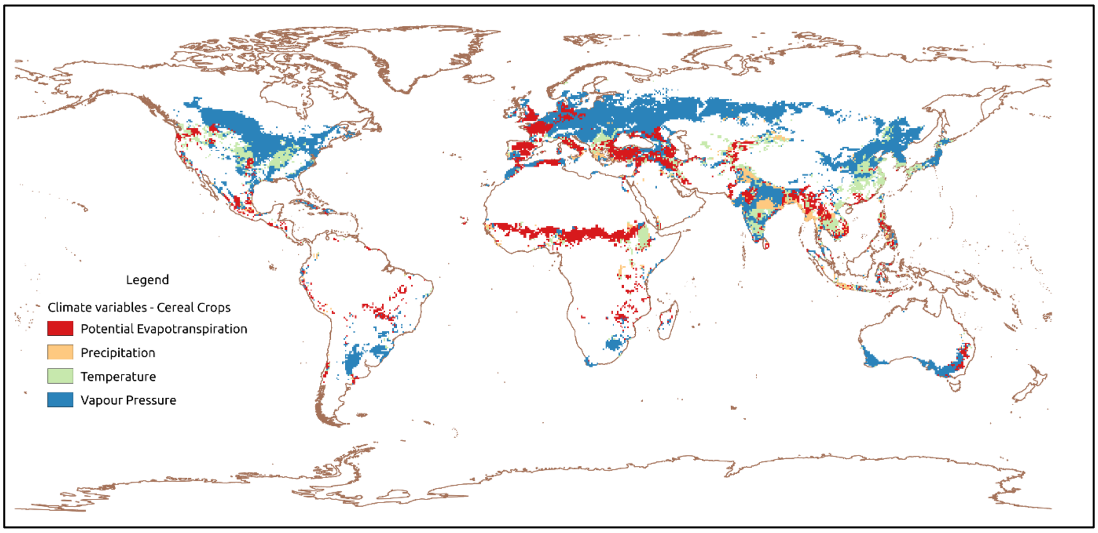



*Figure 20. Spatial distribution of the climate variables with the dominant influence on regional LAI for Cereal Crops.*

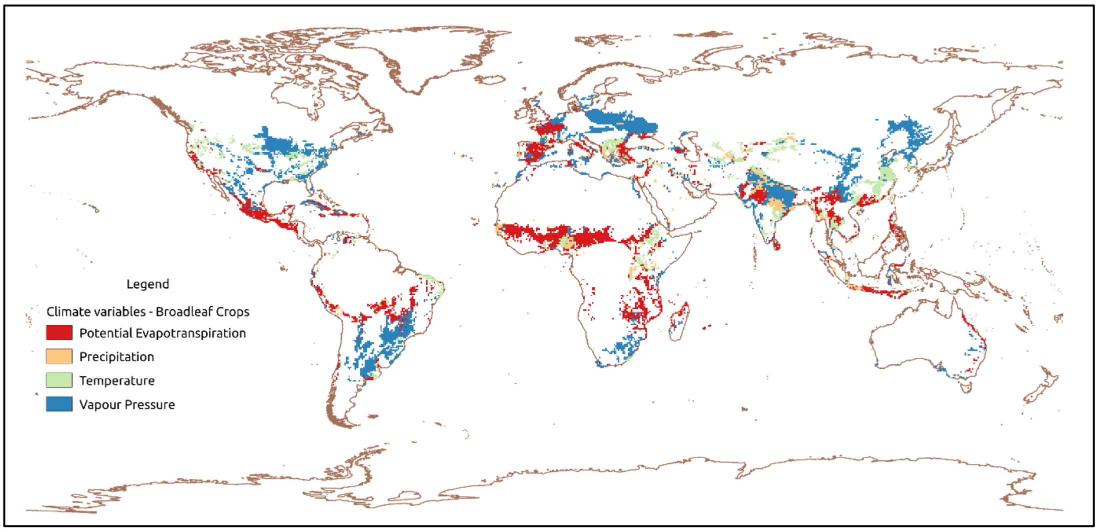



*Figure 21. Spatial distribution of the climate variables with the dominant influence on regional LAI for Broadleaf Crops.*

Crops does not appear at the large scale. The cereal crops covers middle Asia, and middle of North America. In the northern hemisphere the most dominant climate variable is vapour pressure. In the surroundings of the equator the LAI is mostly influenced by potential evapotranspiration.

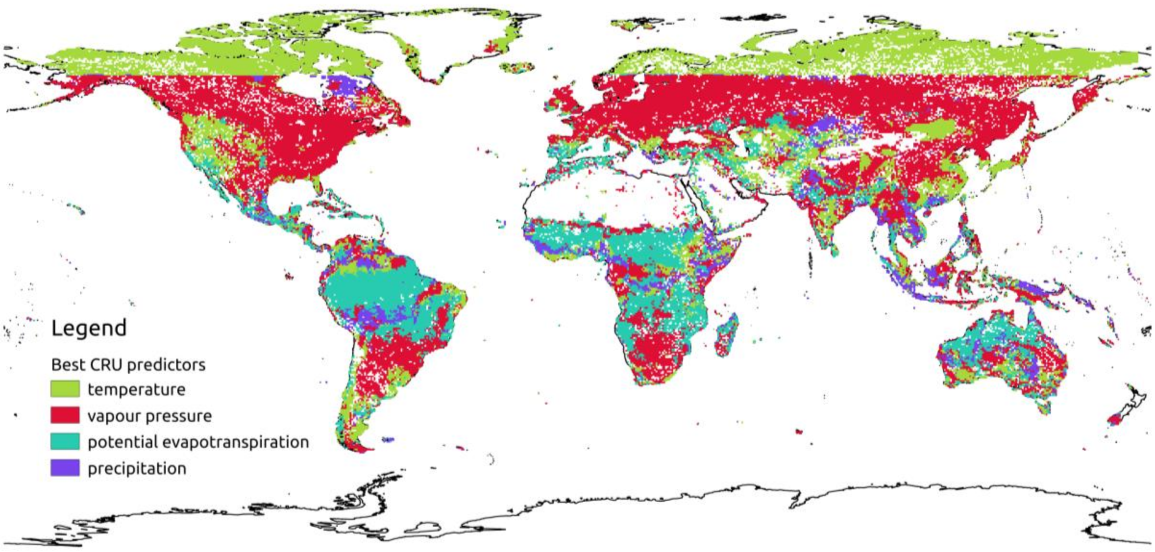



*Figure 22. Spatial distribution of the climate variables with the dominant influence on regional LAI for the plant function type that represents the highest mean LAI (Figure 11 & 12).*

Additionally, was made a map that represents the spatial distribution of the climate variables with the dominant influence on regional LAI for the plant functional type that represents the highest mean LAI (Figure 11 & 12). The temperature has a dominant influence on the north, vapour pressure at the lower degrees, and potential evapotranspiration in the surroundings of the equator. The precipitation is import in the south-eastern Asia.

## 3.3. Predictive model for the LAI with use of climate variables

Predictive models were based on the three formulas: ANPI, Jolly, and SFormula. The purpose of the formulas was to find a direct connections between climate variables and LAI. The formulas predicted LAI over the same 10-year period (2001-2010). The validation of the models was not performed due to underestimated time of prepossessing the datasets. The root-mean-square error (RMSE) and Akaike's Information Criterion (AIC) measured the performance at the local scale, where only RMSE was used to measure the performance of the models at the global scale.

### 3.3.1. Predictive model for five locations (Local scale)

Predictive model was made for two locations in the Temperate Zone and three locations in the Tropical Zone of Köppen's climate zone. The analyse was conducted to reveal the most dominant climate variable in the LAI predictions. For each climate variable of the formulas was determined the variance. The higher determined variance, the higher importance in the prediction of local LAI. The negative values of variances were negatively correlated, thus as smaller negative values, then the higher importance of the local LAI.

*Table 4. The performance of the models for the locations in the Temperate Zone.*

| **Köppen's climate zone**<br>Temperate Zone (30-60°) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Germany | | | | | | | | |
| **Formula** | **Temperature** | **Precipitation** | **Vapour Pressure** | **PET** | **Precipitation one month before** | **Precipitation two months before** | **RMSE** | **AIC** |
| Multivariate Analysis | 0.65 | -0.08 | 0.59 | -0.91 | - | - | 0.34 | 87.23 |
| Jolly | -0.57 | - | -1.16 | - | - | - | 0.36 | 99.34 |
| ANPI | 0.54 gdd | 0.20 | - | - | 0.00 | 0.09 | 0.39 | 118 |
| S Formula | 0.31/ 0.52 gdd | - | -4.3 | -1.00 | - | -0.02 | 0.33 | 86 |
| USA | | | | | | | | |
| **Formula** | **Temperature** | **Precipitation** | **Vapour Pressure** | **PET** | **Precipitation one month before** | **Precipitation two months before** | **RMSE** | **AIC** |
| Multivariate Analysis | 2.04 | -0.21 | 0.29 | -1.56 | - | - | 0.26 | 28.94 |
| Jolly | -0.22 | - | -0.97 | - | - | - | 0.37 | 106 |
| ANPI | 0.77 gdd | 0.19 | - | - | 0.08 | -0.20 | 0.38 | 117 |
| S Formula | 0.95/ 0.95 gdd | - | -0.32 | -1.44 | - | -0.06 | 0.27 | 34 |

*Table 5. The performance of the models for the locations in the Tropical Zone.*

| **Köppen's climate zone**<br>Tropical Zone (0-30˚) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Brazil | | | | | | | | |
| **Formula** | **Temperature** | **Precipitation** | **Vapour Pressure** | **PET** | **Precipitation one month before** | **Precipitation two months before** | **RMSE** | **AIC** |
| Multivariate Analysis | 0.10 | -0.05 | -0.19 | 0.60 | - | - | 0.29 | 50.95 |
| Jolly | 1.25 | - | 0.91 | - | - | - | 0.31 | 67 |
| ANPI | -0.07 gdd | -0.48 | - | - | -0.19 | 0.47 | 0.28 | 41 |
| S Formula | -0.01/-0.01 gdd | - | 0.11 | 0.53 | - | -0.28 | 0.28 | 43 |
| Mexico | | | | | | | | |
| **Formula** | **Temperature** | **Precipitation** | **Vapour Pressure** | **PET** | **Precipitation one month before** | **Precipitation two months before** | **RMSE** | **AIC** |
| Multivariate Analysis | -1.29 | 0.41 | 0.95 | -0.14 | - | - | 0.49 | 174.57 |
| Jolly | -1.77 | - | -1.46 | - | - | - | 0.49 | 173 |
| ANPI | -0.27 gdd | 0.38 | - | - | 0.50 | 1.46 | 0.32 | 76.61 |
| S Formula | -0.34/-0.34 | - | -0.63 | -0.10 | - | 1.64 | 0.32 | 77.14 |
| Malaysia | | | | | | | | |
| **Formula** | **Temperature** | **Precipitation** | **Vapour Pressure** | **Potential Evapotranspiration** | **Precipitation one month before** | **Precipitation two months before** | **RMSE** | **AIC** |
| Multivariate Analysis | 0.61 | -0.15 | -0.03 | -0.46 | - | - | 0.28 | 44.95 |
| Jolly | 0.48 | - | 0.12 | - | - | - | 0.31 | 69 |
| ANPI | 0.23 gdd | -0.02 | - | - | 0.22 | 0.43 | 0.28 | 50 |
| S Formula | 0.29/0.29 gdd | -- | 0.17 | -0.33 | - | -0.35 | 0.26 | 35 |

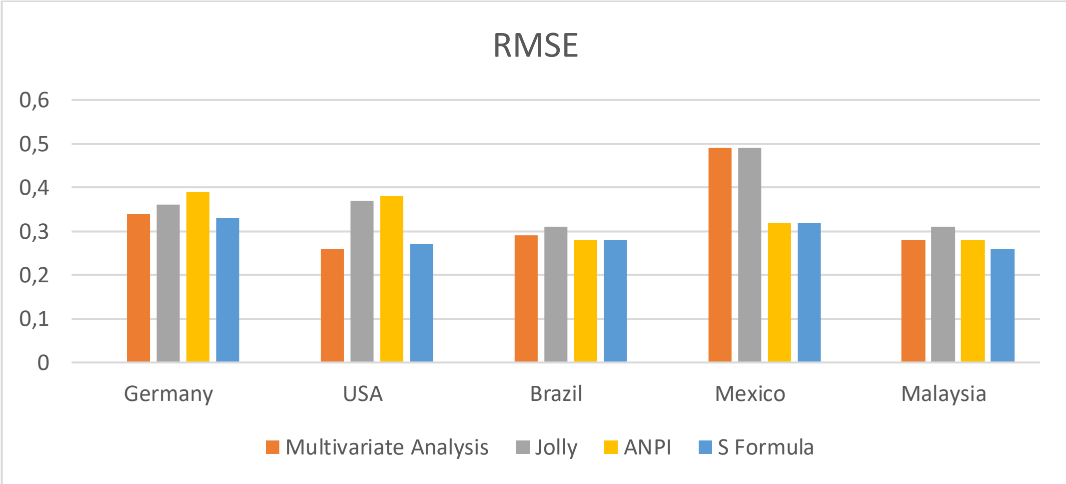



*Figure 23. Comparison of RMSE of the models.*

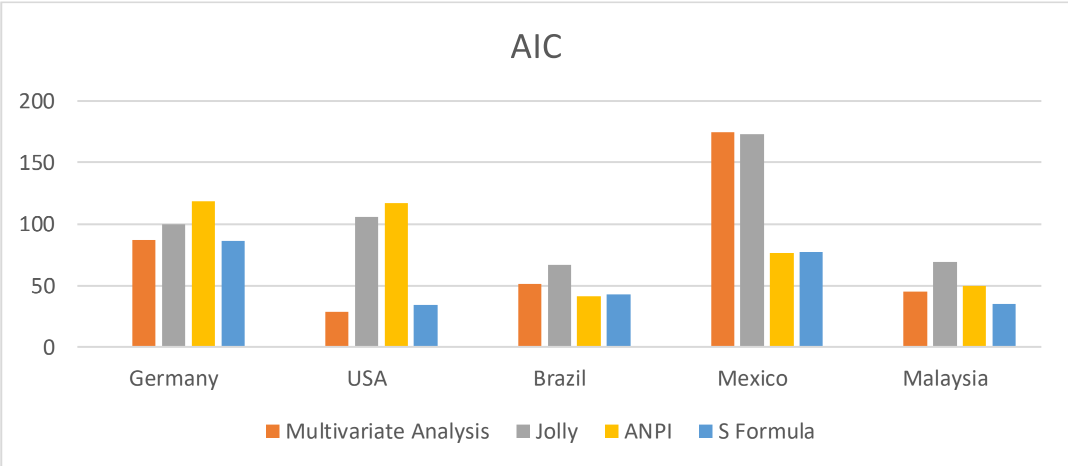



*Figure 24. Comparison of AIC of the models.*

The Temperate Zone has a strong connection and is highly correlated to the formulas that are based on the temperature, thus: Multivariate Analysis and SFormula. Location in Germany has slightly better prediction by using the SFormula. The LAI over the year is dependent on the climate variables, and there is observed more seasonality over the year. Location in USA (Arizona) has the best performance with Multivariate Analysis and it can be assumed that the vegetation is highly correlated to the temperature. It clearly shows that seasonality determines the vegetation grow, thus LAI. The ANPI formula, which depends mainly on the precipitation shows the worst prediction for both locations. That, together with parameters, reveals that precipitation is not high in importance for the temperate zone.

The LAI predictions show high dependency on the precipitation at the locations in Tropical Zone. The ANPI formula had the best performance for the locations in Brazil, and in Mexico, thus the vegetation might be highly dependent on the dry and wet season. Interestingly, Jolly Formula had the best performance for the locations at the Tropical Zone, however AIC classify it as the poorest model.

### 3.3.2. Predictive models of the LAI at the global scale

The predictive formulas were applied globally. For each $0.5^0$ by $0.5^0$ location was predicted LAI based on the all-natural vegetation covers and climate variables. Then the predicted LAI was compared to the historical LAI over the same period (2000-2010). The RMSE shows the average difference between the values for each location.

However, it has to be noticed that MODIS data contains a lot of errors (clouds, limitations of satellite measurements), which appears on the map in the 'patterned shape'.

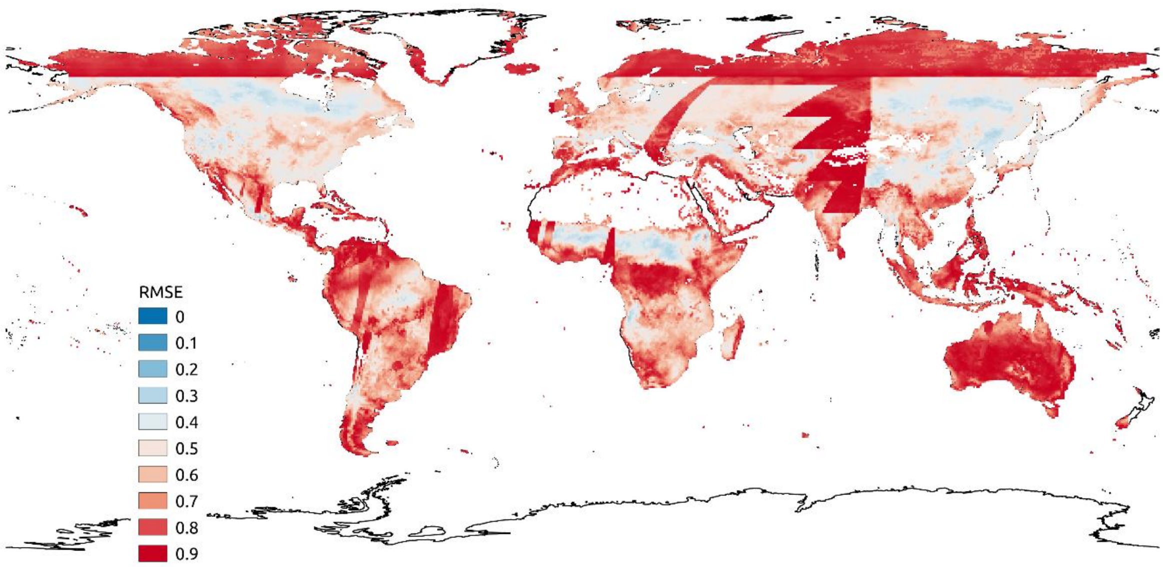



*Figure 25. The ANPI predictive model performance at the global scale.*

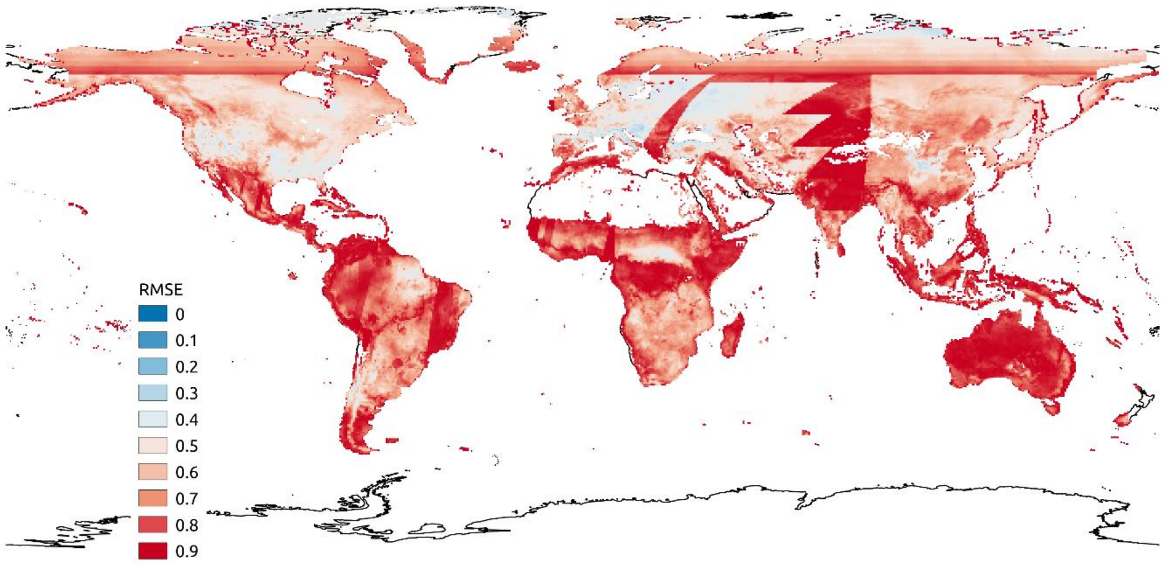



*Figure 26. The Jolly predictive model performance at the global scale.*

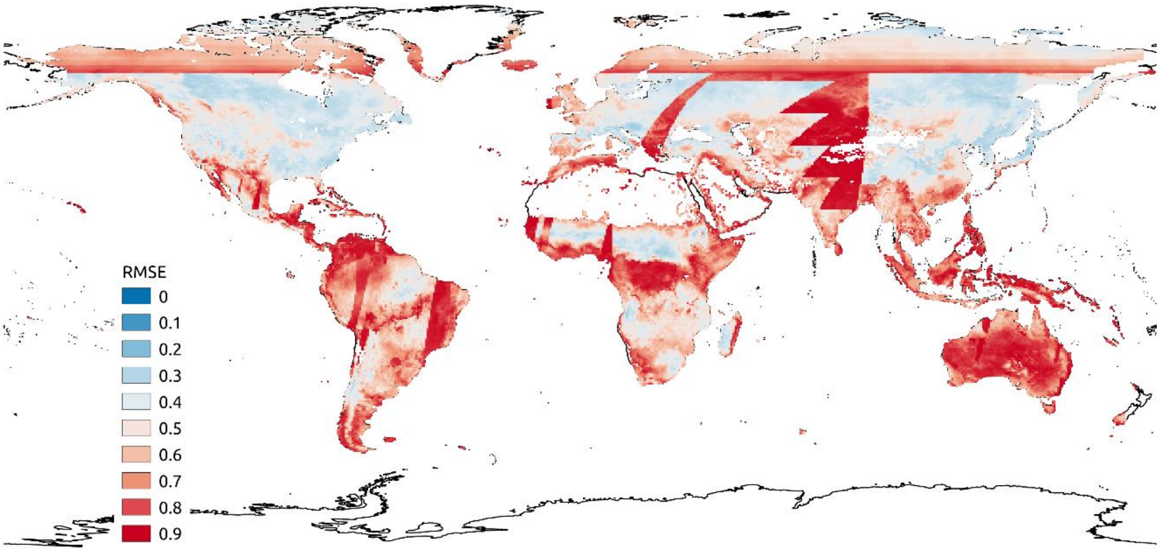


*Figure 27. The SFormula predictive model performance at the global scale.*

The Jolly and ANPI formulas have relatively high RMSE at the most of the globe locations. The Jolly model has better performance at the location where the temperature is a dominant climate variable for the LAI prediction (northern hemisphere). The SFormula predictive mode has the most similar values of predictive LIA in comparison to the historical LAI.

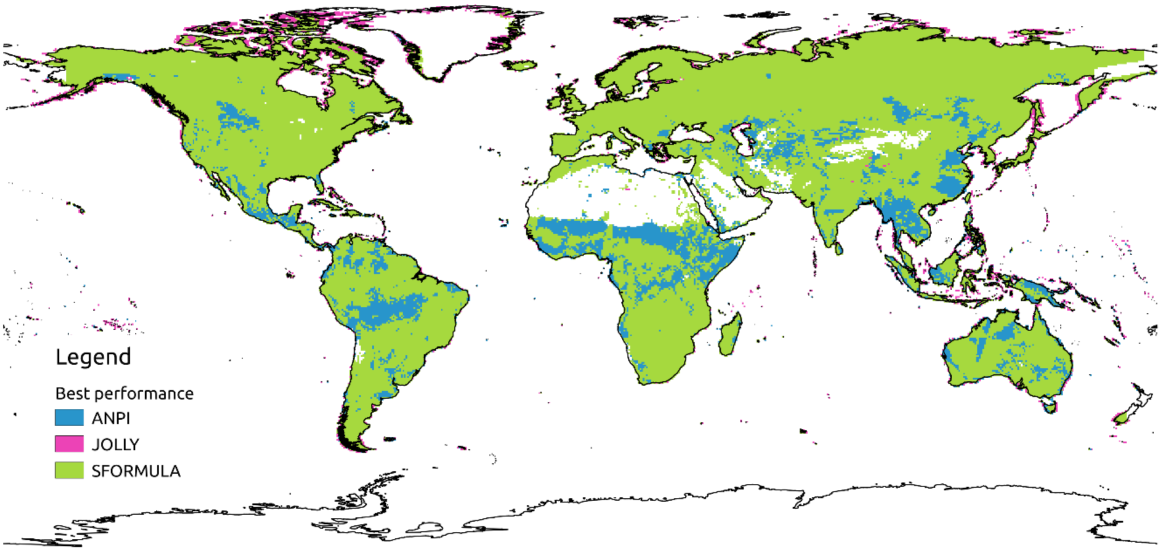


*Figure 28. The best performance for each location of the predictive models at the global scale.*

The comparison of all predictive models were made on the additional map. The map reveals which of the three predictive models had the best result. The map shows that SFormula is the most accurate predictive formula at the global scale. However, at the location of equator, the ANPI formula increases in importance. The same trend has been see in the middle Asia.

## 4. Discussion

The examination of the climate variables reveals that temperature, precipitation, vapour pressure, and potential evapotranspiration are the most influential factors for the LAI prediction at the local and global scale. The local analysis reveals high unexplained variance, which suggests search for the additional factors that could improve performance at the local scale (for instance radiation or altitude). The relationship between the variables and LAI is highly dependent on the climate zone. The local analysis shows better prediction where the climate variables are seasonal, thus Germany and USA. The local LAI at those locations is mostly dependent on the temperature. The local analysis of the Tropical Zone's locations reveals strong connection to the precipitation. The model performed worse predictions of lower degrees locations, in tropical and subtropical zones are more difficult to estimate because other factors are important or most likely the climate stays within reasonable boundaries for plant growth year around. Higher mean LAI values for Evergreen Broadleaf Trees are more difficult to predict from this four climate variables. This is proved by low correlation of predicted LAI to original LAI, and high RMSE, the reason might be that this four climate variables do not matter there as much as climate itself and another variables (solar radiation, soil type, humidity, altitude). The fluctuations are caused by cloudy measurements. The more away from the equator the more useful climatic predictions becomes. The climate changes more due to seasonal changes.

The spatial distribution of the climate variables and Land Use Type shows very clear pattern. The trees covers are influenced by temperature and vapour pressure in the Northern Hemisphere. This might be caused by harder availability of water and lower amount of the exposition to the sun. Moreover, all plant functional types shows very strong connections to the characteristic climate variables of the climate zones. Trees covers were less influent by climate variation then "lower vegetation" for instance grass and crops.

The predictive model, at local scale, revels that formulas that were focused on the temperature and potential evapotranspiration play important role in the local prediction of the LAI. However, above all the formulas it has been found that the SFormula, which uses 'climate memory of the previous months' has the best results in all locations. This is connected to the statement that vegetation is not dependent on the current meteorological conditions, but it is influenced by longer periods of climate data. In another words, the LAI predictions over specific area in September are dependent on the meteorological conditions of the previous months (for instance June-July-August).

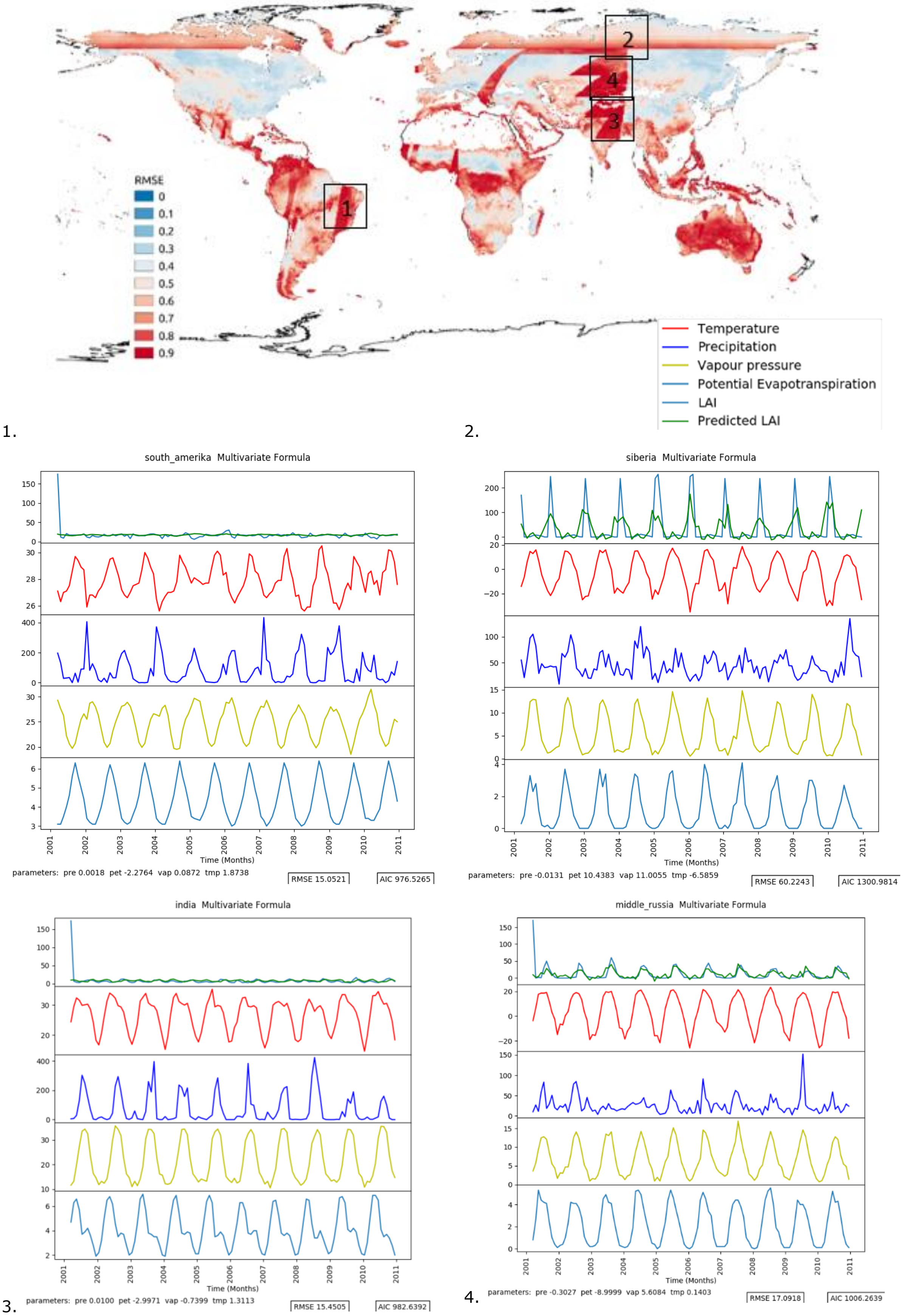



*Figure 29. Comparison of the time series of climate variables at the locations with errors in the measurements.*

It has be noticed, that most of the results at the global scale (spatial distribution of climate variables, as well as, predictive models), shows errors. The additional analysis was made at the locations were those measurements occurs (Figure 29). The multivariate analysis and time series of climate variables were studied to explain this 'unexpected pattern'. The location in South America, middle Russia, and India shows very high peak of the LAI in the year 2001 (at the beginning of the time series). The land use area at that location was not classified as a natural vegetation area (it was water, ice, or agriculture area). Moreover, the locations in the north Asia (Siberia) have a long period of winter and sun does not reach some of the areas. This might cause a errors in the calculations. Furthermore, the satellite data presents the LAI measurements as a raw dataset, thus it is suggested to use good filter to remove extreme LAI values (above 200).

## 5. Conclusion

The prediction of the local vegetation, as a Leaf Area Index (LAI), based on the climate variables is very complex. The literature and previous researches reveals that precipitation, temperaturel, potential evapotranspiration, and vapour pressure might be used for the determination of the dynamic relationship between LAI. Those variables are correlated and dependent.

The studies at local and global scal reveal that temperature and vapour pressure are the most dominant climate variables that determine LAI at the Temperate Zone of Köppen's Classification. This is due to the harder conditions of those places (less water availabilty). Those regions also have clear seasonality the climate variables, thus predicitons of LAI are more accurate then in the Tropical Zone of Köppen's Classification. In this zone the predictions of the LAI depends mostly on the potential evapotranspiration and precipitation. Moreover, the average LAI of this zone is higher and does not differ much over the year. Those characteristics make predictions of the local LAI more difficult and suggest that addition variable are also important (type of soil, radiation, altitude).

The spatial distribution of the most dominant climate variable for each plant functional type reveals that those variables are more dependent on the climate zone than on the plant functional type. However, 'higher vegetation' for instance Trees are less sensitive to meteorological data then 'lower vegetation' for instance scrubs.

The predictive models of the LAI were based on four formulas. The best performance had a SFormula that had a 'meteorological memory'. The vegetation is highly sensitive to the wet and dry periods, thus cannot be only modelled based on the current meteorological conditions. The most general dynamic formula for the global LAI predictions should be highly dependent on the potential evapotranspiration, which is highly correlated to the precipitation and temperature. Furthermore, the additional variables are very important in the prediction of the regional LAI ( for instance altitude, radiation).

## 6. Outlook and Recommendations

In this research several subjects needs an additional explanation and recommendations. First of all, the model was supposed to be apply in the global hydrological model PCR-GLOBWB. However, the additional pre-processing of the satellite data (MODIS) consumed enormous amount of the data not included at the beginning of the research. The satellite data of LAI (MODIS15AH) is obtained as a raw grid of a certain area. It has to be noticed that those grids are different sizes, thus a special script had to written to match this dataset with the meteorological dataset.

Furthermore, some additional comments and recommendations has to be added to the model. The (predictive) model was made based on the 10-year data period of climate variables, which however is rather brief time for the LAI prediction. The model could be better trained by using longer period of datasets (for instance 50 years). Moreover, the satellite data (MODIS/NASA) itself is very sensitive due to the extreme values, which are caused by obstacles like clouds or errors in the measurements. This suggests that more focus should be put on the filtering the datasets. The LAI dataset v006 of the entire world over a 10-year period is around 512Gb, thus the size of the datasets have to counted in the similar future research. Moreover, Land Use dataset can be used for every year over the same period as LAI dataset. Furthermore, the meteorological dataset is very low compare for the LAI and this makes rather impossible to make an accurate model. In a small area the climate can still change a lot due to the altitude or local precipitation, thus it is recommended to find a higher resolution of the meteorological dataset.

Last but not least development of the predictive model of the vegetation based on the climate (meteorological data) requires finding a compromise between the simplification and accuracy.

## References

Baret, F., Guyot, G., 1991, Potentials and limits of vegetation indices for LAI and APAR assessment. Volume 35, Issues 2–3, 161-173

Betts, R. A., Cox, P. M., Lee, S. E., Woodward, I. F., 1997, Contrasting physiological and structural vegetation feedbacks in climate change simulations. Nature volume 387, 796–799.

Chen, J. J., Per T., Masayuki G., Zhihui M., Bunkei R., 2004, A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sensing of Environment. 91. 332-344.

Jolly, W. M., Nemani, R., Running, S. W., 2005, Including the dynamic relationship between climatic variables and leaf area index in a hydrological model to improve streamflow prediction under a changing climate. Hydrol. Earth Syst. Sci., 19, 2821–2836.

Mohammad A., 2013. Climate Change Impacts on Plant Biomass Growth, North South University, Dhaka, Bangladesh.

Neilson, R.P., Marks, D., 1994. A global perspective of regional vegetation and hydrologic sensitivities and risks from climatic change. J. Veget. Sci. 5, 715–730.

Sala, O. E., Chapin, F.S., Armesto, J.J., 2000. Global biodiversity scenarios for the year 2100. Science. 287 (5459): pp. 1770–4.

Savitzky, A., Golay, M. J. E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36, pp. 1627–1639

Splitt M., Horel J., 2002, Use of Multivariate Linear Regression for meteorological data analysis and quality assessment in complex terrain. Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

Stephenson, N. L., Climatic Control of Vegetation Distribution: The Role of the Water Balance, 1990, The American Naturalist 135, no. 5, 649-670.

Rind, D., Hansen, J. E., Takahashi, T., 2013. The Influence of Vegetation on the Hydrologic Cycle in a Global Climate Model. Climate Processes and Climate Sensitivity, 1

Tang H., Brolly, M., Feng, Z., Strahler A. H., Schaaf C. L. Ganguly, S., Zhang G., Dubayah R., 2014, Deriving and validating Leaf Area Index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA. Remote Sensing of Environment, Volume 143, 131-141.

Tesemma Z. K., Y. Wei, M. C. Peel, and A. W. Western, 2015, Including the dynamic relationship between climatic variables and leaf area index in a hydrological model to improve streamflow prediction under a changing climate, Hydrol. Earth Syst. Sci., 19,

Van Beek, L.P.H., M.F.P. Bierkens, 2008, The Global Hydrological Model PCRGLOBWB: Conceptualization, Parameterization and Verification, Report. Department of Physical Geography, Utrecht University, Utrecht, The Netherlands. http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf

Watson, D.J., 1947, Comparative physiological studies on the growth of field crops: I. Variation in net assimilation rate and leaf area between species and varieties and within and between years. Annals of Botany. 11: 41–76.

Website:

www.lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd15a2h_v006 available: 22/04/2018

https://modis-land.gsfc.nasa.gov/MODLAND_grid.html available: 08/06/2018

## Appendix A: Parameters for climate variables in multivariate analysis for five locations

| Climate variable / Location | Brazil | | Mexico | | Malaysia | | Germany | | USA | |
|---|---|---|---|---|---|---|---|---|---|---|
| Units | [-] | % | [-] | [-] | % | [-] | % | % | [-] | % |
| Temperature | 0.03 | 0.02 | 1.08 | 0.44 | 0.352 | 1.03 | 0.61 | 0.20 | 0.66 | 0.35 |
| Precipitation | 0.48 | 0.29 | 0.19 | 0.05 | 0.04 | 0.04 | 0.02 | 0.03 | 0.25 | 0.13 |
| Potential evapotranspiration | 0.33 | 0.20 | 0.09 | 0.1 | 0.08 | 0.21 | 0.12 | 0.02 | 0.07 | 0.04 |
| Vapour pressure | 0.04 | 0.02 | 1.23 | 0.34 | 0.272 | 0.06 | 0.03 | 0.22 | 0.04 | 0.02 |
| Unexplained variance | 0.77 | 0.47 | 2.93 | 0.32 | 0.256 | 0.32 | 0.19 | 0.53 | 0.86 | 0.46 |
| **SUM** | **1.65** | **1** | **5.52** | **1.25** | **1** | **1.66** | **1** | **1** | **1.88** | **1** |

**Appendix B: Correlation coefficient between climate variables for five locations**

GERMANY

| Climate variable | Temperature | Vapour pressure | Precipitation | Potential evapotranspiration |
|---|---|---|---|---|
| Temperature | 1 | 0.74 | 0.05 | 0.59 |
| Vapour pressure | 0.74 | 1 | 0.44 | 0.04 |
| Precipitation | 0.05 | 0.44 | 1 | -0.50 |
| Potential evapotranspiration | 0.59 | 0.04 | -0.50 | 1 |

BRAZIL

| Climate variable | Temperature | Vapour pressure | Precipitation | Potential evapotranspiration |
|---|---|---|---|---|
| Temperature | 1 | 0.87 | 0.37 | 0.57 |
| Vapour pressure | 0.87 | 1 | 0.66 | 0.16 |
| Precipitation | 0.37 | 0.66 | 1 | -0.31 |
| Potential evapotranspiration | 0.57 | 0.16 | -0.31 | 1 |

MEXICO

| Climate variable | Temperature | Vapour pressure | Precipitation | Potential evapotranspiration |
|---|---|---|---|---|
| Temperature | 1 | 0.96 | 0.06 | 0.90 |
| Vapour pressure | 0.96 | 1 | 0.16 | 0.81 |
| Precipitation | 0.06 | 0.16 | 1 | -0.06 |
| Potential evapotranspiration | 0.90 | 0.81 | -0.06 | 1 |

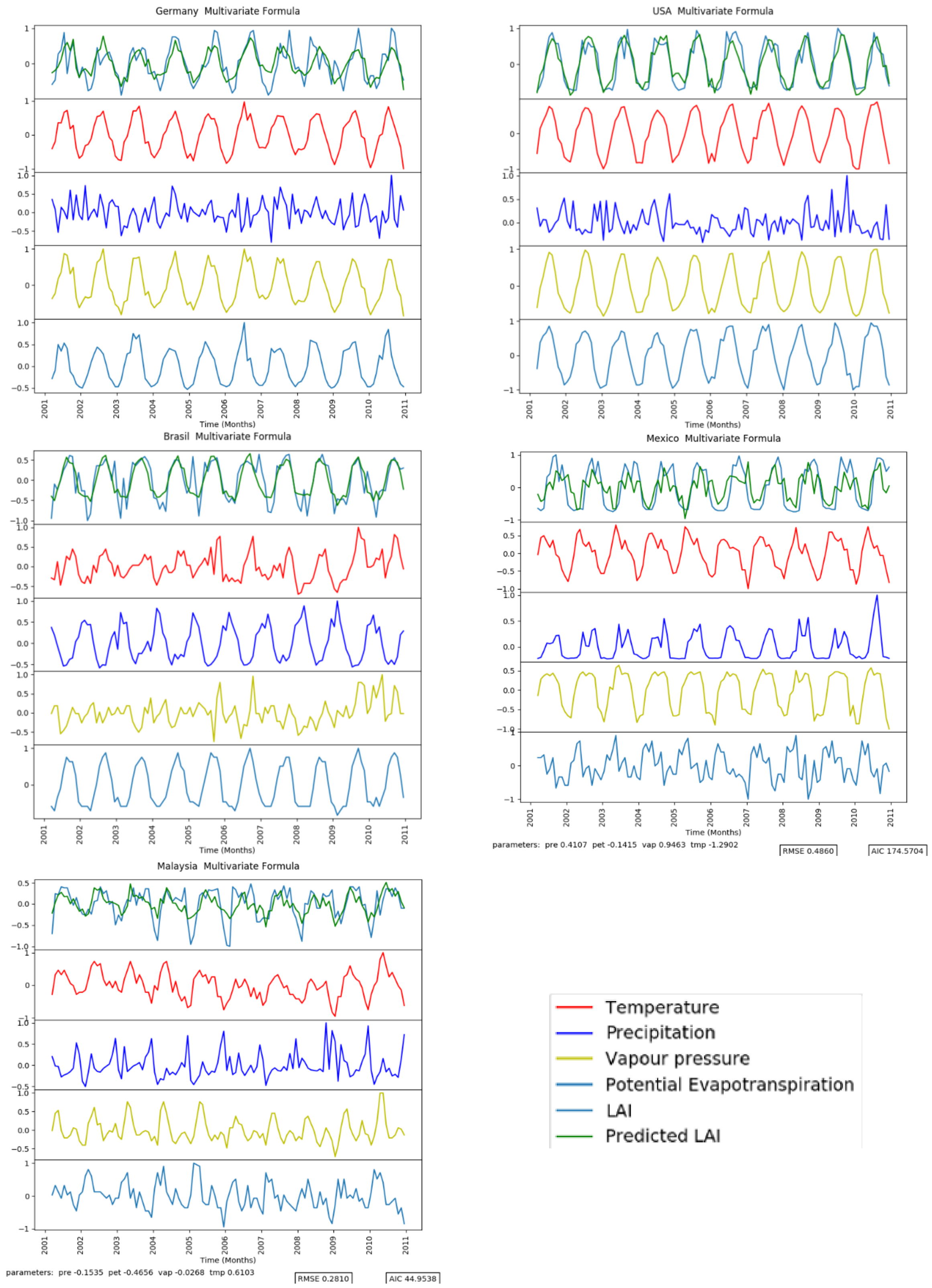
USA

| Climate variable | Temperature | Vapour pressure | Precipitation | Potential evapotranspiration |
|---|---|---|---|---|
| Temperature | 1 | 0.97 | -0.19 | 0.95 |
| Vapour pressure | 0.97 | 1 | -0.17 | 0.91 |
| Precipitation | -0.19 | -0.17 | 1 | -0.28 |
| Potential evapotranspiration | 0.95 | 0.91 | -0.28 | 1 |

MALAYSIA

| Climate variable | Temperature | Vapour pressure | Precipitation | Potential evapotranspiration |
|---|---|---|---|---|
| Temperature | 1 | 0.64 | -0.41 | 0.46 |
| Vapour pressure | 0.64 | 1 | -0.01 | 0.33 |
| Precipitation | -0.41 | -0.01 | 1 | -0.52 |
| Potential evapotranspiration | 0.46 | 0.33 | -0.52 | 1 |

## Appendix C: Time series of climatic variables in multivariate analysis for five locations



## Appendix D: Multivariate analysis in a python script

*The whole script can be found on the github.com annateuerle named LAI thesis*

```python
"""
- Predict LAI based on the climatic variables.

- Fit a line, y = ax + by + cz + dq + constant
  for each cru lon, lat location on a modis map

- Plot the end result on a map
"""

import argparse
import matplotlib.pyplot as plt
import matplotlib as mpl
from mpl_toolkits.basemap import Basemap
import logging
import numpy as np
import os

import math
import modis_map
import read_modis
from datetime import datetime

import settings
from settings import conf
import extract_CRU
import create_lai_cube
import plot_predictors

from plot_map_progress import plot_lai

log = logging.getLogger(__name__)
log.setLevel(logging.DEBUG)
logging.basicConfig(level=logging.DEBUG)


CRU_IDX = ('tmp', 'vap', 'pet', 'pre')

OPTIONS = {
    'debug': False
}


def solver_function_multi(
        lcru, llai, timestamps, predictors=('tmp', 'vap', 'pet', 'pre'),
        label='all', showplot=False):
    """
    Fit a line, y = ax + by + cz + dq + constant,
    through some noisy data-points

    :param lcru:  cru at location (raw , option to still do calculations)
    :param llai:  lai at location (normalized)
    :param timestamps datetimes of time period. used to make graph.
    :param predictors  the dataset names we use to predict.
```

```
:param label  we store this predicted lai under pred_{label}
:return:  best symbol values and rmse for prediction function in settings.
"""
measurements = []
plot_predictor_labels = []

for ds_key in predictors:
    if type(ds_key) is str:
        input_ar = lcru[:, CRU_IDX.index(ds_key)]
        plot_predictor_labels.append(ds_key)
    else:
        input_ar = ds_key(lcru)
        plot_predictor_labels.append(ds_key.__name__)

    input_ar = normalize(input_ar)
    measurements.append(input_ar)

y = llai

measurements.append(np.ones(len(y)))

# We can rewrite the line equation as y = Ap,
# where A = [[x 1]] and p = [[m], [c]].
# Now use lstsq to solve for p:
A = np.vstack(measurements).T  # [[x  y z q 1]]

try:
    parameters = np.linalg.lstsq(A, y, rcond=None)[0]
except ValueError:
    # log.error('missing cru?')
    return

predictor_params = "parameters: "
for l, p in zip(plot_predictor_labels, parameters):
    log.debug(l)
    log.debug(p)
    predictor_params += ' %s %.4f ' % (l, p)

m = measurements
y_pred = np.zeros(120)

# for i, p in enumerate(parameters[:-1]):   # we skip K.
for i, p in enumerate(parameters):
    # log.debug('p %s', p)
    y_pred += p * m[i]

for v in y_pred:
    print(v)

rmse = calc_rmse(y, y_pred)
log.info('%s RMSE: %s', label, calc_rmse(y, y_pred))

if not showplot:
    return label, rmse

# datasets[f'pred_{label}'] = y_pred
plot_predictors.plot(
```

```
        timestamps, y, y_pred,
        measurements, predictors=plot_predictor_labels, p_label=label,
        text=predictor_params)

    calculate_ss(y, y_pred, measurements, plot_predictor_labels, parameters)


def calculate_ss(y, y_pred, measurements, plot_predictor_labels, parameters):
    """Calculate SSerr SStot, SSreg, R2,

    not means are always zero since we have standardized the data.
    """
    log.info(conf['groupname'])
    log.info(plot_predictor_labels)
    log.info(parameters)
    ss_err = np.power(y - y_pred, 2).sum() / 120
    ss_tot = np.power(y, 2).sum() / 120
    ss_tot_p = np.power(y_pred, 2).sum() / 120

    log.info('ss_err %.3f', ss_err)
    log.info('ss_tot %.3f', ss_tot)
    log.info('ss_tot_p %.3f', ss_tot_p)

    sum_r = 0
    for m, l, p in zip(measurements, plot_predictor_labels, parameters):

        # mss_tot = np.power(m, 2).sum() / 120
        mss_reg = np.power(m * p, 2).sum() / 120
        # mss_reg = np.power(mp, 2).sum() / 120
        log.info('%s %.2f ss_reg %.3f', l, p, mss_reg)
        sum_r += mss_reg

    log.info('sum regression %s', sum_r)
    log.info('sum regression + err %.2f', sum_r + ss_err)
    # fraction of vaiance unexplained.
    fvu = ss_err / ss_tot

    log.info('fvu %.3f', fvu)
    log.info('R2 %.3f', 1 - fvu)


def make_local_plot(grid, lai, cru, timestamps, geotransform, projection):
    """Plot LAI, predicted LAI and predictors (temp, vap, pre, pet)
    of lon lat location defined in settings for current group
    """
    group = conf['groupname']
    lon = settings.locations[group]['lon']
    lat = settings.locations[group]['lat']
    x, y = read_modis.determine_xy(geotransform, projection, lon, lat)

    log.debug('%s %s', x, y)
    lai_at_location = lai[:, int(y), int(x)]
    lai_at_location_org = np.copy(lai_at_location)
    lai_at_location = normalize(lai_at_location)

    log.info('MEAN_LAI %d', lai_at_location_org.mean())
    log.info('MEAN_LAI N %d', lai_at_location.mean())
```

```python
    log.info('STD LAI STD %d', lai_at_location_org.std())
    log.info('STD LAI STD N %d', lai_at_location.std())
    for v in lai_at_location:
        print(v)
    log.info('---' * 15)

    # log.debug(lai_at_location)
    # find closest cru data
    min_i = -1
    min_d = 999999999999

    for i, (glon, glat) in enumerate(grid):
        distance = (lon-glon)**2 + (lat-glat)**2
        if distance < min_d:
            min_d = distance
            min_i = i
            # log.debug(f'Distance**2 = {min_d} {glon}, {glat} -> {lon} {lat}')

    cru_at_location = cru[min_i, :, :]

    for label, predictors in MODEL_OPTIONS.items():
        solver_function_multi(
            cru_at_location, lai_at_location, timestamps,
            predictors=predictors,
            label=f'{group}-{label}', showplot=True)


def normalize(arr):
    """standardize arr with average around 0
    """
    mean = np.mean
    std = np.std
    normalized_data = (arr - mean(arr, axis=0)) / std(arr, axis=0)
    # normalized_data = normalized_data / abs(normalized_data).max()
    return normalized_data


def calc_rmse(predictions, targets):
    differences = predictions - targets      # the DIFFERENCES.
    differences_squared = differences ** 2   # the SQUARES of ^
    mean_of_differences_squared = differences_squared.mean()  # the MEAN of ^
    rmse_val = np.sqrt(mean_of_differences_squared)       # ROOT of ^
    if np.isnan(rmse_val):
        log.error('P %s', predictions)
        log.error('T %s', targets)
    return rmse_val


def valid_box(box):
    """Check if box is valid
    """

    if not box:
        return False

    if any([c is None for c in box]):
        return False
```

```
    if len(box) != 4:
        return False

    # if box[1][1] == box[2][1]:
    #     log.debug('same y')
    #     return False

    # if box[0][0] == box[3][0]:
    #     log.debug('same x')
    #     return False

    return True


def extract_grid_data(box, lai, green_m, grid, debug=False):
    """For grid location extract relavant data
    """

    l, r, d, u = box

    log.info('%d:%d %d:%d', u[1], d[1], l[0], r[0])
    cube_lai_at_location = lai[:, u[1]:d[1], l[0]:r[0]]
    # p4grid.extend([pr, pl, pu, pd])
    g_at_loc = green_m[u[1]:d[1], l[0]:r[0]]

    def plot(data):
        valid_px = grid_to_pixels(grid)
        # plot the cru pixel we are working on
        for j, (x, y) in enumerate(valid_px):
            plt.plot(x, y, 'r+')

        plt.imshow(data)
        geotransform, projection, bbox = create_lai_cube.extract_lai_meta()
        plt.show()

    if debug:
        # debug indexing.
        # plot what we are doing.
        # and check if sliceing is going ok
        dlai = np.copy(lai[20, :, :])
        dlai[u[1]:d[1], l[0]:r[0]] = 30
        plot(dlai)

        green_c = np.copy(green_m)
        green_c[u[1]:d[1], l[0]:r[0]] = 30
        plot(green_c)

        plt.imshow(g_at_loc)
        # plt.colorbar()
        plt.show()
        # green_mask = np.logical_and(g_at_loc > 0
        plt.imshow(g_at_loc)
        # plt.colorbar()
        plt.show()

    return cube_lai_at_location, g_at_loc
```

```python
def _rmse_one_location(
        models, grid, i, g, box,
        cru, lai, green_m, timestamps,
        grid_model_rmse, invalid):
    """
    Find lowest rmse for models of one cru grid location in map/dataset
    """

    cru_at_location = cru[i, :, :]

    cube_lai_at_location, green_mask = extract_grid_data(
        box, lai, green_m, grid, debug=OPTIONS['debug'])

    # validate green mask
    if not np.any(green_mask):
        invalid.append((g, 'nogreen'))
        return

    # we want at least more then 10km²
    if np.sum(green_mask) < 10:
        invalid.append((g, 'nogreen'))
        return

    # create a 3d / cube green mask
    green_mask3d = np.zeros(cube_lai_at_location.shape, dtype=bool)
    green_mask3d[:, :, :] = green_mask[np.newaxis, :, :]
    # set ALL LAI values not in green mask are ZERO
    cube_lai_at_location[~green_mask3d] = 0
    assert cube_lai_at_location.shape == green_mask3d.shape
    # Sum lai values in each layer (month) to array of 120 values
    sum_colums_lai = cube_lai_at_location.sum(axis=1)
    sum_array_lai = sum_colums_lai.sum(axis=1)

    # we have summed up lai values of 120 months
    # assert sum_array_lai.size == 120

    # print(sum_array_lai)
    # normalize input data

    # cru_at_location = normalize(cru_at_location)
    avg_lai_at_location = normalize(sum_array_lai)

    # print(avg_lai_at_location)
    if np.isnan(avg_lai_at_location).any():
        log.error('%s %s', g, sum_array_lai)
        invalid.append((g, 'nolai'))
        return

    for label, p_labels in models.items():
        # label = '%s %s' % (label, g)
        answer = solver_function_multi(
            cru_at_location, avg_lai_at_location,
            timestamps, predictors=p_labels, label=label)

        if not answer:
```

```
            log.error(answer)
            invalid.append((g, 'normse'))
            return

        m, rmse = answer

        grid_model_rmse[g].append((rmse, m))


def calculate_models_for_grid(
        models, cru, lai, green, timestamps, grid, boxes):
    """
    For location in box. which is a cru location, find cru data,

    Parameters

    models:      Dict of models evaluate for each location
    cru:         925*120*4 of cru information
    lai:         1200*1200*120 cube of lai information
    green:       1km² locations of green
    timestamp:   10 years in months (120)
    param grid:  all cru lon, lat of current lai data location of world
    param boxes: 4 points in pixel location, up, down, left, right.
    """
    invalid = []
    grid_model_rmse = {}
    green_m = np.logical_and(green > 0, green < 5)

    for i, g in enumerate(grid):
        # add default value.
        g = tuple(g)
        # values should be between 0, 1, 2 will be masked
        grid_model_rmse[g] = [(2, '')]
        box = boxes[i*4:i*4+4]
        # print(box)

        if not valid_box(box):
            invalid.append((g, 'bbox'))
            continue

        log.error('%d %d %s', i, len(grid), box)

        log.debug('OK %s', box)

        _rmse_one_location(
            models, grid, i, g, box, cru,
            lai, green_m, timestamps,
            grid_model_rmse, invalid)

    return grid_model_rmse, invalid


def aic_criterion(models_to_make, datasets):
    # load hdf5 measurement data.
    lai = datasets['lai']
    for p, ds_label in models_to_make.items():
        p_label = f'pred_{p}'
```

```python
        predicted_lai = datasets[p_label]
        R = np.square(lai - predicted_lai).sum()
        # print(R)
        m = len(ds_label)  # len variables
        n = len(lai)       # measurements
        A = n * math.log((2*math.pi)/n) + n + 2 + n * math.log(R) + 2 * m
        print('%s %.4f' % (p, A))


def tmp_gdd(lcru):
    """
    temperature below 5.
    """
    tmp = np.copy(lcru[:, CRU_IDX.index('tmp')])
    tmp[tmp < 5] = 0
    return tmp


def tmp_one(lcru):
    tmp = np.copy(lcru[:, CRU_IDX.index('tmp')])
    tmpone = np.roll(tmp, 1)
    return tmpone


def pet_one(lcru):
    tmp = np.copy(lcru[:, CRU_IDX.index('pet')])
    tmpone = np.roll(tmp, 1)
    return tmpone


def vap_one(lcru):
    tmp = np.copy(lcru[:, CRU_IDX.index('vap')])
    tmpone = np.roll(tmp, 1)
    return tmpone


def pre_one(lcru):
    pre = np.copy(lcru[:, CRU_IDX.index('pre')])
    preone = np.roll(pre, 1)
    return preone


def find_green_location_mask(green):
    g = green
    m = np.logical_and(g > 0, g < 5)
    return m


def make_box_grid(grid):
    """
    For each point find 4 neighboring points to calculate ~exact area to
    take green points from.
    """
    # for grid item find 4 neigbouring points in middle.
    p4grid = []

    for (lon, lat) in grid:
```

```python
        pr = (lon + 0.15, lat)
        pl = (lon - 0.15, lat)
        pu = (lon, lat + 0.20)
        pd = (lon, lat - 0.20)
        p4grid.extend([pl, pr, pd, pu])
        # log.debug(f'{lon}:{lat}: {pr}{pl}{pu}{pd}')

    return p4grid


def make_basemap(extent):
    # create map using BASEMAP
    m = Basemap(
        llcrnrlon=extent[0], llcrnrlat=extent[3],
        urcrnrlon=extent[1], urcrnrlat=extent[2],
        # lat_0=(lat_max - lat_min) / 2,
        # lon_0=(lon_max - lon_min) / 2,
        projection='cyl',
        # projection='cyl',
        resolution='h',
        # area_thresh=10000.,
    )

    m.drawcoastlines(linewidth=0.5)
    m.drawcountries(linewidth=0.5)
    parallels = np.arange(49., 79., .5)
    # labels = [left,right,top,bottom]
    m.drawparallels(parallels, labels=[False, True, True, False])
    meridians = np.arange(0., 49., .5)
    m.drawmeridians(meridians, labels=[False, True, True, False])

    plt.tight_layout()

    return m


def plot_errors(m, invalid):
    """Plot data errors on map
    """

    for (lon, lat), reason in invalid:
        if reason == 'bbox':
            m.scatter(lon, lat, c='yellow', latlon=True, marker='x')
        elif reason == 'nogreen':
            m.scatter(lon, lat, c='green', latlon=True, marker='8')
        elif reason == 'nolai':
            m.scatter(lon, lat, c='blue', latlon=True, marker='v')
        elif reason == 'normse':
            m.scatter(lon, lat, c='red', latlon=True, marker='>')
        else:
            raise ValueError('unknown reason')


def _plot_and_save(m, cmap, data_g, title):
    """Plot world map and save output to a location
    """
    data_g = data_g.ReadAsArray()
```

```
    d = np.flipud(data_g)
    m.imshow(d, vmin=1, vmax=5, cmap=cmap)
    plt.title(title)
    manager = plt.get_current_fig_manager()
    manager.resize(*manager.window.maxsize())
    fig1 = plt.gcf()
    plt.show()
    d = datetime.now()
    date = f'{d.year}-{d.month}-{d.day}'
    imgtarget = os.path.join('imgs', conf['groupname'], f'{date}-{title}.png')
    fig1.savefig(imgtarget)


def plot_models(models, extent, lons, lats, data_g, invalid, title):

    m = make_basemap(extent)

    model_keys = list(MODEL_OPTIONS.keys())
    model_keys.sort()

    data = []

    for mdl in models:
        if mdl in model_keys:
            data.append(model_keys.index(mdl))
        else:
            data.append(-1)

    models_data = np.array(data)
    mdata = models_data.reshape(len(lons), len(lats))

    cmap = plt.get_cmap('RdBu', 4)
    cmap.set_under('0.8')
    cmap.set_bad('0.8')
    cmap.set_over('0.8')

    norm = mpl.colors.Normalize(vmin=0, vmax=np.max(mdata), clip=True)

    valid = np.ma.masked_where(mdata == -1, mdata)

    im = m.pcolormesh(
        # make sure suqares are over the crosses
        lons - 0.25, lats - 0.25, valid, vmin=-.5, vmax=len(model_keys)-.5,
        norm=norm,
        # edgecolors='None',
        latlon=True, cmap=cmap, alpha=0.6)

    cbar = m.colorbar(im)
    cbar.ax.set_ylabel(" ".join(model_keys))

    plot_errors(m, invalid)

    _plot_and_save(m, cmap, data_g, title)


def plot_scores(scores, extent, lons, lats, data_g, invalid, title):
```

```
    m = make_basemap(extent)

    mscores = np.ma.masked_where(scores > 1, scores)

    im = m.pcolormesh(
        # make sure suqares are over the crosses
        lons - 0.25, lats - 0.25, mscores, vmin=0, vmax=1,
        latlon=True, cmap='RdBu_r', alpha=0.4)

    m.colorbar(im)

    cmap = plt.cm.Greens
    cmap.set_under('0.8')
    cmap.set_bad('0.8')
    cmap.set_over('0.8')

    plot_errors(m, invalid)

    scores[scores > 1] = 0
    mean_rmse = scores.mean()

    _plot_and_save(m, cmap, data_g, f'{title} - mean {mean_rmse}')


def plot_model_map(
        green, grid_model_rmse, invalid, title='rmse', version=None):

    x_size = 1200
    y_size = 1200

    if version:
        geotransform, projection, bbox =  \
            create_lai_cube.extract_lai_meta_v006()
        x_size = 2400
        y_size = 2400
    else:
        geotransform, projection, bbox = create_lai_cube.extract_lai_meta()

    grid = grid_model_rmse.keys()
    grid_model_rmse.values()

    data_g = modis_map.reproject_dataset(
        green, geotransform=geotransform, x_size=x_size, y_size=y_size)

    geo, pro = read_modis.get_meta_geo_info(data_g)

    lons = []
    lats = []

    for x, y in grid:
        lons.append(x)
        lats.append(y)

    lons, lats = np.meshgrid(lons, lats)
    scores = []
    models = []
```

```
    for x, y in zip(lons.flatten(), lats.flatten()):
        rmse_model_list = grid_model_rmse[tuple((x, y))]
        # order by score
        rmse_model_list.sort()
        # lowest rmse
        scores.append(rmse_model_list[0][0])
        # extract the bet model
        models.append(rmse_model_list[0][1])

    scores = np.array(scores)
    scores = scores.reshape(len(lons), len(lats))

    extent = [
        geo[0], geo[0] + data_g.RasterXSize*geo[1],
        geo[3], geo[3] + data_g.RasterYSize*geo[5]]

    # plot_scores(scores, extent, lons, lats, data_g, invalid, "rmse scores")
    plot_models(models, extent, lons, lats, data_g, invalid, "models")


def grid_to_pixels(grid):

    geotransform, projection, bbox = create_lai_cube.extract_lai_meta()

    valid_px = extract_CRU.find_xy_cru_grid(
        geotransform, projection, 1200, 1200, grid)

    return valid_px


def plot_layer(x, lai, green, grid):
    """
    Plot layer fromm lai cube with 2 projections
    each with lon, lat grid of cru data
    """
    onelayer = lai[x, :, :]

    geotransform, projection, bbox = create_lai_cube.extract_lai_meta()

    data = modis_map.reproject_dataset(
        onelayer, geotransform=geotransform, x_size=1200, y_size=1200)

    data_g = modis_map.reproject_dataset(
        green, geotransform=geotransform, x_size=1200, y_size=1200)

    geo, pro = read_modis.get_meta_geo_info(data)

    points4326 = [read_modis.coord2pixel(geo, lon, lat) for lon, lat in grid]

    boxgrid = make_box_grid(grid)

    # valid = []

    # for i in range(len(grid)):
    #     x, y = points4326[i]
    #     if 0 > x  or x > data.RasterXSize:
    #         continue
```

```
#     if 0 > y or y > data.RasterYSize:
#         continue
#     valid.append(grid[i])

valid_px = extract_CRU.find_xy_cru_grid(
    geotransform, projection, 1200, 1200, grid)

box_px = extract_CRU.find_xy_cru_grid(
    geotransform, projection, 1200, 1200, boxgrid)

# no projection raw data
plot_lai(onelayer, green, valid_px, title='one layer of cube')

points4326px = extract_CRU.cru_filter(
    points4326, data.RasterXSize, data.RasterYSize)

extent = [
    geo[0], geo[0] + data.RasterXSize*geo[1],
    geo[3], geo[3] + data.RasterYSize*geo[5]]

# create map using BASEMAP
m = Basemap(
    llcrnrlon=extent[0], llcrnrlat=extent[3],
    urcrnrlon=extent[1], urcrnrlat=extent[2],
    # lat_0=(lat_max - lat_min) / 2,
    # lon_0=(lon_max - lon_min) / 2,
    projection='cyl',
    # projection='cyl',
    resolution='h',
    # area_thresh=10000.,
)

cmap = plt.cm.gist_rainbow
cmap.set_under('0.8')
cmap.set_bad('0.8')
cmap.set_over('0.8')

d = data.ReadAsArray()
data_g = data_g.ReadAsArray()
control = np.copy(d)
d = np.flipud(d)
m.imshow(d, vmin=0, vmax=40, cmap=cmap)

m.drawcoastlines(linewidth=0.5)
m.drawcountries(linewidth=0.5)
parallels = np.arange(49., 79., .5)
# labels = [left,right,top,bottom]
m.drawparallels(parallels, labels=[False, True, True, False])
meridians = np.arange(0., 49., .5)
m.drawmeridians(meridians, labels=[False, True, True, False])

plt.tight_layout()
plt.show()

plot_lai(control, data_g, points4326px)

return box_px, boxgrid
```

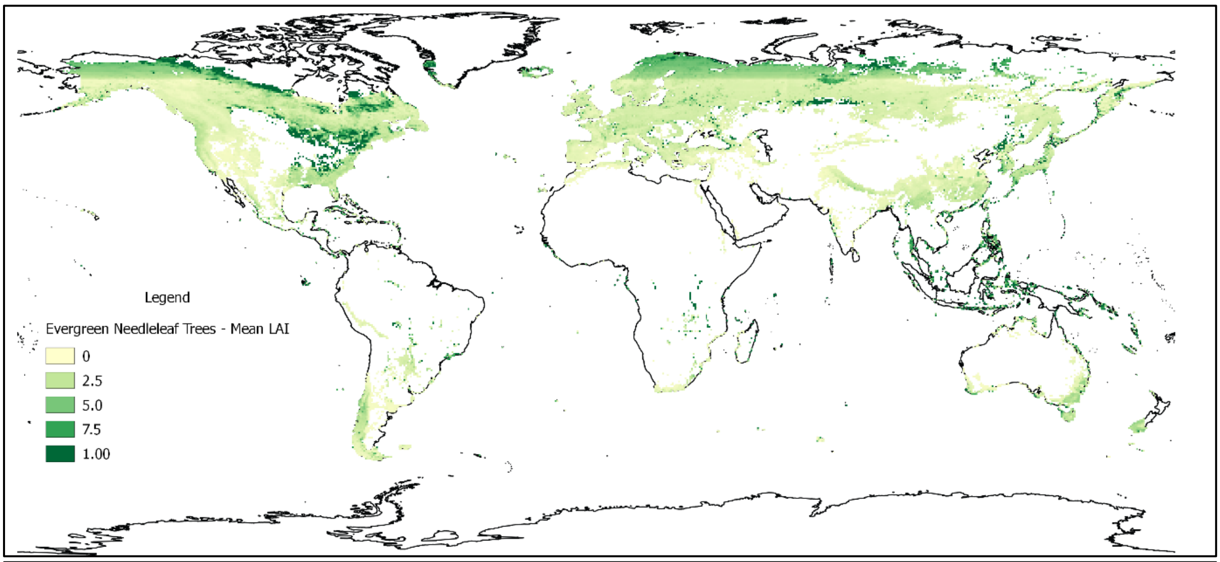
```python
def _plot_rmse_each_model(
        model_options, cru, lai, green, timestamps, grid, box_px):

    for k, v in model_options.items():
        title = k
        model_option = {k: v}
        locations_model_rmse, invalid = calculate_models_for_grid(
            model_option, cru, lai, green, timestamps, grid, box_px)

        # print(len(locations_model_rmse))
        plot_model_map(green, locations_model_rmse, invalid, title=title)


MODEL_OPTIONS = {
    'p4': ['pre', 'pet', 'vap', 'tmp'],
    # 'p3_tmp-vap-pet': ['tmp', 'vap', 'pet'],
    # 'p2_vap_pre': ['vap', 'pre'],
    # 'p3_vap_pre_tmp_one': ['vap', 'pre', tmp_one],
    # 'p2_vap_pet': ['vap', 'pet',],
    # 'p2_pet_pre': ['pet', 'pre',],
    # 'p2_tv': ['tmp', 'vap'],
    # 'p3_pre_pre_pet': [pre_one, 'pre', 'pet'],
    # 'gdd2': [tmp_gdd, pre_one],
    # 'tmp_pet_vap_tmp5': ['vap', 'tmp', 'pet', pet_one, tmp_one, tmp_gdd],
    # 'p6': ['tmp', 'pre', 'vap', 'pet', pre_one],
    # 'gdd_tmp': [tmp_gdd, 'tmp'],
    # 'gdd_tmp_vap': [tmp_gdd, 'tmp', 'vap'],
    # 'tmp': ['tmp'],
    # 'vap': ['vap'],
    # 'pre': ['pre'],
    # 'pet': ['pet'],
}


def main_world(plotlocation=False):
    """
    """
    model_options = MODEL_OPTIONS
    import h5util

    # h5util.print_paths()
    grid = h5util.load_dataset('grid')
    green = h5util.load_dataset('green')
    lai = h5util.load_dataset('lai/smooth_month')
    # lai = h5util.load_dataset('lai/month')
    cru = h5util.load_dataset('cru')
    timestamps = [
        datetime.fromtimestamp(t) for t in h5util.load_dataset('months')]

    geotransform, projection, bbox = create_lai_cube.extract_lai_meta_v006()

    if OPTIONS['debug']:
        box_px, box_lon_lat = plot_layer(20, lai, green, grid)

    boxgrid = make_box_grid(grid)
```

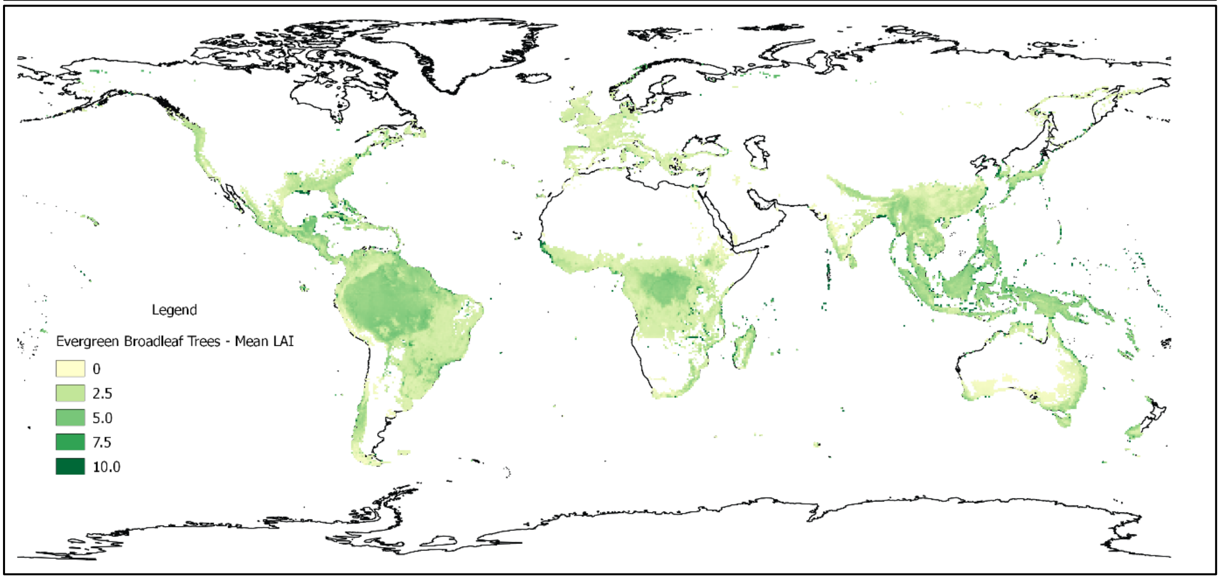
```
    box_px = []

    # plot graph of single location
    if plotlocation:
        make_local_plot(
            grid,
            lai, cru, timestamps,
            geotransform, projection)
        return

    # plot_layer(20, lai, green, grid)
    # return
    for lon, lat in boxgrid:
        x, y = read_modis.determine_xy(geotransform, projection, lon, lat)
        box_px.append((int(x), int(y)))

    for i, (x, y) in enumerate(list(box_px)):
        if 0 > x or x > 1199:
            box_px[i] = None
            continue
        if 0 > y or y > 1199:
            box_px[i] = None
            continue

    assert len(grid)*4 == len(boxgrid), f'{len(grid)} {len(boxgrid)}'

    # make sure grid and boxgrid match
    for g, b in zip(grid, boxgrid[::4]):
        log.debug('%s %s', g, b)

    # aic_criterion(model_options, datasets)
    #  _plot_rmse_each_model(
    #   model_options, cru, lai, green, timestamps, grid, box_px)

    # return

    locations_model_rmse, invalid = calculate_models_for_grid(
        model_options, cru, lai, green, timestamps, grid, box_px)

    # print(len(locations_model_rmse))
    plot_model_map(
        green, locations_model_rmse, invalid,
        title='lowest rmse', version='006')


def main(debug=False, plotlocation=False):

    if debug:
        OPTIONS['debug'] = True

    model_options = MODEL_OPTIONS
    import h5util

    # h5util.print_paths()
    grid = h5util.load_dataset('grid')
    green = h5util.load_dataset('green')
    lai = h5util.load_dataset('lai/smooth_month')
```

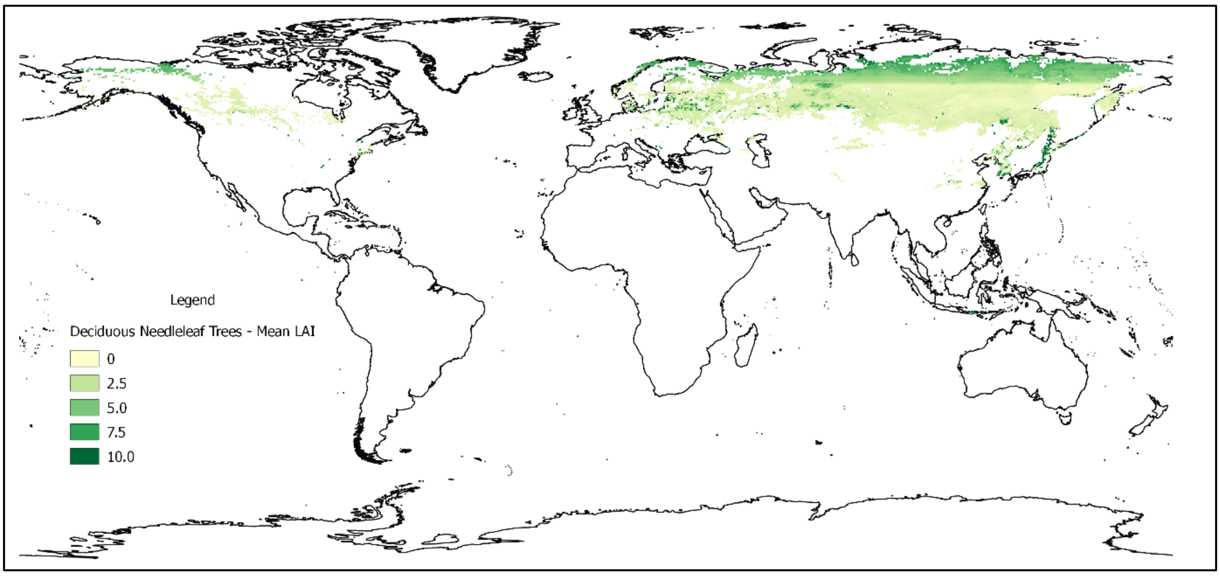
```
    # lai = h5util.load_dataset('lai/month')
    cru = h5util.load_dataset('cru')
    timestamps = [
        datetime.fromtimestamp(t) for t in h5util.load_dataset('months')]

    geotransform, projection, bbox = create_lai_cube.extract_lai_meta()

    if OPTIONS['debug']:
        box_px, box_lon_lat = plot_layer(20, lai, green, grid)

    boxgrid = make_box_grid(grid)
    box_px = []

    # plot graph of single location
    if plotlocation:
        make_local_plot(
            grid,
            lai, cru, timestamps,
            geotransform, projection)
        return

    # plot_layer(20, lai, green, grid)
    # return
    for lon, lat in boxgrid:
        x, y = read_modis.determine_xy(geotransform, projection, lon, lat)
        box_px.append((int(x), int(y)))

    for i, (x, y) in enumerate(list(box_px)):
        if 0 > x or x > 1199:
            box_px[i] = None
            continue
        if 0 > y or y > 1199:
            box_px[i] = None
            continue

    assert len(grid)*4 == len(boxgrid), f'{len(grid)} {len(boxgrid)}'

    # make sure grid and boxgrid match
    for g, b in zip(grid, boxgrid[::4]):
        log.debug('%s %s', g, b)

    # aic_criterion(model_options, datasets)
    #  _plot_rmse_each_model(
    #    model_options, cru, lai, green, timestamps, grid, box_px)

    # return

    locations_model_rmse, invalid = calculate_models_for_grid(
        model_options, cru, lai, green, timestamps, grid, box_px)

    # print(len(locations_model_rmse))
    plot_model_map(green, locations_model_rmse, invalid, title='lowest rmse')


if __name__ == '__main__':
    desc = "Create GREEN matrix cubes of LAI"
    inputparser = argparse.ArgumentParser(desc)
```

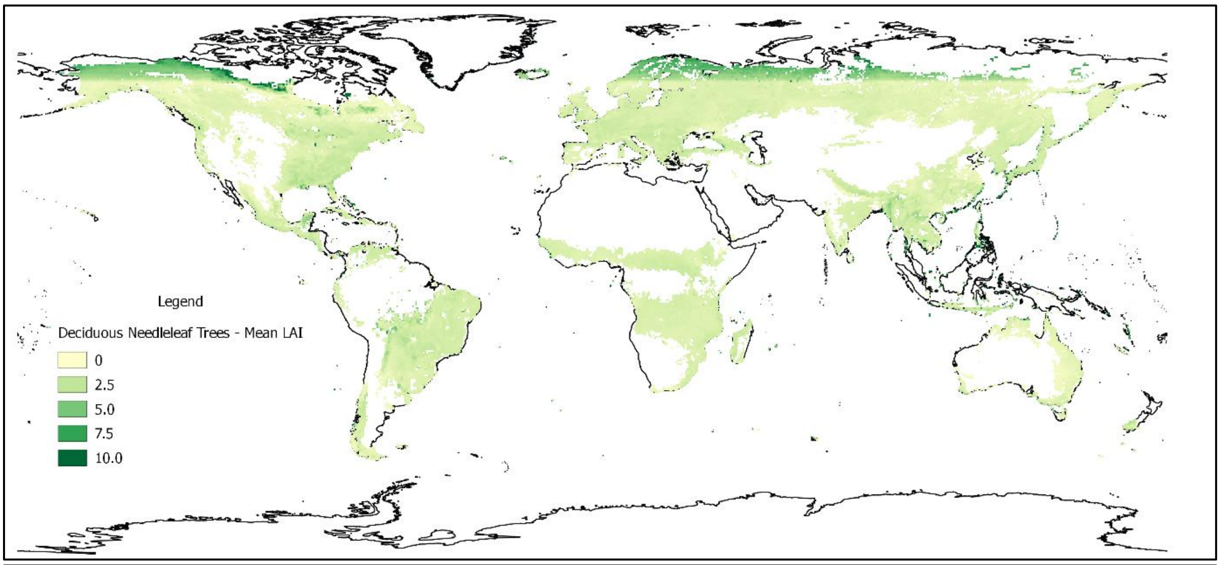
```python
inputparser.add_argument(
    '--debug',
    action='store_true',
    default=False,
    help="Print raw hdf landuse data from direcotory")

inputparser.add_argument(
    '--plotlocation',
    action='store_true',
    default=False,
    help="Make predictor plots of locations definded in settings")

args = inputparser.parse_args()
main(debug=args.debug, plotlocation=args.plotlocation)
```

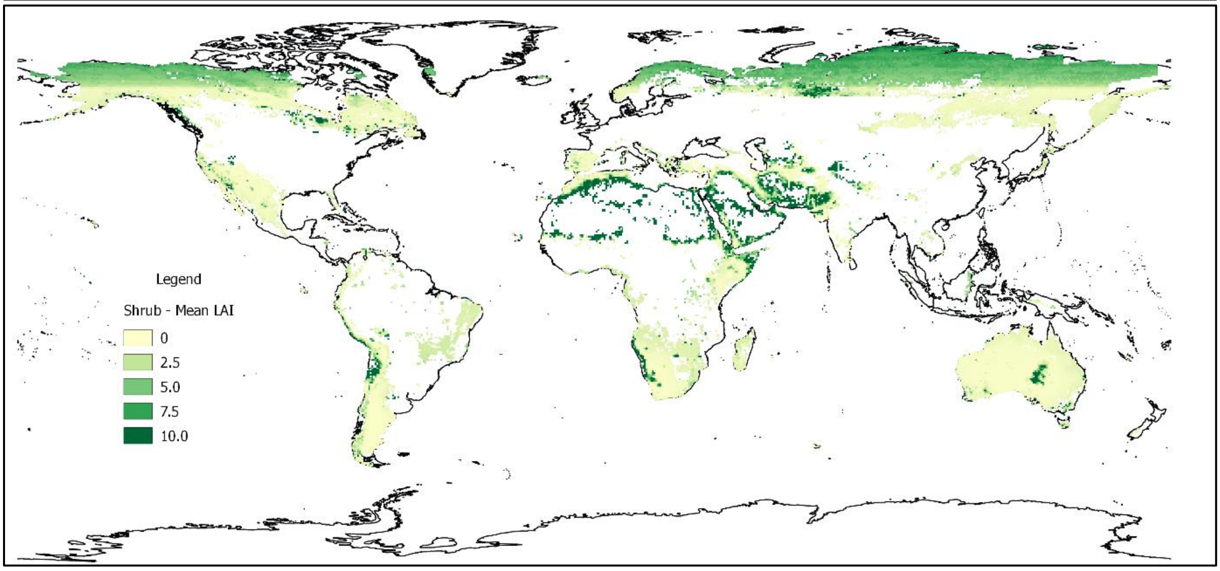
## Appendix E: Global maps of the mean LAI of the plant functional type per grid cell

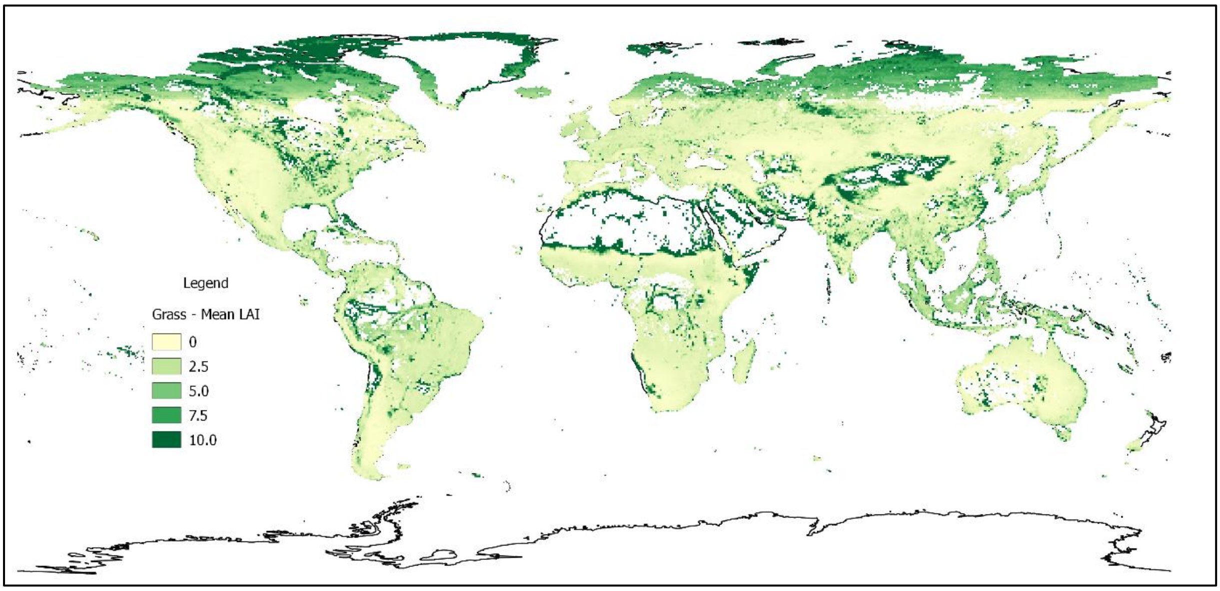


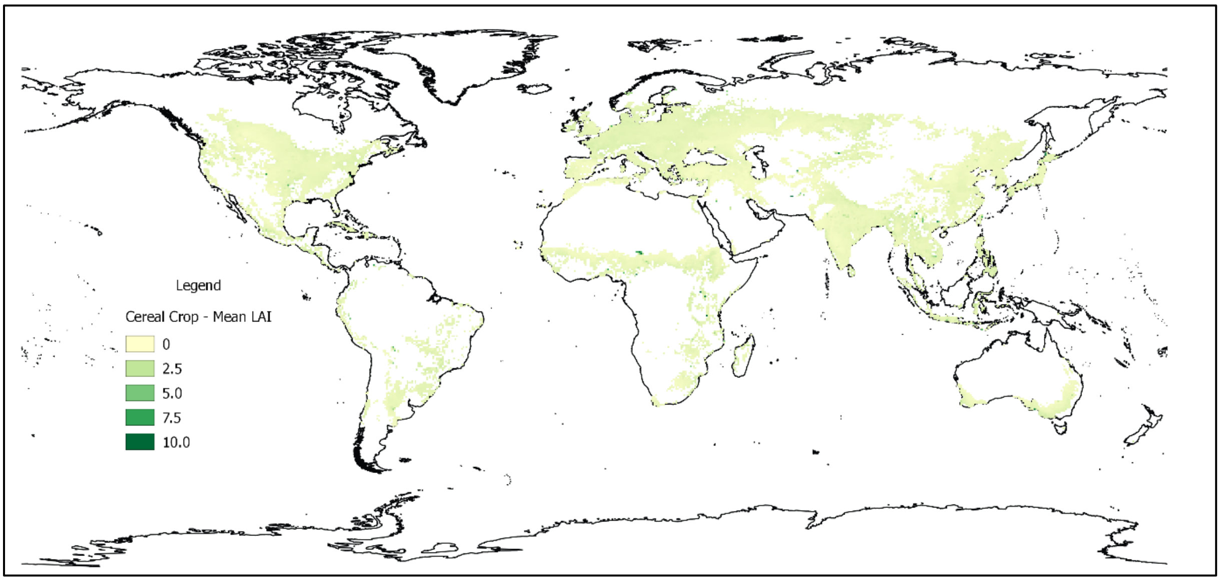



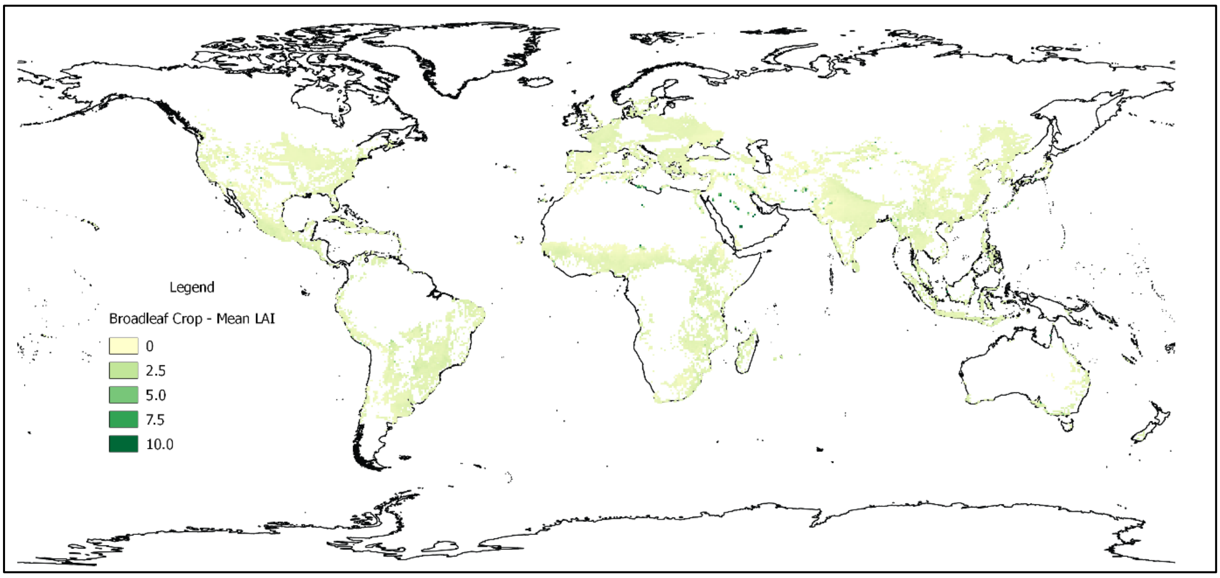




Legend
Deciduous Needleleaf Trees - Mean LAI
0
2.5
5.0
7.5
10.0

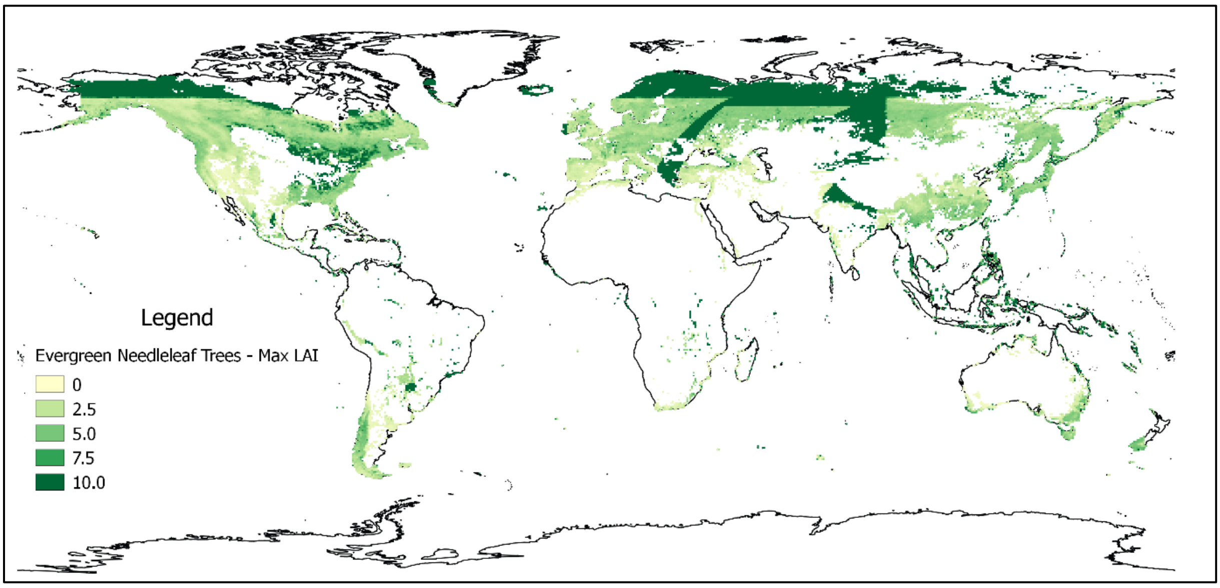

Legend
Shrub - Mean LAI
0
2.5
5.0
7.5
10.0

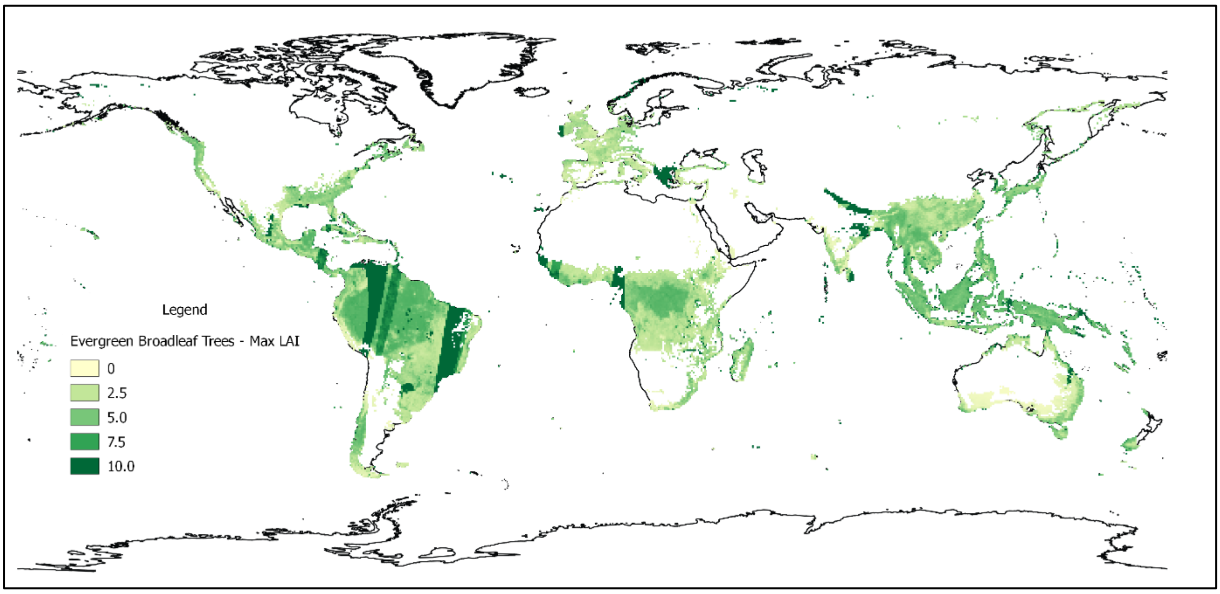

Legend
Grass - Mean LAI
0
2.5
5.0
7.5
10.0

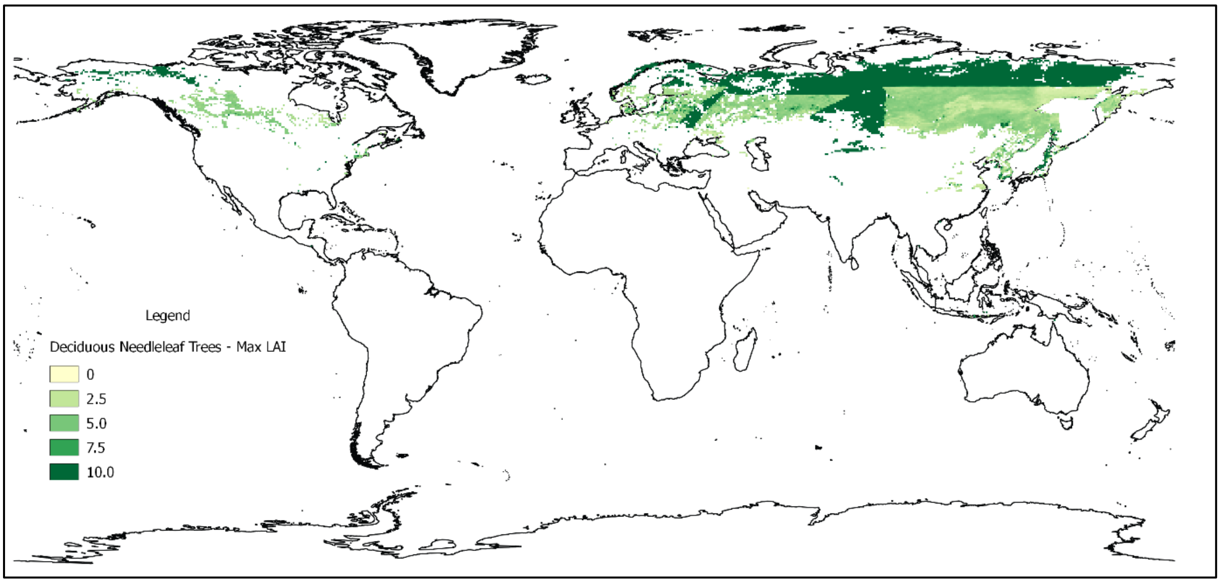

Legend
Cereal Crop - Mean LAI
0
2.5
5.0
7.5
10.0

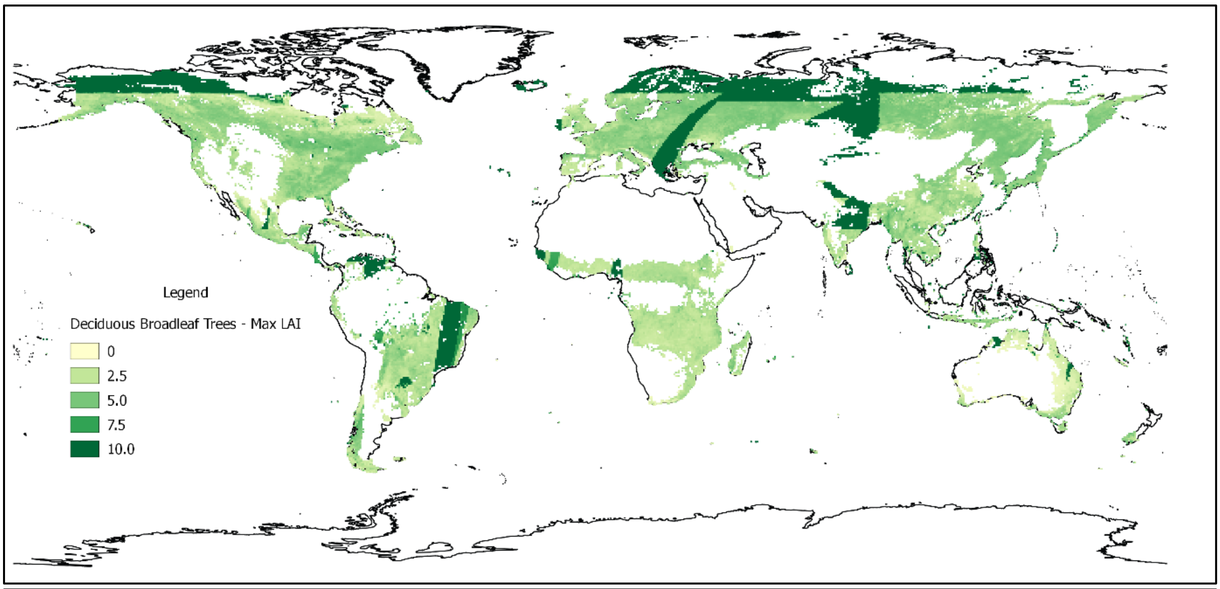

Legend
Broadleaf Crop - Mean LAI
0
2.5
5.0
7.5
10.0

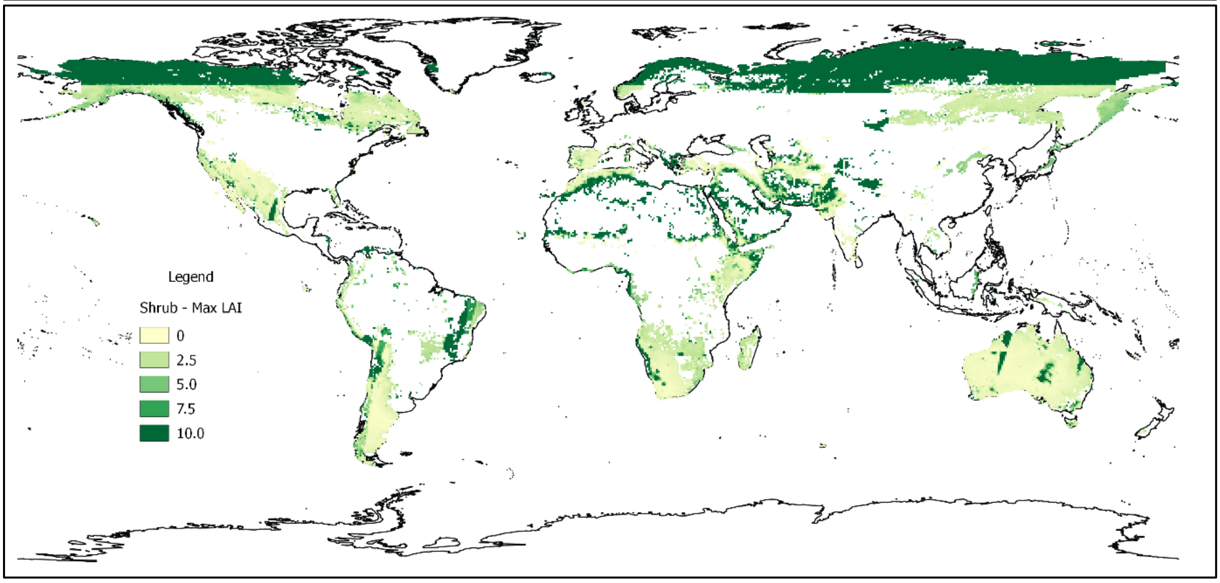
## Appendix F: Global maps of the maximum LAI of the plant functional type per grid cell

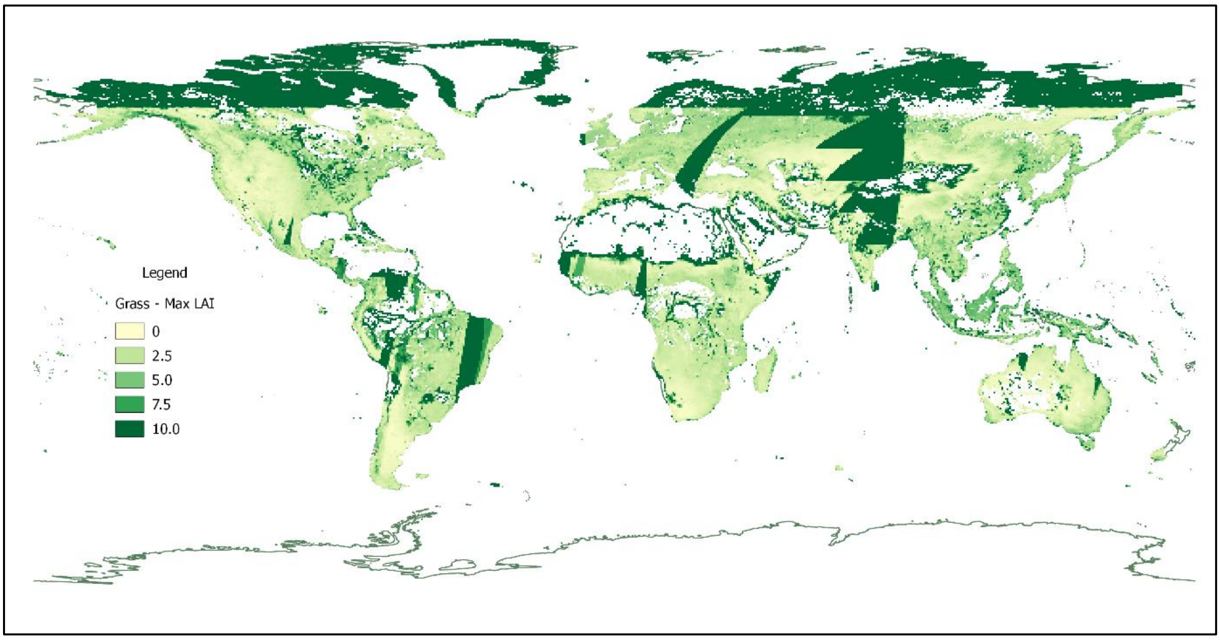


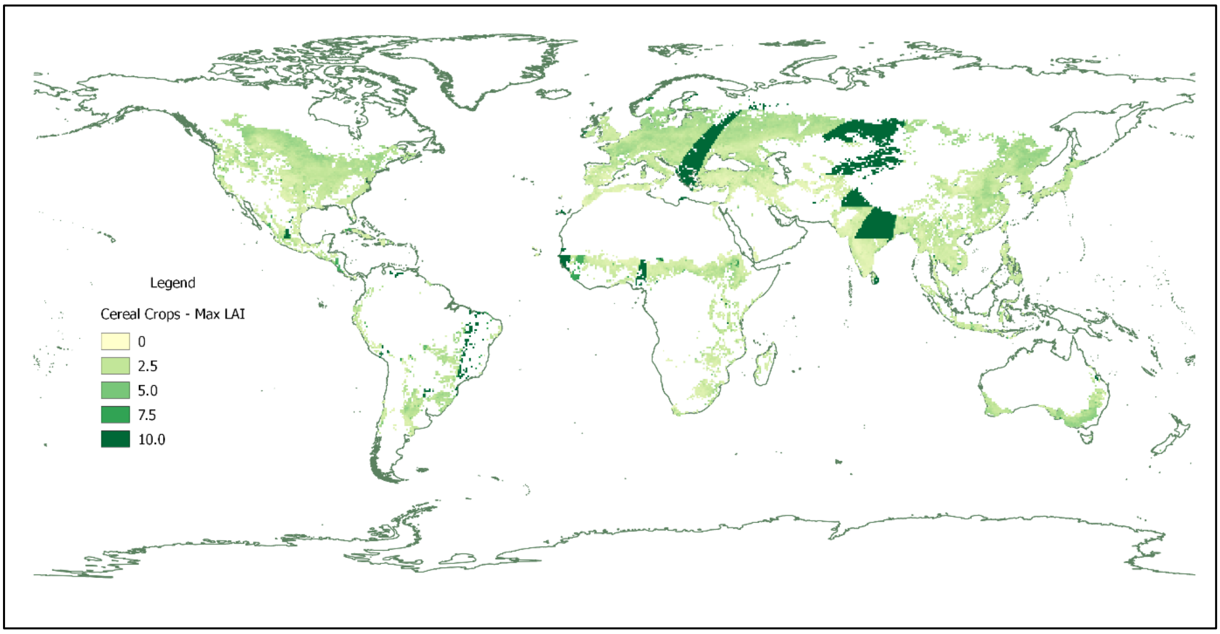



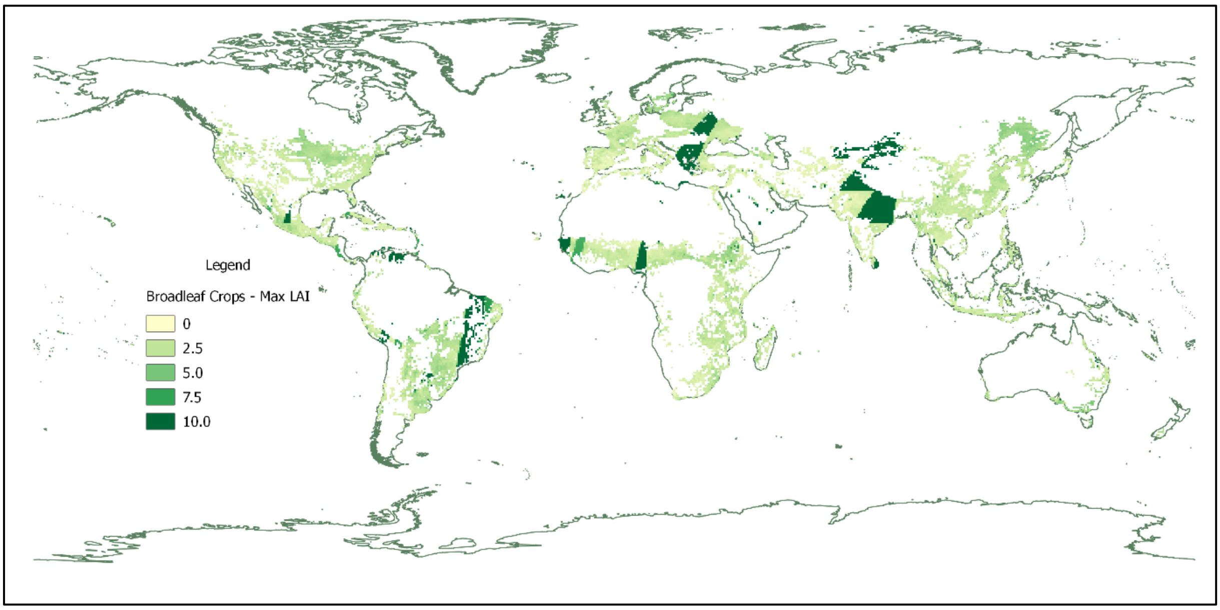

Legend
Deciduous Needleleaf Trees - Max LAI
0
2.5
5.0
7.5
10.0


Legend
Deciduous Broadleaf Trees - Max LAI
0
2.5
5.0
7.5
10.0

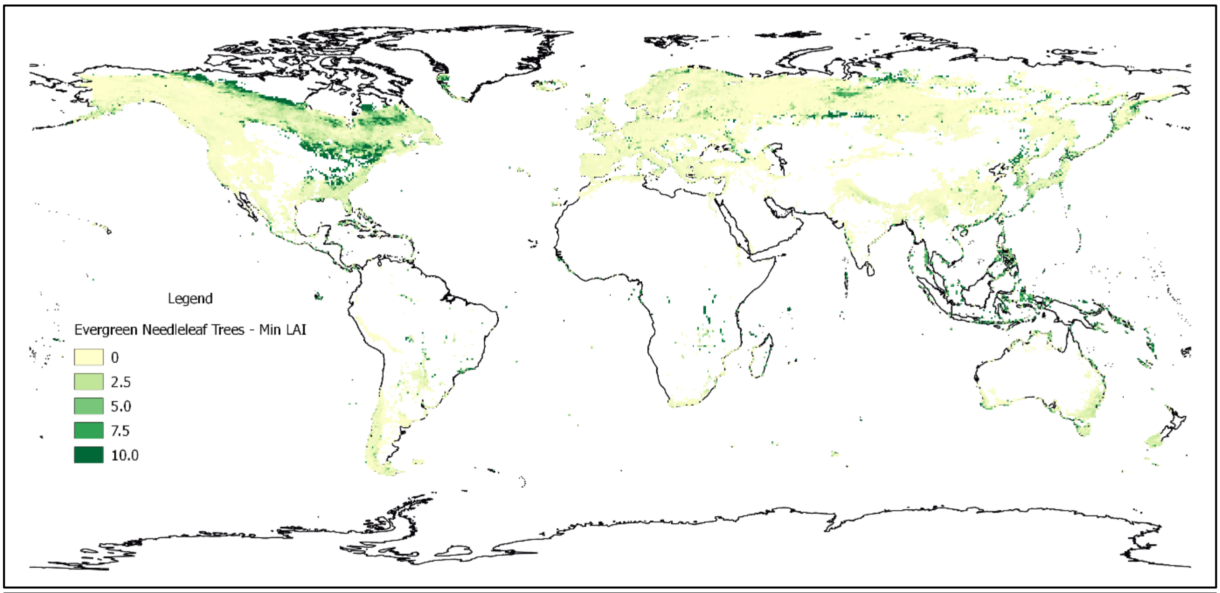

Legend
Shrub - Max LAI
0
2.5
5.0
7.5
10.0

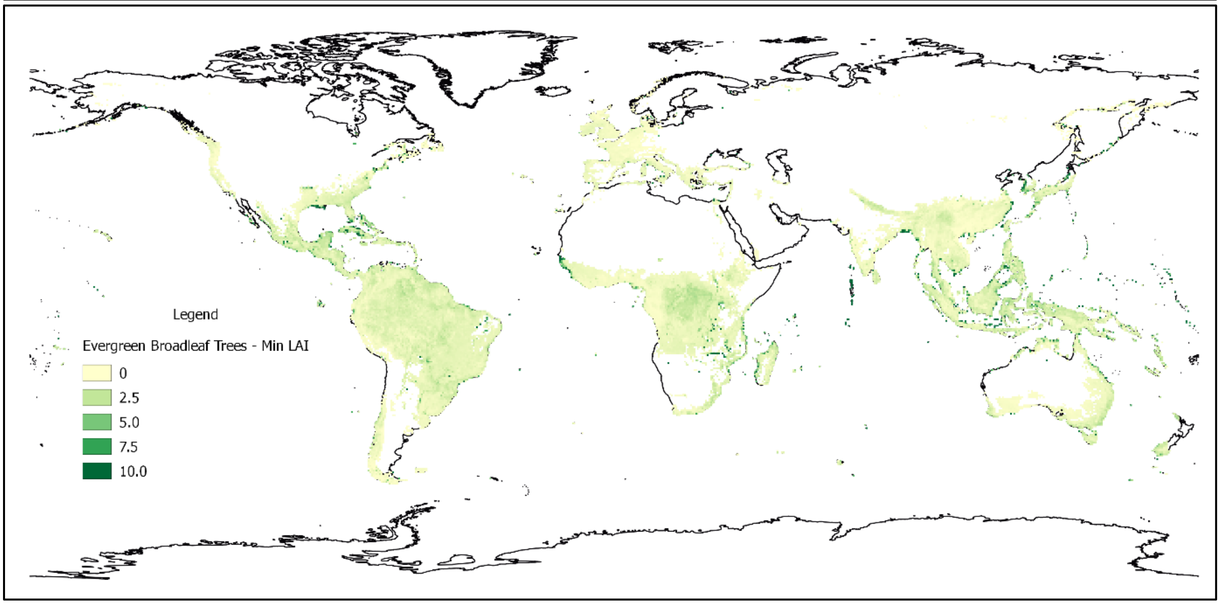

Legend
Grass - Max LAI
0
2.5
5.0
7.5
10.0

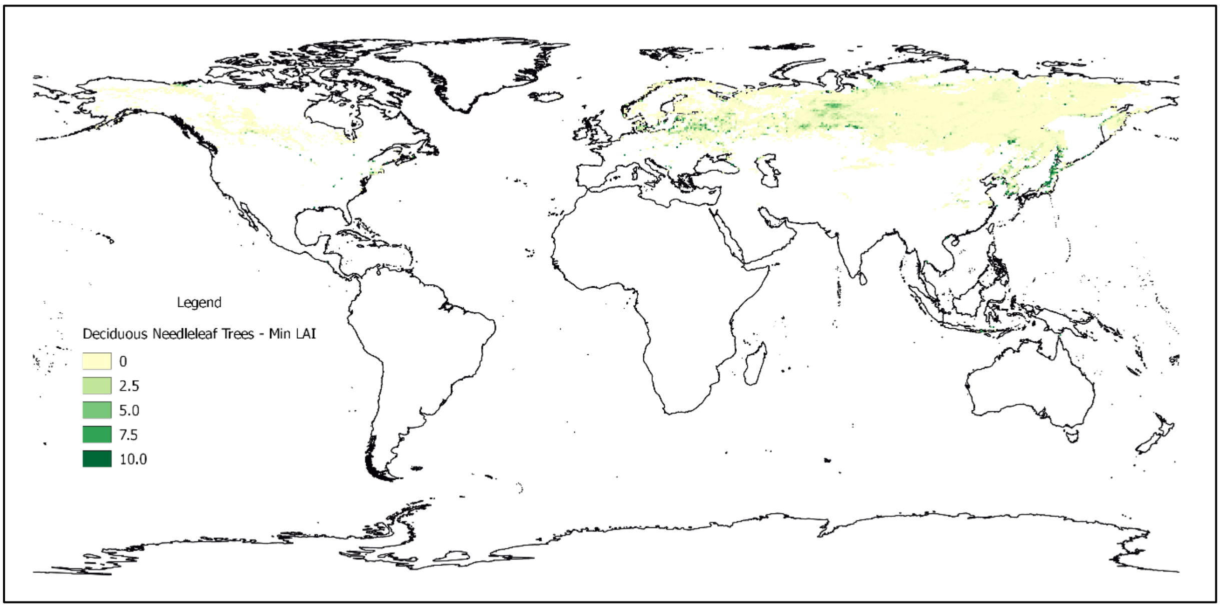


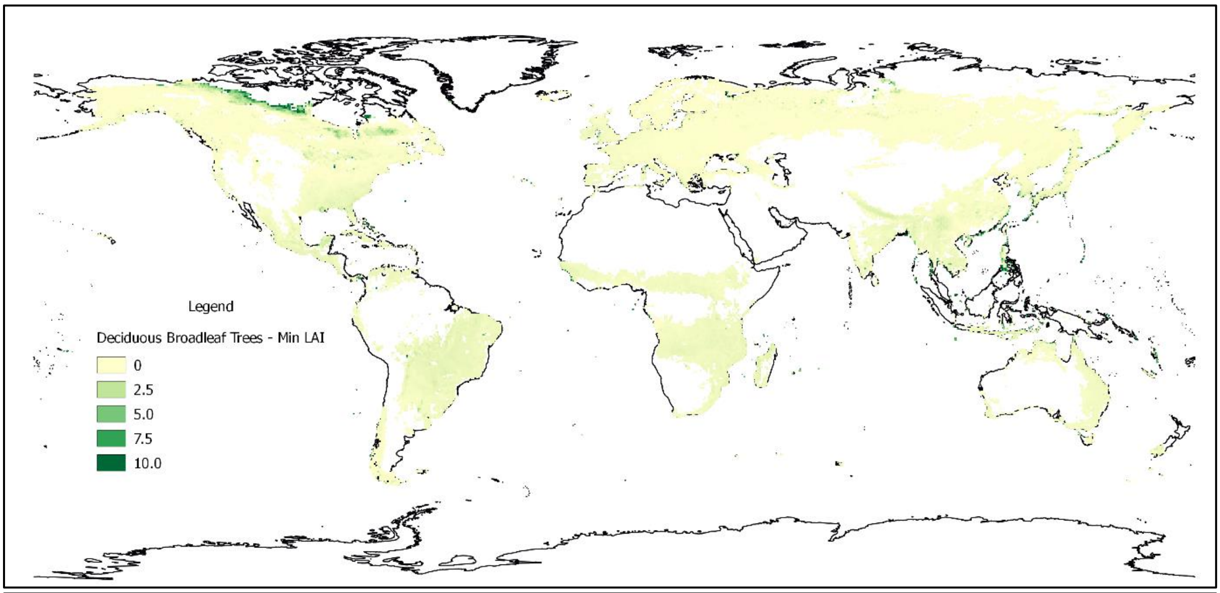



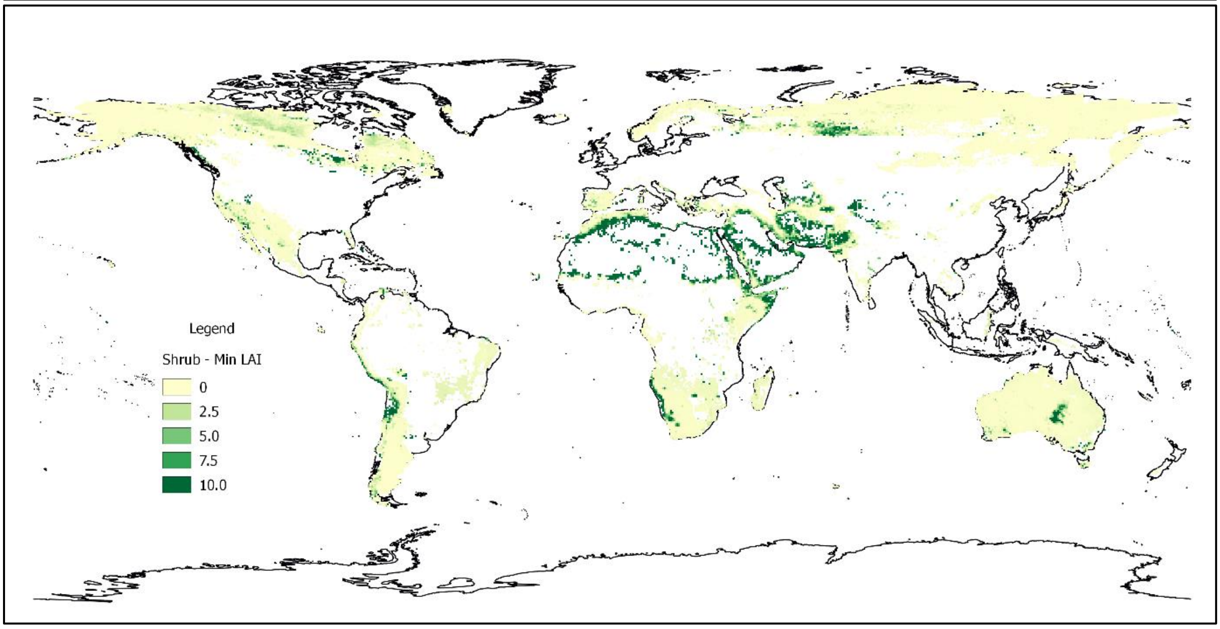

**Appendix G: Global maps of the minimum LAI of the plant functional type per grid cell**

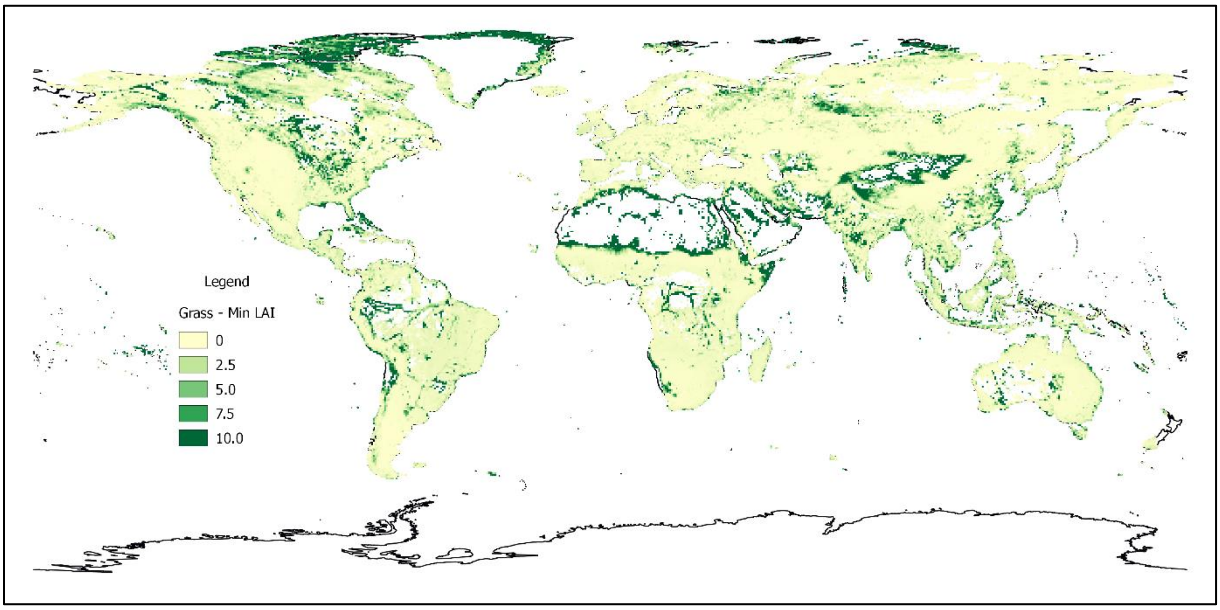

Legend
Evergreen Needleleaf Trees - Min LAI
0
2.5
5.0
7.5
10.0

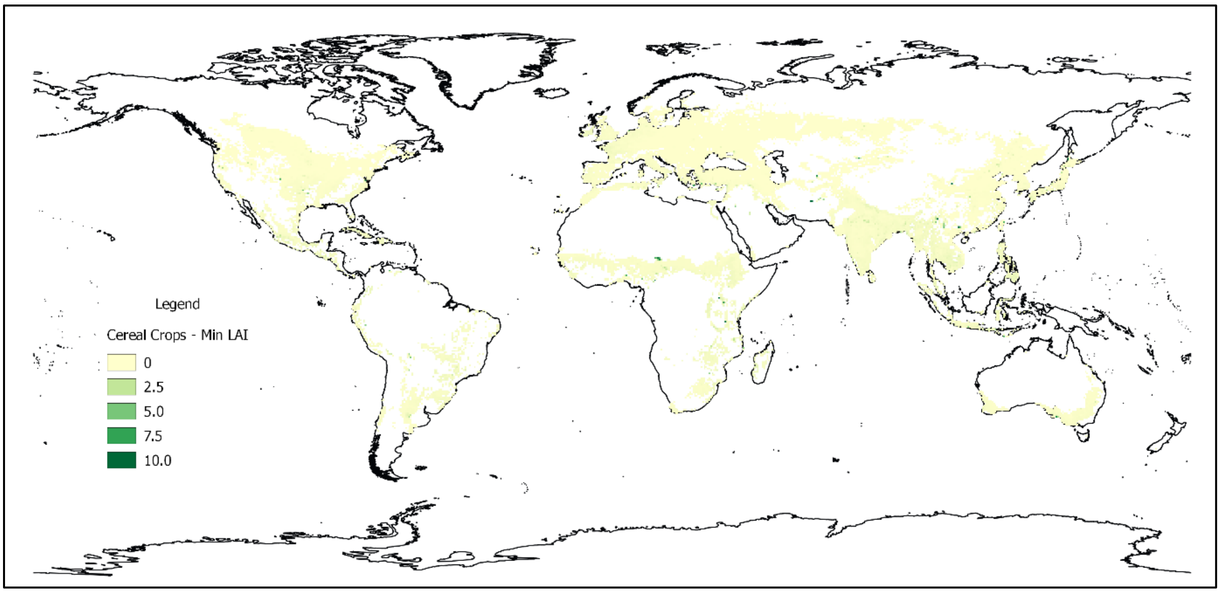

Legend
Evergreen Broadleaf Trees - Min LAI
0
2.5
5.0
7.5
10.0

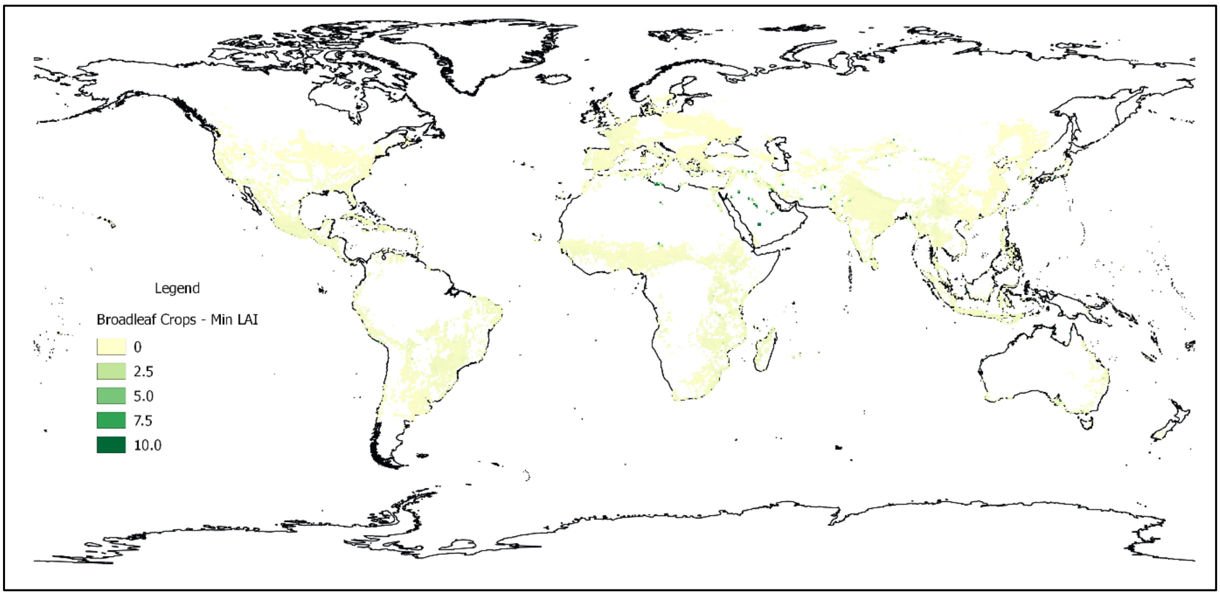

Legend
Deciduous Needleleaf Trees - Min LAI
0
2.5
5.0
7.5
10.0


Legend
Deciduous Broadleaf Trees - Min LAI
0
2.5
5.0
7.5
10.0

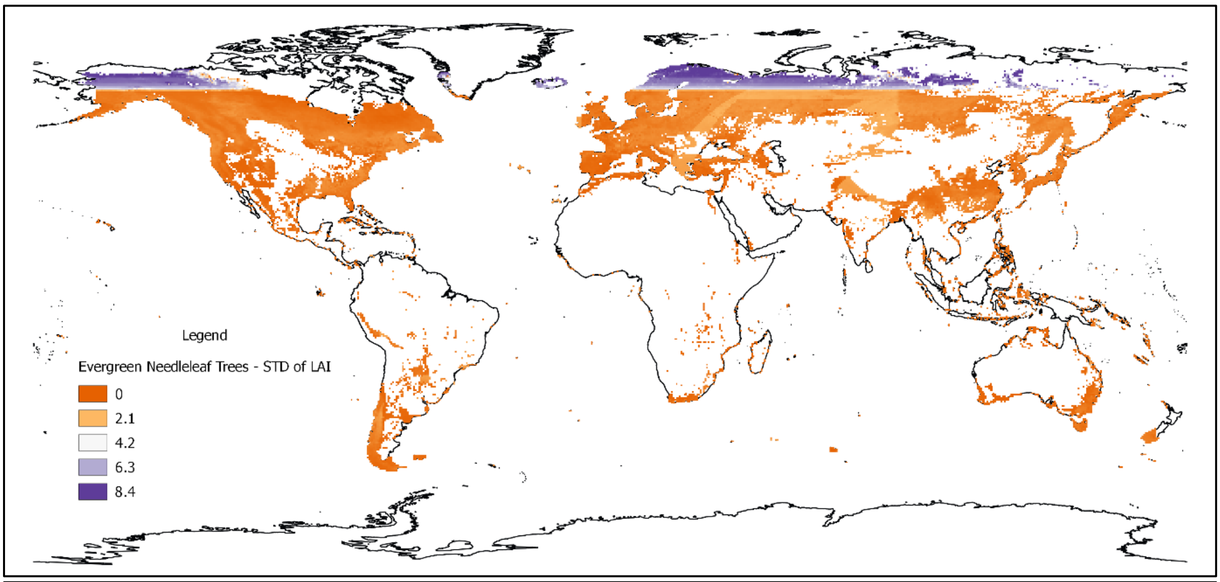

Legend
Shrub - Min LAI
0
2.5
5.0
7.5
10.0

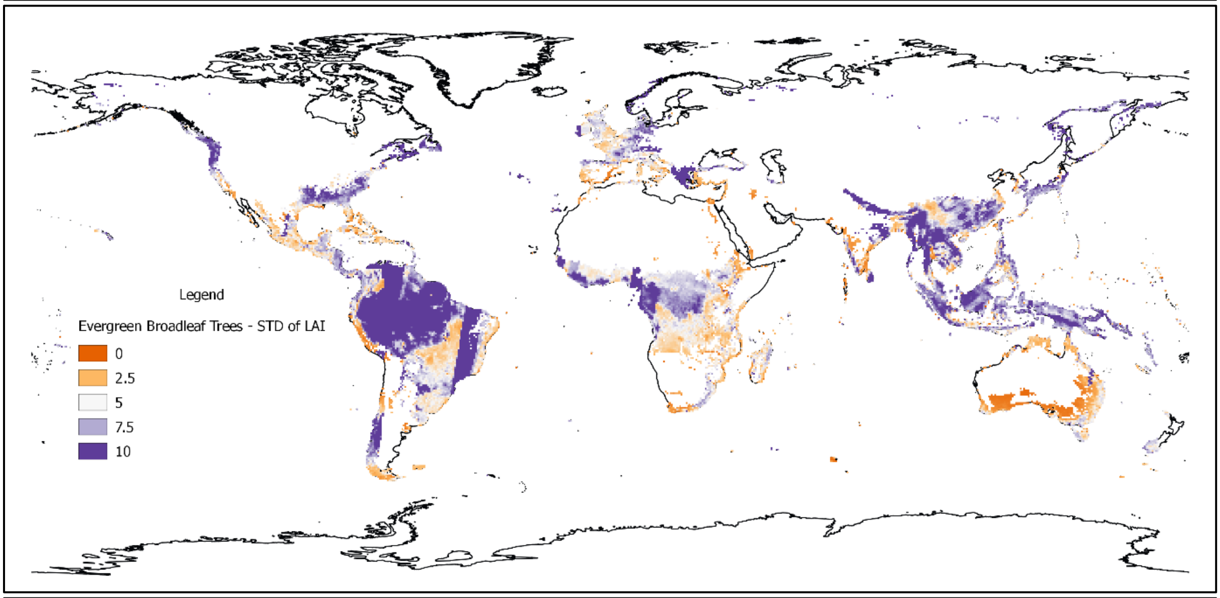

Legend
Grass - Min LAI
0
2.5
5.0
7.5
10.0

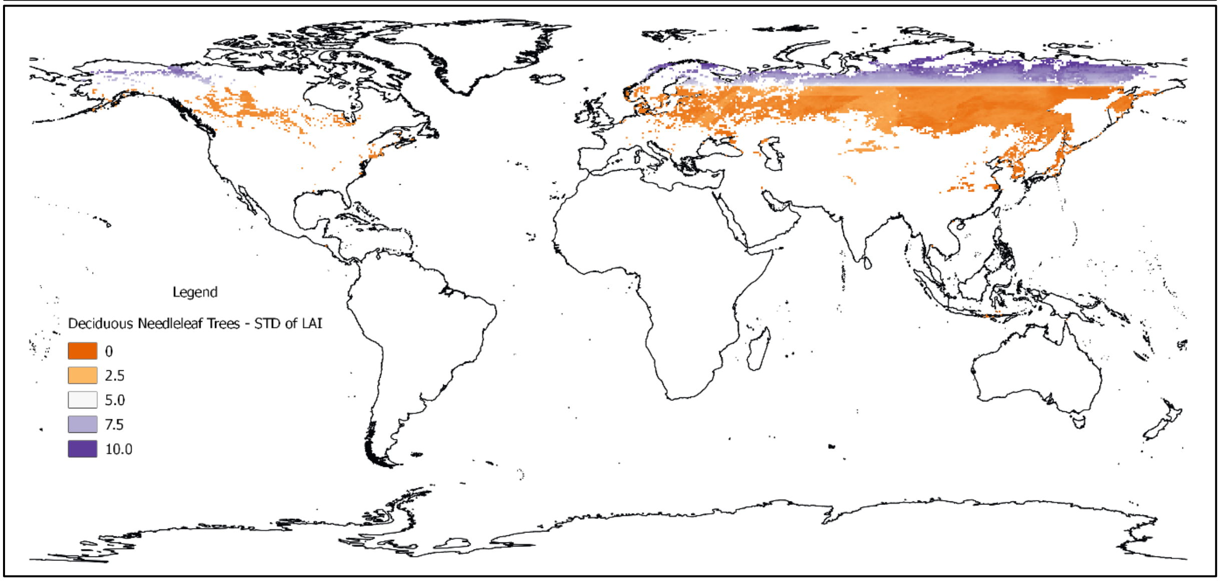

Legend
Cereal Crops - Min LAI
0
2.5
5.0
7.5
10.0

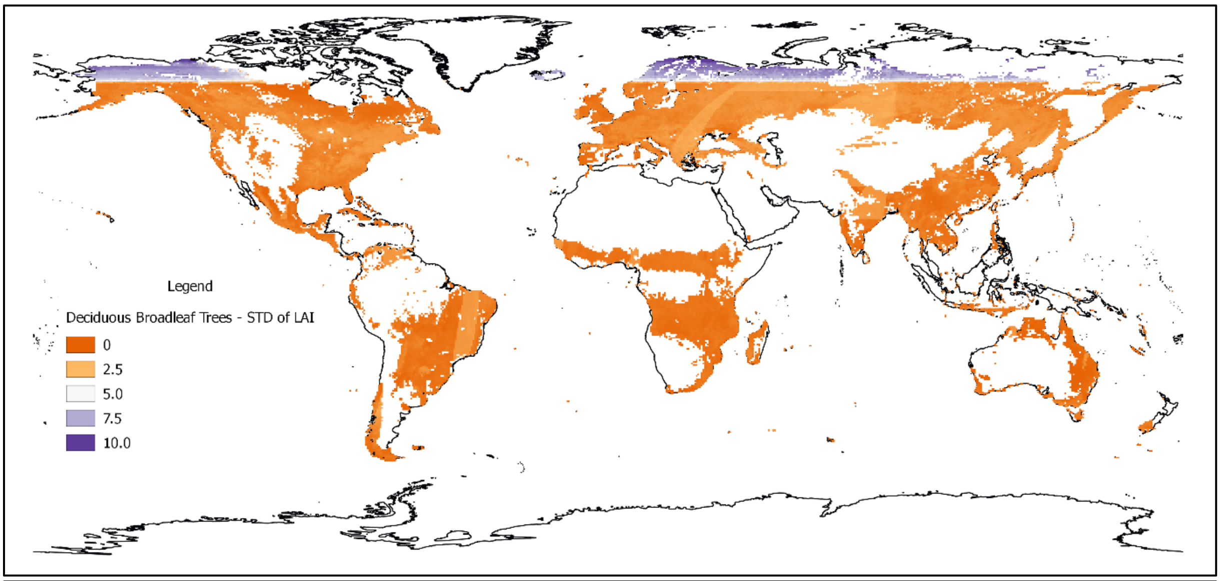

Legend
Broadleaf Crops - Min LAI
0
2.5
5.0
7.5
10.0

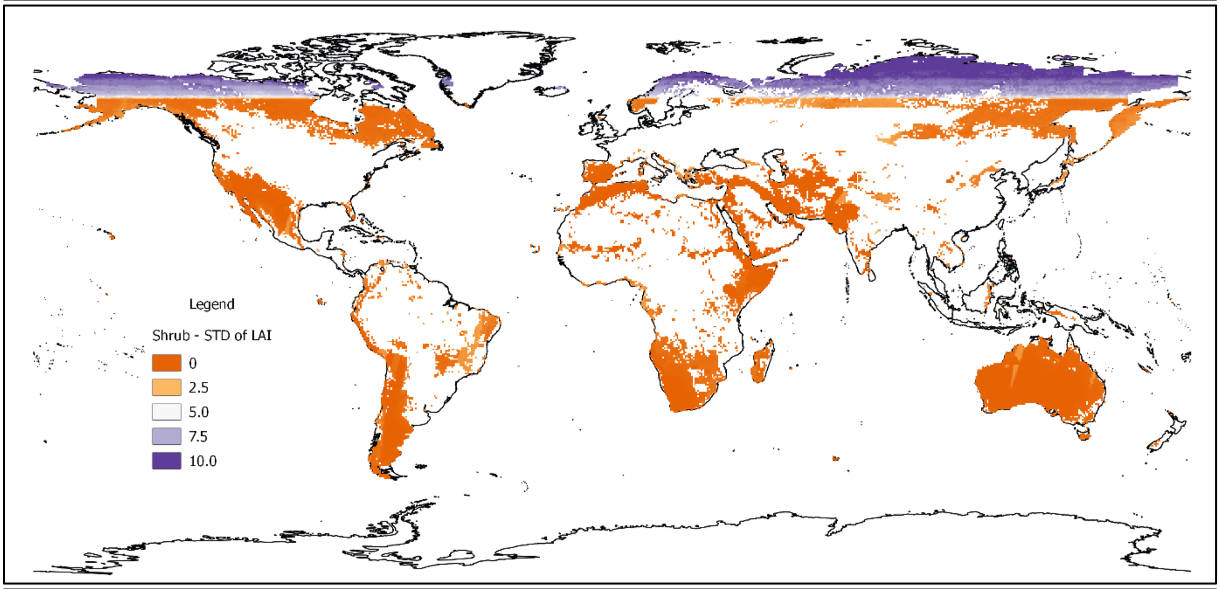
**Appendix H: Global maps of the standard deviation of LAI of the plant function type per grid cell**

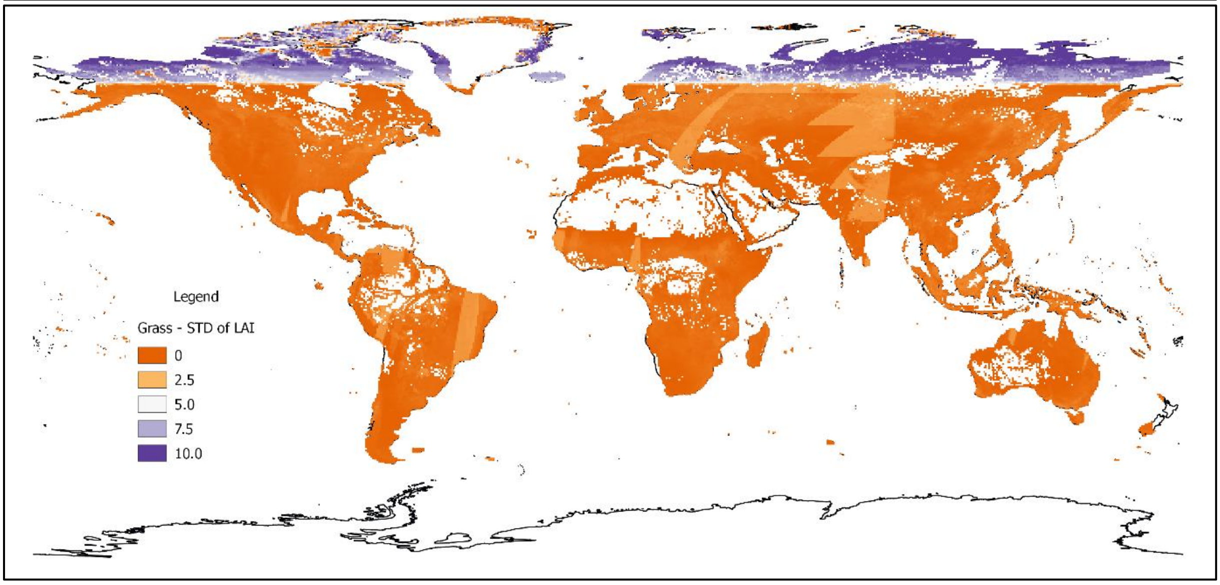


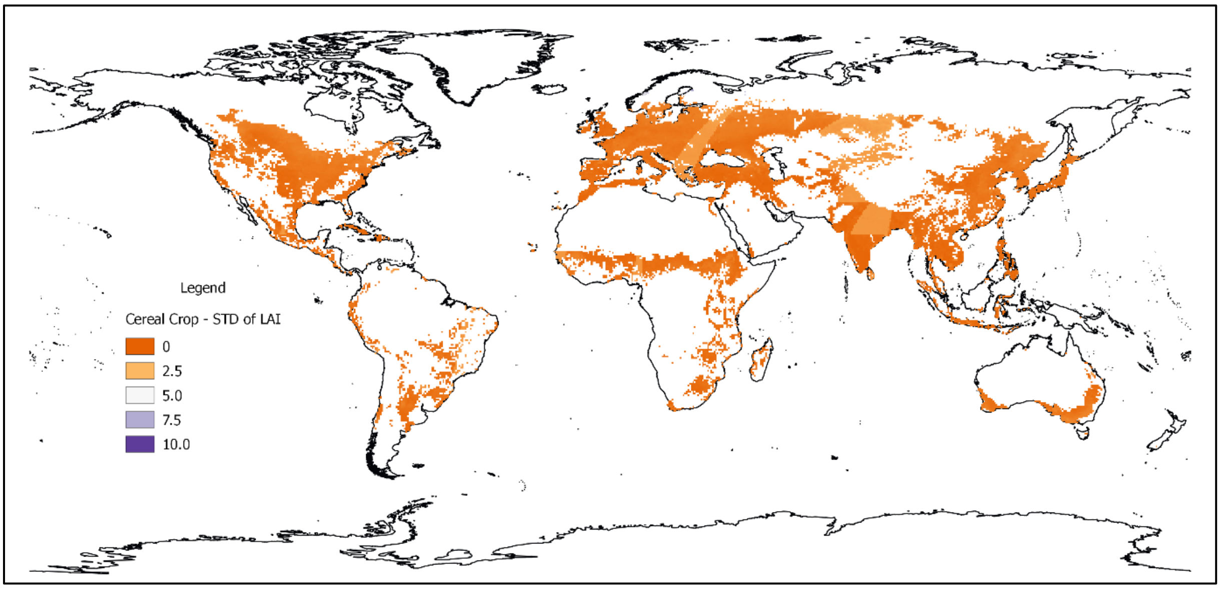



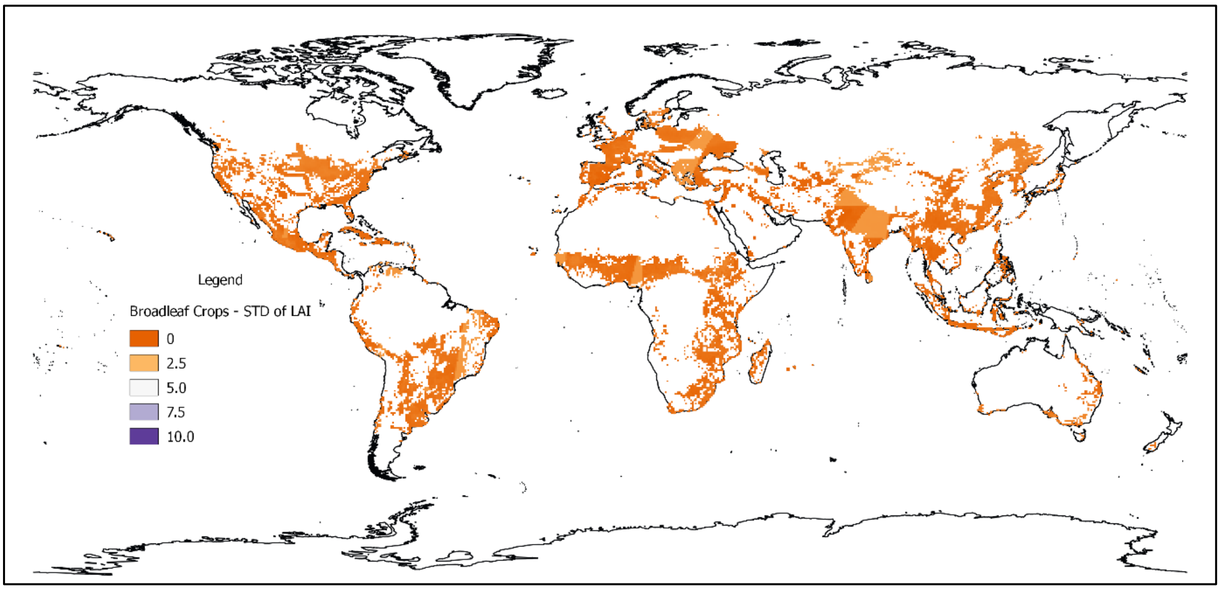




Legend
Deciduous Broadleaf Trees - STD of LAI
0
2.5
5.0
7.5
10.0

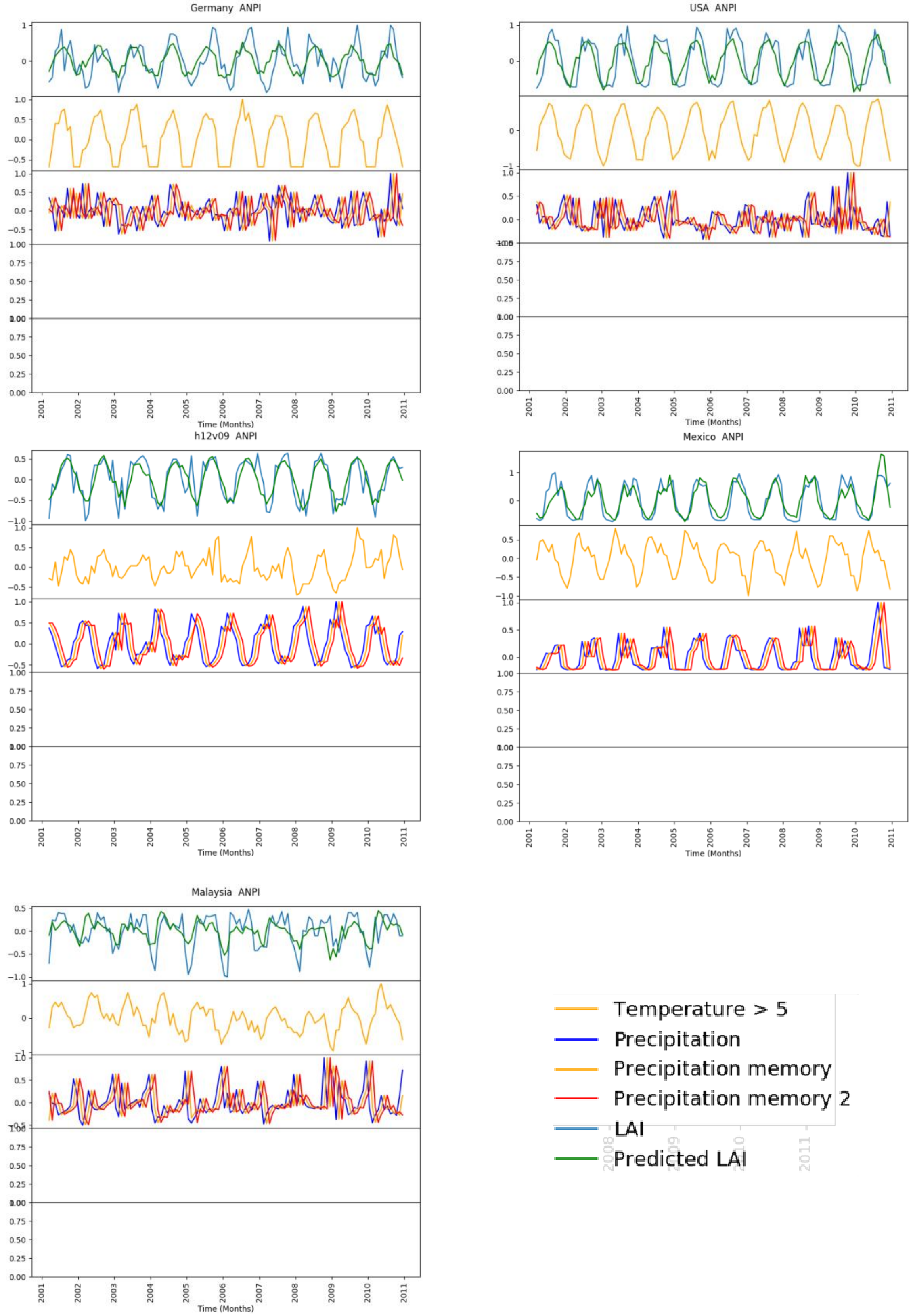

Legend
Shrub - STD of LAI
0
2.5
5.0
7.5
10.0


Legend
Grass - STD of LAI
0
2.5
5.0
7.5
10.0

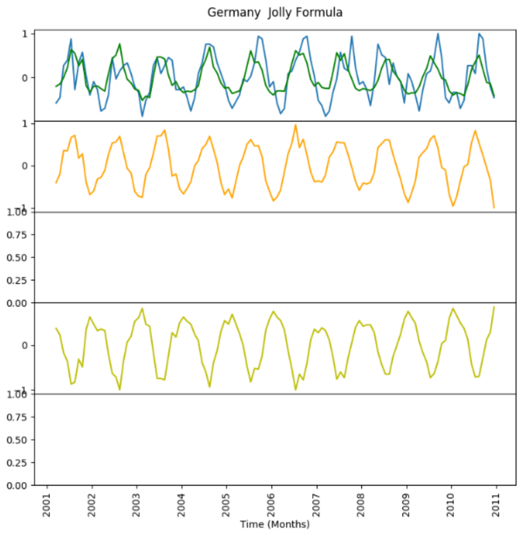

Legend
Cereal Crop - STD of LAI
0
2.5
5.0
7.5
10.0

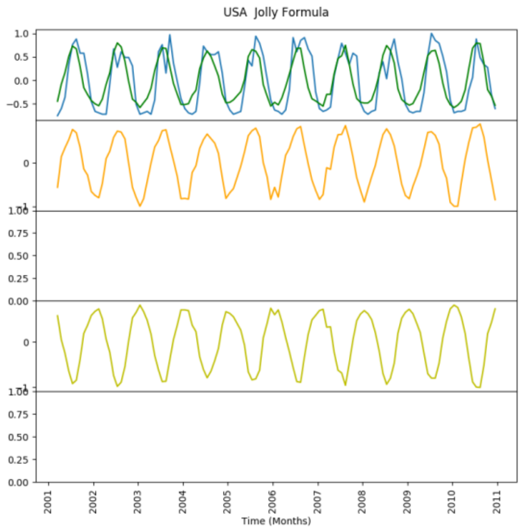

Legend
Broadleaf Crops - STD of LAI
0
2.5
5.0
7.5
10.0

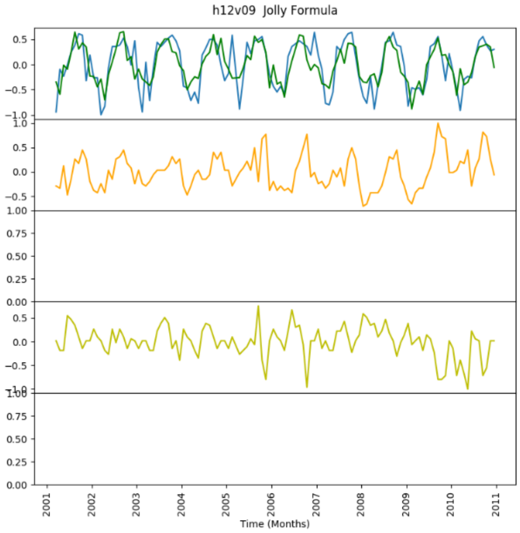
## Appendix I: Time series of climatic variables in ANPI formula for five locations

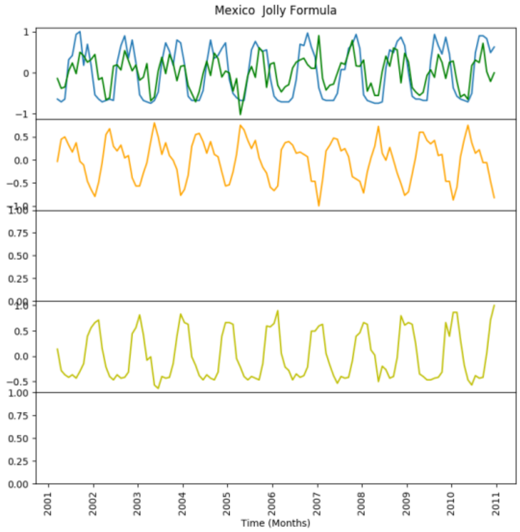


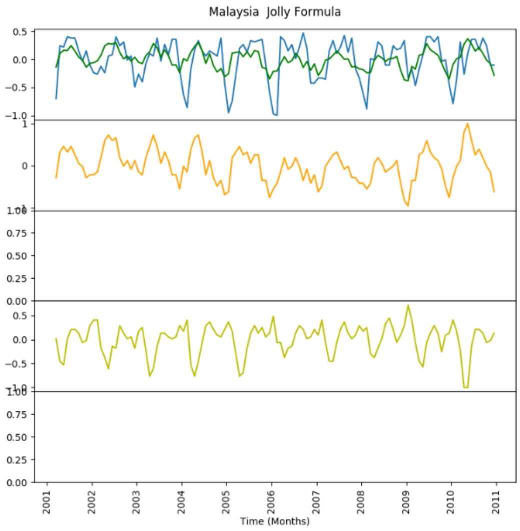
# Appendix J: Time series of climatic variables in Jolly formula for five locations

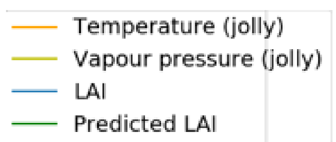


## Appendix K: Time series of climatic variables in S Formula for five locations