Temporal variability of probability density functions of
the leaf area index for boreal and temperate forests

A.J. van Raaij

March 2010
Front page picture: MODIS-based leaf area index values in Europe and Asia, projected on a globe. The darker the color, the higher the leaf area index value.
Temporal variability of probability density functions of the leaf area index for boreal and temperate forests

A.J. van Raaij

Registration number 8603270677040

Supervisors:

Ir. Valérie Laurent
Dr. Ir. Jan Clevers

A thesis submitted in partial fulfillment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands

March 15th 2010
Wageningen, The Netherlands

Thesis code number:    GRS-80436
Thesis Report:         GIRS-2010-04
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing
FOREWORD

This research report is part of an MSc thesis project, which is one of the requisites for obtaining the MSc degree in Geo-Information. The subject for this thesis project is provided by the Centre for Geo-Information, part of Wageningen University & Research centre (WUR).
## TABLE OF CONTENTS

FOREWORD ................................................................................................................................ v  
LIST OF FIGURES.................................................................................................................. viii  
LIST OF TABLES ....................................................................................................................... ix  
ABBREVIATIONS ...................................................................................................................... x  
ABSTRACT ................................................................................................................................. xi  

1. INTRODUCTION ............................................................................................................... 1  
1.1. Background ................................................................................................................... 1  
1.2. Problem definition ....................................................................................................... 2  
1.3. Research Objective & Questions ................................................................................ 3  
1.4. Thesis outline ................................................................................................................ 3  

2. MATERIALS ......................................................................................................................... 4  
2.1. Biome product .............................................................................................................. 4  
2.1.1. Boreal coniferous forest ....................................................................................... 6  
2.1.2. Temperate coniferous forest ............................................................................... 6  
2.1.3. Temperate broadleaf and mixed forest ............................................................. 6  
2.2. Leaf Area Index product ............................................................................................. 6  
2.3. Land cover product ..................................................................................................... 8  
2.4. Tree cover product ....................................................................................................... 9  

3. METHODOLOGY ............................................................................................................. 10  
3.1. Data preprocessing .................................................................................................... 10  
3.2. Biome specific temporal analysis ............................................................................. 12  
3.3. Tree cover analysis ..................................................................................................... 12  
3.4. LAI algorithm analysis .............................................................................................. 12  
3.5. Land cover analysis ................................................................................................... 12  

4. RESULTS ............................................................................................................................. 14  
4.1. Biome specific temporal analysis ............................................................................. 14  
4.1.1. Boreal coniferous forest ..................................................................................... 14  
4.1.2. Temperate coniferous forest ............................................................................. 16  
4.1.3. Temperate broadleaf and mixed forest ............................................................. 18  
4.2. Tree cover analysis ..................................................................................................... 20  
4.2.1. Boreal coniferous forest ..................................................................................... 20  
4.2.2. Temperate coniferous forest ............................................................................. 21  
4.2.3. Temperate broadleaf and mixed forest ............................................................. 23
4.3. LAI algorithm analysis .............................................................................................. 24
4.4. Land cover analysis ................................................................................................. 24
  4.4.1. Boreal coniferous forest .................................................................................... 25
  4.4.2. Temperate coniferous forest ............................................................................. 25
  4.4.3. Temperate broadleaf and mixed forest ........................................................... 26

5. DISCUSSION ................................................................................................................. 28
  5.1. Selected source data .............................................................................................. 28
  5.2. Methodological aspects ......................................................................................... 29
  5.3. Interpretation of PDF’s .......................................................................................... 30
    5.3.1. Double peak feature ....................................................................................... 30
    5.3.2. Sensitivity of the PDF to tree cover ................................................................. 32

6. CONCLUSIONS & RECOMMENDATIONS .................................................................. 33
  6.1. Conclusions ........................................................................................................... 33
  6.2. Recommendations .................................................................................................. 34

REFERENCES ....................................................................................................................... 35
APPENDIX 1: Box plot of LAI for temperate broadleaf and mixed forests in 2003..... 38
LIST OF FIGURES

Figure 1: Spatial extent of three biomes defined by the WWF .....................................................5
Figure 2: Data processing chain: Operations in green indicate preprocessing. Operations in red indicate selecting LAI subset .....................................................................................................11
Figure 3: Spatial distribution of high and low quality pixels for the boreal region in 2003....14
Figure 4: Monthly average PDF of LAI for boreal forest..............................................................15
Figure 5: Yearly summer average PDF of LAI for boreal forest.....................................................15
Figure 6: Statistical measures characterizing LAI values for boreal forest during summer. ..16
Figure 7: Multiyear average PDF of LAI for boreal forests based on summer periods. ............16
Figure 8: Probability distribution of LAI for temperate coniferous forest in 2003. ...............17
Figure 9: Probability distribution of LAI for temperate coniferous forest in 2008. ...............17
Figure 10: Differences in PDF between 2008 and 2003 for temperate coniferous forest ......18
Figure 11: Probability distribution of LAI for temperate broadleaf and mixed forests in 2003. ...................................................................................................................................................18
Figure 12: Probability distribution of LAI for temperate broadleaf and mixed forests in 2008. ...................................................................................................................................................19
Figure 13: Difference in PDF between 2008 and 2003 for temperate broadleaf and mixed forests...................................................................................................................................................20
Figure 14: PDF’s for boreal coniferous forest in July 2003 using different percentages of tree cover ...................................................................................................................................................20
Figure 15: Contour view of PDF’s for boreal coniferous forest in July 2003 using different percentages of tree cover ...................................................................................................................................................21
Figure 16: Relative distribution of pixels among the 10 tree cover classes for boreal coniferous forest ...................................................................................................................................................21
Figure 17: PDF’s for temperate coniferous forest in July 2003 using different percentages of tree cover ...................................................................................................................................................22
Figure 18: Contour view of PDF’s for temperate coniferous forest in July 2003 using different percentages of tree cover ...................................................................................................................................................22
Figure 19: Relative distribution of pixels among the 10 tree cover classes for temperate coniferous forest ...................................................................................................................................................22
Figure 20: PDF’s for temperate broadleaf and mixed forest in July 2003 using different percentages of tree cover ................................................................................................................... 23

Figure 21: Contour view of PDF’s for temperate broadleaf and mixed forest in July 2003 using different percentages of tree cover ................................................................................................................... 23

Figure 22: Relative distribution of pixels among the 10 tree cover classes for temperate broadleaf and mixed forest ................................................................................................................... 23

Figure 23: PDF of LAI for boreal forest in July 2003 using different algorithms ................................................................................................................... 24

Figure 24: PDF of LAI for boreal forest in July 2003 for the four largest MODIS land cover classes ................................................................................................................... 25

Figure 25: PDF of LAI temperate coniferous forest in July 2003 for the four largest MODIS land cover classes ................................................................................................................... 26

Figure 26: PDF of LAI temperate broadleaf and mixed forest in July 2003 for the four largest MODIS land cover classes ................................................................................................................... 27

Figure 27: Spatial distribution of temperate broadleaf & mixed forest LAI values for August 2003 in Europe. Green pixels are part of the first peak, red pixels are part of the second peak. ................................................................................................................... 31

LIST OF TABLES

Table 1: Biomes (Major Habitat Types) defined by the WWF. ................................................................................................................... 5

Table 2: Description QC-values ........................................................................................................................................................................ 8

Table 3: MODIS land cover classes used by RT-algorithm ........................................................................................................................................ 8

Table 4: Number of pixels per MODIS land cover class for boreal forest in July 2003 ................................................................................................................... 25

Table 5: Number of pixels per MODIS land cover class for temperate coniferous forest in July 2003 ................................................................................................................... 26

Table 6: Number of pixels per MODIS land cover class for temperate broadleaf and mixed forest in July 2003 ................................................................................................................... 27
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)ATSR</td>
<td>(Advanced) Along-Track Scanning Radiometer</td>
</tr>
<tr>
<td>FS-USDA</td>
<td>Forest Service of the United States Department of Agriculture</td>
</tr>
<tr>
<td>GCM</td>
<td>General Circulation Model</td>
</tr>
<tr>
<td>HDF</td>
<td>Hierarchical Data Format</td>
</tr>
<tr>
<td>LAI</td>
<td>Leaf Area Index</td>
</tr>
<tr>
<td>LP DAAC</td>
<td>Land Processes Distributed Active Archive Center</td>
</tr>
<tr>
<td>LUT</td>
<td>Look-Up Table</td>
</tr>
<tr>
<td>MERIS</td>
<td>MEdition Resolution Imaging Spectrometer</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate-resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>QC</td>
<td>Quality Control</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>RT-algorithm</td>
<td>Radiative Transfer algorithm</td>
</tr>
<tr>
<td>TC</td>
<td>Tree Cover</td>
</tr>
<tr>
<td>VCF</td>
<td>Vegetation Continuous Field</td>
</tr>
<tr>
<td>WGS1984</td>
<td>World Geodetic System 1984</td>
</tr>
<tr>
<td>WWF</td>
<td>World Wide Fund for nature, alt: World Wildlife Fund</td>
</tr>
</tbody>
</table>
ABSTRACT

The Leaf Area Index (LAI) is one of the basic quantities used in estimation of net primary production and modeling of terrestrial carbon cycle processes. Several methods have been developed to measure the LAI. One of the currently most widely used methods is the remote sensing based method. Because satellite systems produce large extent data and a spatially continuous representation of the land surface, this method is nowadays preferred over other methods. One of the remote sensing systems currently used is the MODIS device on board the spaceborne Terra and Aqua platforms. Until now, most analyses of MODIS LAI data were based on site or biome specific average values and the changes of these averages over time. In this research, the possibility of using a MODIS-based probability density function (PDF) to analyze temporal variations of the LAI was explored. Three biomes were selected for analysis: boreal coniferous forest, temperate coniferous forest, and temperate broadleaf and mixed forest. Results showed a large probability peak around LAI = 1 during winter months, which shifts during growing season towards higher LAI values around LAI = 2.4. A second peak appears during summer around LAI = 5.2. In between these two probability peaks, a dip fixed at LAI = 4.3 is present. This dip is present in the PDF’s of all summer months of all three studied biomes and is thought to be caused by the MODIS radiative transfer algorithm, the algorithm upon which most LAI values in the MODIS dataset are based. Although it was found that the MODIS LAI dataset describes temporal variation of LAI reasonably well, it is suggested that the algorithm should be further improved to eliminate the algorithm’s double peak behavior during summer.
1. INTRODUCTION

This chapter provides a general introduction related to the context of this research. In section 1.1 an overview of the background of the topic is presented, followed by the definition of the problem that was tackled, described in section 2.2. Using this problem definition, a main objective was formulated in section 2.3, from which the main research question was deduced. This main research question and several additional research questions formed the basis of the research described in this report. To conclude this introduction chapter, an outline of the report is given in section 2.4.

1.1. Background

Numerous climate change models have been developed in order to simulate and model physical processes that take place on the Earth’s surface and in its atmosphere (Gholz et al., 1991). The core of these climate change models, often referred to as General Circulation Models (GCM’s), describes climate dynamics by using mathematical equations. The core GCM can be extended with various components such as land surface models, to produce increasingly complex models (Solomon et al., 2007). The quality of the output of these complex models depends not only on the internal physics of the model, but also heavily on the quality of the models’ input variables (Bounoua et al., 2006; Forest et al., 2008). To achieve good quality, the input variables for land surface models must be collected frequently for a long period of time (Knyazikhin et al., 1998; Wang et al., 2004).

One of the input variables that is used in many land surface models is the Leaf Area Index (LAI) (Liu et al., 2008; Tian et al., 2003). The LAI is one of the basic quantities used in estimation of net primary production and modeling of terrestrial carbon cycle processes (Chason et al., 1991; Fuentes et al., 2008; Sellers et al., 1986). The LAI is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as half the total needle surface area per unit ground area in coniferous canopies (Chen et al., 1997; Fuentes et al., 2008; Sellers et al., 2008; Sellers et al., 1997; Wang et al., 2004; Yang et al., 2006a; Yang et al., 2006b).

The LAI can be measured using direct and indirect methods (Chen et al., 1997; Fuentes et al., 2008). The direct methods measures the actual leaf area of the canopy in relation to the ground surface it covers (Daughtry and Hollinger, 1984). Although this method can yield a high accuracy, it is a time consuming and destructive way of measuring this variable (Fuentes et al., 2008). Furthermore, this method does not easily allow representation of LAI at a regional scale in terms of spatial and temporal variation (Chason et al., 1991). The field alternative to this direct method is to estimate the LAI indirectly via field optical sensors. Several methods have been developed based on the principle of measuring light intensity underneath the canopy in order to calculate the probability of light passing un-intercepted through the plant canopy, the so called gap fraction methods (Chason et al., 1991; Chen et al., 1997). Although these methods have proven to produce accurate LAI estimates on patch scale, it still remains a time consuming effort if large regions have to be covered.

Since the introduction of Earth observation satellites it has become possible to measure the spectral reflectance of the vegetation on a large scale and to relate red/near-infrared ratios to
parameters such as the LAI. Because satellite systems can produce data at a large extent, with a frequent revisit time and with a spatially continuous representation of the land surface, this method is nowadays preferred over the direct and gap fraction methods (Li et al., 2005). However, because the derived LAI values are only estimates of the physical variable, it is important to ensure that the algorithm used to calculate the LAI from spectral reflectance is as accurate as possible (Yang et al., 2006c).

Currently, several governmental and commercial organizations are developing Earth observation satellite systems. United States’ governmental organization NASA (National Aeronautics and Space Administration), together with the national aerospace agencies of Canada and Japan, initiated a project in which a coordinated series of polar-orbiting satellites was launched into space for long-term global observations of the Earth’s land surface (WWW1, 2009). In late 1999, the Terra platform was launched with several Earth observation sensor systems on board. One of these sensors is MODIS (Moderate-resolution Imaging Spectroradiometer) which measures the spectral reflectance in 36 spectral bands. Another operational MODIS sensor became available when the Aqua platform was launched in 2002.

Since the MODIS LAI products became available in 2000, they were subjected to extensive validation and analysis procedures (Baret et al., 2006; Cohen et al., 2003; Wang et al., 2004; Yang et al., 2006c). Although the radiative transfer algorithm used in MODIS products produced accurate LAI estimates for many regions on Earth, LAI values for complex biomes like boreal coniferous forests were proven to deviate from field data (Yang et al., 2006a; Yang et al., 2006b). The canopy structure of these boreal coniferous forests consists of many needle layers, making it difficult for the radiative transfer algorithm to simulate the radiation transfer inside the canopy. Because these forests are mostly situated in the northern part of the northern hemisphere, also cloud and snow coverage form an obstacle for the algorithm to produce reliable output. However, based on validation results the quality of the algorithm has been improved several times. After each refining of the algorithm, the complete dataset was reprocessed which resulted in the currently available Collection 5 dataset with 1 km resolution for the entire planet, containing all LAI data from 2002 until the present. This dataset is not only based on spectral reflectance of the Terra-MODIS sensor, but also on the data of the Aqua-MODIS sensor. This yields a 10% - 20% higher accuracy for coniferous forests when compared to only using the Terra-MODIS data as input (Yang et al., 2006b).

1.2. Problem definition

Until now, most analyses of MODIS LAI data were based on site or biome specific average values and the changes of these averages over time. Current climate modeling uses the LAI average values and their standard deviations as inputs (Blümel and Reimer, 2009; Martin, 1998). However, additional information could be derived from the probability density function of LAI values within a biome. Having this probability density function, which gives the relative number of pixels assigned to each LAI value, could be beneficial for climate and land surface modelers, as it provides besides the mean also additional
information about extreme and frequently occurring values (Alexander et al., 2006; Katz and Brown, 1992; Perkins et al., 2007). Not only the probability density function itself can provide valuable information, but also the temporal variability and thus the stability of the probability density function over time can help climate modelers develop their models.

Because boreal and temperate forests form a crucial element in the biosphere, as their carbon storage has a profound impact on the carbon cycle (Kellomäki and Wang, 2000), it is particularly important to have accurate information about LAI values for these biomes and their variation over time. However, the probability density functions of LAI values for boreal and temperate forest regions and therefore its variation over time remains unknown. When these probability density functions of the LAI are available, climate and land surface modelers can test the benefits of using a PDF compared to LAI mean and standard deviation.

1.3. Research Objective & Questions

Objective:

1. To determine the temporal variation in the Probability Density Functions (PDF’s) of the Leaf Area Index (LAI) for boreal and temperate forests by analyzing MODIS Collection 5 LAI datasets (2003-2008).

Main research question:

1. What is the temporal variation in the Probability Density Functions of the Leaf Area Index for boreal and temperate forests between 2003 and 2008?

Additional research questions:

1. What is the effect on a PDF if not only LAI values calculated by the radiative transfer algorithm are used, but if also LAI values calculated by the backup algorithm are used?
2. What are the differences in seasonal variation of the PDF’s between years?
3. What are the differences in seasonal variation between different biomes?
4. What is the contribution of different vegetation types to a PDF?

1.4. Thesis outline

In the following chapters, the research project and its results will be explained and discussed. Chapter two describes the different datasets that were used to provide the input data. This is followed by chapter three which describes the methodology used. Subsequently, an overview of the results is presented in Chapter 4. In Chapter 5, these results are interpreted and discussed. Finally, in Chapter 6, an overview is given of the main conclusions together with recommendations for future research on this topic.
2. MATERIALS

In this research four datasets were used: a global biome dataset to define the study areas, a global LAI dataset to provide source data for the PDF’s, a land cover dataset to analyze the effect of land cover on a PDF, and a tree cover dataset to make sure that only forest pixels were selected. In sections 2.1 to 2.4 each of these datasets is described in terms of its background and spatial properties.

2.1. Biome product

For many years, the spatial distribution of vegetation was studied by ecologists in order to define global vegetation patterns and zones. In 1939, ecologists F.E. Clements and V.E. Shelford introduced the concept of biomes in order to describe these patterns (Smith and Smith, 2001). The terms biome and land cover are often mixed and used interchangeably. Land cover classification is often only related to vegetation type, while biome classification can be also based on geographic location, fauna and climate. The biomes of Clements and Shelford were characterized by a uniform life form of vegetation, such as tropical forest or tundra. However, because boundaries between biomes are continuous and indistinct, several more detailed classification schemes have been developed. The Holdridge Life Zone System (Holdridge, 1967) involves three levels of classification, all based on climatological parameters like temperature and humidity. A more recent approach of classification is the concept of ecoregions, largely developed by Bailey (1978). It subdivides terrestrial and oceanic ecosystems based on the interaction between climate, soils and topography (Bailey and Hogg, 1986).

Recently, several institutes and organizations developed datasets containing the spatial extent of global biomes and ecoregions. The most widely used datasets are the datasets of the United Nations Environmental Program (UNEP), the Forest Service of the United States Department of Agriculture (FS-USDA), and the World Wide Fund for nature (WWF). All provide global biome data, but use different classification schemes and projection systems. As the biome and ecoregion dataset of the WWF is currently widely used and contains the best accessible metadata, it was selected for use in this research.

The ecoregions and biomes defined by the WWF were initially based on the Pielou (1979) and the Udvardy et al. (1975) system of biogeographic regions. These regions have been modified and complemented with several other datasets in order to determine the extent of the biomes (Olson et al., 2001):

- The Dinerstein et al. (1995) and Ricketts et al. (1999) biome classification.
- Global maps of floristic or zoo geographic provinces.
- Global and regional maps of units based on the distribution of selected groups of plants and animals.
- The world’s biotic province maps
- Global maps of broad vegetation types.

Although ecoregions provide much more detailed information than biomes, this ecoregion subdivision is often not being used for large scale climate and land surface modeling, but for nature conservation policies (Olson et al., 2001). Instead of using ecoregions in climate and land surface modeling, often only a differentiation in vegetation type is used which is
represented sufficiently by a biome classification (Martin, 1998). Therefore only the spatial extent of the biomes will be used from the WWF Ecoregions database.

In the WWF Ecoregions dataset fourteen biomes (Table 1), called Major Habitat Types, are defined. Three of these biomes were used in this research: temperate broadleaf and mixed forests, temperate coniferous forests and boreal forests.

The spatial extent of these three biomes is shown in Figure 1. Sections 2.1.1 to 2.1.3 describe the spatial and climatological properties of these three biomes.

Table 1: Biomes (Major Habitat Types) defined by the WWF.

<table>
<thead>
<tr>
<th>Biome</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ Tundra</td>
</tr>
<tr>
<td>♦ Boreal forests/taiga</td>
</tr>
<tr>
<td>♦ Temperate coniferous forests</td>
</tr>
<tr>
<td>♦ Temperate broadleaf and mixed forests</td>
</tr>
<tr>
<td>♦ Temperate grasslands, savannah, and shrublands</td>
</tr>
<tr>
<td>♦ Mediterranean forests, woodlands, and shrub</td>
</tr>
<tr>
<td>♦ Tropical and subtropical coniferous forests</td>
</tr>
<tr>
<td>♦ Tropical and subtropical moist broadleaf forests</td>
</tr>
<tr>
<td>♦ Tropical and subtropical dry broadleaf forests</td>
</tr>
<tr>
<td>♦ Tropical and subtropical grasslands, savannah, and shrublands</td>
</tr>
<tr>
<td>♦ Montane grasslands and shrublands</td>
</tr>
<tr>
<td>♦ Deserts and xeric shrublands</td>
</tr>
<tr>
<td>♦ Mangrove</td>
</tr>
<tr>
<td>♦ Flooded grasslands and savannah</td>
</tr>
</tbody>
</table>

Figure 1: Spatial extent of three biomes defined by the WWF.
2.1.1. Boreal coniferous forest

The boreal coniferous forest (also called Taiga) is the largest of the three biomes, covering more than 15 million km² of the Earth’s surface (11% of land surface). The forests are located between 50° and 65° N latitude, covering large parts of Scandinavia, Russia, Alaska and central Canada (Figure 1). They consist mainly of evergreen coniferous trees like spruce (Picea), fir (Abies) and pine (Pinus). Also, near the south of the biome, small leaved deciduous tree occur: larch (Larix) or tamarack (Larix). The continental climate is characterized by a high temperature range, with temperatures varying from −54°C to 27°C throughout the year. The summer period, and thus the growing season, is short and also relatively wet because precipitation (200-600 mm annually) occurs mainly during summer as rain, and evapotranspiration rates remain low (Molles, 2010).

2.1.2. Temperate coniferous forest

Compared to the boreal coniferous forests, the temperate coniferous forests are situated in a less extreme environment. The winters are less cold and the annual precipitation is higher (>2000 mm). However, this precipitation is not evenly distributed over the year, causing droughts during the summer months (Molles, 2010). The temperate coniferous forests cover 4 million km² of the Earth’s surface and are mostly situated on the west coast of North America and in Asia (Figure 1).

2.1.3. Temperate broadleaf and mixed forest

The temperate broadleaf and mixed forests consist of deciduous trees like oak (Quercus), beech (Fagus), maple (Acer) and birch (Betula) and coniferous trees like pine (Pinus), fir (Abies) and spruce (Picea). The biome covers almost 13 million km² and covers large parts of Europe, North America and eastern Asia (Figure 1). Because the climate is moderate, the precipitation (500-1200 mm annually) is evenly distributed throughout the year (Molles, 2010). Because winters are not as long and cold as the on higher latitudes situated boreal forests, the growing season can last up to six months.

2.2. Leaf Area Index product

Several institutes have created global LAI datasets with the purpose to serve as input for climate and land surface models. The following datasets were assessed for suitability for this research:

1. EcoClimap, based on AVHHR data.
2. GlobCarbon, based on SPOT/VEGETATION and ATSR/AATSR data.
3. CYCLOPES based on SPOT/VEGETATION data.
4. MODIS LAI, based on the Terra/MODIS and Aqua/MODIS data.
5. MERIS LAI, based on ENVISAT/MERIS data.
In order to select the most suitable dataset, it was necessary to define several requirements that the dataset should fulfill:

1. **Temporal resolution:** As the main objective of this research is to analyze the temporal variation of the PDF of LAI, the LAI dataset should have a high temporal coverage. Because vegetation changes fast during growing season, the LAI dataset should have preferably a temporal resolution of less than one month.

2. **Spatial resolution:** Because the spatial extent of biomes is relatively large (millions of km$^2$), a spatial resolution of 1 km is sufficient.

3. **Costs:** The LAI dataset should be available at no cost.

4. **Accessibility:** The dataset and its metadata should be accessible via the internet, preferably downloadable from an FTP server.

Although it is suggested by Garrigues et al. (2008) that the CYPLOPES product corresponds best with ground truth data, the MODIS LAI product is selected to be used in this research because it is the only dataset which meets the accessibility and costs requirements.

The MODIS LAI dataset is based on reflectance data of the MODIS (Moderate-resolution Imaging Spectroradiometer) sensors on board the Terra and Aqua platforms. The sensors measure the spectral reflectance in 36 different bands, 7 of which are used by the main algorithm to compute LAI estimates (Tian et al., 2000). This main algorithm is based on the principle of Radiative Transfer (RT) and was specially developed to work with the spectral and spatial properties of the MODIS reflectance data. The RT-algorithm simulates the reflectance of the 3-dimensional canopy structure using several inputs: canopy parameters (including LAI), sun and view directions, band uncertainties, and eight biome land cover classes (Section 2.3) (Tian et al., 2000; Yang et al., 2006a). Subsequently, an inverse modeling technique is applied to the radiative transfer algorithm to calculate the LAI values. In some cases, the RT-algorithm is not able to estimate the LAI properly; this happens when the combination of spectral reflectance data and other input parameters does not contain enough information for the RT-algorithm to generate one unique output. In this case many outputs are generated over a wide range of LAI values, all having an equal probability of occurrence. When this happens, the retrieved LAI is said to be part of the saturation domain, indicating that the retrieved LAI value is less accurate (Tian et al., 2000). In case the RT algorithm completely fails to generate an output, a back-up algorithm is used to calculate the LAI values instead. This backup algorithm uses biome-specific empirical relationships between the Normalized Difference Vegetation Index (NDVI) and LAI, but is proven to be less reliable than the RT algorithm (Yang et al., 2006c).

During the period 1999 to 2002, only the Terra platform (launched in 1999) was operational, resulting in only one available MODIS LAI dataset. However, as of 2002, a second MODIS sensor became operational shortly after the launch of the Aqua platform. From 2002 onwards, two independent reflectance datasets were available, resulting in two independent LAI datasets. Research indicated that combining the two MODIS datasets could yield more accurate LAI estimates (Yang et al., 2006b). For this reason, an additional LAI dataset was created based on both MODIS sensors. This combined dataset (MODIS MCD15A2) is used as LAI source data for the calculation of the PDF’s in this research.
The MODIS LAI dataset, produced by the MODIS Land Science team, is composited every 8 days at 1-kilometer resolution on a sinusoidal grid. The latest Collection 5 version of the LAI dataset is downloadable from NASA’s Land Process Distributed Active Archive Center (LP DAAC) in tiled HDF-format or in monthly global composites made available by the Climate and Vegetation Research Group of the Department of Geography of the Boston University. Because the dataset of the Boston University already contains a global mosaic of the LAI values, it was used as the main data source in this research.

Each monthly global LAI composite is accompanied by a Quality Control (QC) layer. For each pixel, the QC layer describes the number of 8-day periods on which it is based upon. The QC-values 1 to 4 indicate that the pixel is based on the unsaturated output of the main algorithm and computed by averaging 1 up to 4 eight-day periods (Table 2). The higher the value, the higher the quality is. The QC-values 5 to 8 represent pixels which are based on the output of the backup algorithm or on the saturated output of the main algorithm. A higher value indicates a higher quality.

Table 2: Description QC-values.

<table>
<thead>
<tr>
<th>Value:</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Main algorithm, unsaturated output, Average of 1 eight-day period</td>
</tr>
<tr>
<td>2</td>
<td>Main algorithm, unsaturated output, Average of 2 eight-day periods</td>
</tr>
<tr>
<td>3</td>
<td>Main algorithm, unsaturated output, Average of 3 eight-day periods</td>
</tr>
<tr>
<td>4</td>
<td>Main algorithm, unsaturated output, Average of 4 eight-day periods</td>
</tr>
<tr>
<td>5</td>
<td>Backup algorithm &amp; saturated main algorithm output, Average of 1 eight-day period</td>
</tr>
<tr>
<td>6</td>
<td>Backup algorithm &amp; saturated main algorithm output, Average of 2 eight-day periods</td>
</tr>
<tr>
<td>7</td>
<td>Backup algorithm &amp; saturated main algorithm output, Average of 3 eight-day periods</td>
</tr>
<tr>
<td>8</td>
<td>Backup algorithm &amp; saturated main algorithm output, Average of 4 eight-day periods</td>
</tr>
</tbody>
</table>

2.3. Land cover product

One of the inputs of the RT algorithm is a land cover product (MOD12Q1, Collection 4). This land cover product, derived from MODIS data, contains a classification scheme with 8 distinct classes (Table 3). The product is based on one full year of MODIS reflectance data.

Table 3: MODIS land cover classes used by RT-algorithm.

<table>
<thead>
<tr>
<th>Class</th>
<th>Class name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Grasses/Cereal Crops</td>
</tr>
<tr>
<td>2</td>
<td>Shrubs</td>
</tr>
<tr>
<td>3</td>
<td>Broadleaf Crops</td>
</tr>
<tr>
<td>4</td>
<td>Savannah</td>
</tr>
<tr>
<td>5</td>
<td>Broadleaf Forest</td>
</tr>
<tr>
<td>6</td>
<td>Coniferous Forest</td>
</tr>
<tr>
<td>7</td>
<td>Unvegetated</td>
</tr>
<tr>
<td>8</td>
<td>Urban</td>
</tr>
</tbody>
</table>
Because land cover is assumed to be relatively static compared to the LAI values, only one year, 2001, was used in the analysis. The data is available in 500 meter resolution in a sinusoidal projection. The global composite was made available by the Department of Geography of the University of Boston, USA.

2.4. Tree cover product

In order to make sure that within each biome only forest pixels were selected, a tree cover product was used. The MODIS Vegetation Continuous Fields (VCF) product (MODIS MOD44B) contains such a tree cover estimate and was already available as a yearly composite in sinusoidal projection with 500 meter spatial resolution. This product has some additional value compared to the land cover product, as it not only indicates where forest is located, but also in what quantity. This VCF product, which is based on Terra/MODIS data, contains three data layers: tree cover (sometimes the broader term ‘woody vegetation’ is used), herbaceous vegetation and bare soil. Each layer contains a number which represents the percentage of coverage by that vegetation type. For each pixel the sum of the percentages in the three layers adds up to hundred percent. As this research only focused on forest, only the tree cover layer was used for the analyses.

The VCF product could be downloaded in tile format via the LP DAAC or via the ftp-server of the University of Maryland, USA, which also composited the 2001 Collection 3 data into continental tiles (Goode Homolosine projection). These 2001 Collection 3 continental tiles were used in this research for further analyses.
3. METHODOLOGY

In order to create and analyze the PDF’s, several data preprocessing steps were necessary. Section 3.1 describes these data preprocessing steps for each dataset. Subsequently, in section 3.2 to 3.5, the methodology of the analyses is explained in detail.

3.1. Data preprocessing

The data selected for this research (Chapter 2) was acquired in different spatial resolutions, data types (raster/vector) and different projection systems. For each type of analysis, the datasets first had to be converted to one format. Because the spatial extent of the biome and tree cover datasets were assumed to be static over the period 2003-2008, they consisted of only one timeframe. The LAI dataset however, consisted of monthly composites. Multiplying the 12 months of the year by the 6 years the dataset covers, results in a considerable larger dataset for LAI than for both the biome and the tree cover dataset. For time efficiency reasons, the biome and tree cover dataset were converted to match the spatial properties of the LAI dataset. This meant that several preprocessing steps had to be done (Figure 2).

**Biome dataset:**

The WWF ecoregion dataset and the LAI dataset were both obtained in a different coordinate reference system, respectively in a WGS 1984 geographic coordinate system and a projected sinusoidal reference system. Since overlaying datasets in different coordinate reference systems is not possible, the ecoregion dataset was reprojected to a world sinusoidal projection system (operation 1, Figure 2). Reprojecting the dataset resulted in an offset between coordinates of 20 to 40 km depending on the longitude and latitude of the pixel. This problem was mostly solved by georeferencing the ecoregions dataset to the LAI dataset. First, the ecoregion dataset was converted into raster format. To do so, first, all ecoregions were dissolved into biome polygons (operation 2). These dissolved biome polygons were rasterized, setting the cell size to 1000 meters and the assignment type to ‘cell center’ (operation 3). This last setting meant that the new raster cell would get the attribute value of the polygon that covers the center of the new raster cell. In order to correct for the offset in pixel location, which was the result of the reprojection procedure, the biome dataset was georeferenced to the LAI dataset. This was done by selecting 12 control points, mostly situated near the edges of continents and islands, evenly spread over the Earth’s surface (operation 4). A 3rd order polynomial transformation was selected to rectify the image. This resulted in a total RMS error of 864 meter. Given the fact that the pixel size is 1000 meter this was considered to be acceptable.

**Tree cover dataset:**

The MODIS VCF product, containing the tree cover (TC) dataset, was obtained in Goode's Homolosine projection system. Therefore, after mosaicing the continental tiles (operation 5) into one global dataset, the dataset had to be reprojected to a world sinusoidal projection (operation 6). In order to match the spatial resolution of the LAI dataset, the spatial resolution of the TC dataset was first geometrically transformed to a 1000-meter cell size
(operation 7). This was done by using a bilinear interpolation method. This method uses the value of the four nearest input cell centers to determine the value on the output raster. The new value for the output cell is a weighted average of these four values, adjusted to account for their distance from the center of the output cell. Because similar problems as with the biome dataset occurred after reprojecting to sinusoidal projection system, also this dataset was georeferenced to the LAI dataset using 12 control points resulting in a RMS error of 799 meter (operation 8).

![Data processing chain](image)

Figure 2: Data processing chain: Operations in green indicate preprocessing. Operations in red indicate selecting LAI subset.
3.2. Biome specific temporal analysis

After all datasets had been standardized to one data format, several conditional operations have been carried out. In order to produce biome specific PDF’s, the biome dataset was used to select only pixels from the monthly LAI data subsets which fulfilled the conditional statement (operation 9) ‘biome = biome#’, where biome# represents the number of the biome in the dataset. Subsequently, a similar conditional statement was used to select only LAI pixels that fulfilled the QC requirement; All PDF’s in the temporal analysis are based on main algorithm output so therefore the statement ‘QC ≤ 4’ was used in operation 10 (Table 2). Because only pixels that represent forests had to be selected, a third conditional statement ‘TC ≥ 40’ was used (operation 11). This eliminated all LAI pixels with less than 40% tree cover, so only LAI values of relatively dense forests were used. These operations resulted in a selection of LAI pixels from which the LAI values from the attribute table were exported to a database file (operation 12). This database file could directly be used to create monthly PDF’s (operation 13). For each biome these PDF’s were visualized in a 3D surface chart. This surface chart interpolates the 12 monthly PDF’s into a smooth surface for visual analysis.

3.3. Tree cover analysis

One timeframe of the LAI dataset was used for selecting data to analyze the effect of tree cover on PDF’s: July 2003. This analysis has been carried out three times, once for each of the three biomes. Again the QC layer was used in a conditional statement (operation 10) to select only high quality pixels: ‘QC ≤ 4’. Using this selection of pixels, a batch of conditional statements (tree cover ≥ 0% AND Tree cover < 10%, ... tree cover ≥ 90% AND tree cover < 100%) was used in operation 11 to create 10 different scenarios for the July 2003 timeframe, each scenario being based on a distinct slice covering 10% of the total tree cover range (0%-100%). The 10 different selections of LAI pixels were used to generate 10 different PDF’s (operation 12 & 13). Again a 3D surface chart was produced to visualize the results.

3.4. LAI algorithm analysis

Also in this analysis only one timeframe was used: July 2003. Operation 9 was set to select only boreal forest pixels. To analyze the difference between the two MODIS algorithms, two different conditional statements were used in operation 10: ‘QC ≤ 4’ to select only output of the main algorithm and ‘QC ≥ 5’ to select only output of the backup algorithm. The TC dataset was used in operation 11 to select only LAI pixels having a tree cover of 40% or higher. The two resulting PDF’s were plotted to visually assess the differences between the algorithm’s outputs.

3.5. Land cover analysis

The radiative transfer algorithm uses not only the raw spectral reflectance data of the MODIS sensor as input, but also a preprocessed land cover dataset (MOD12Q1). For each land cover class, the algorithm uses different Look-Up Tables (LUT’s) to calculate the appropriate LAI estimate. These LUT’s contain vegetation type specific properties such as canopy transmittance, reflectance, and absorptance.
Because the spatial extent of a biome defined by the WWF contains multiple MODIS land cover classes, it is important to know the relative contribution of each class to the PDF. An analysis was made of the differences in PDF if only pixels belonging to one specific MODIS land cover class were selected. This analysis was carried out for July 2003 for all three biomes. Only high quality pixels were selected in operation 10 (Figure 2) and only pixels with more than 40% tree cover were selected in operation 11. Subsequently an additional statement was included, operation 14, to select from the MODIS Land Cover product one single land cover class. For each biome, PDF’s of the four biggest classes were plotted in a graph to visually assess the differences.
4. RESULTS

4.1. Biome specific temporal analysis

For three biomes, a temporal analysis of the LAI was performed. The results of these three analyses are presented in sections 4.1.1 to 4.1.3.

4.1.1. Boreal coniferous forest

Except for the months June, July and August, all available LAI data contain missing values above a latitude of around 63° North. The latitude above which these missing values occur, shifts during the winter period because of the changing tilt of the Earth and thus the changing illumination conditions. This considerably reduces the number of usable LAI pixels. Figure 3 shows the spatial distribution of high quality and low quality pixels for the boreal region for three different months in 2003.

The number of high quality pixels is at a low of 4% in January. During the year, this number increases gradually to 96% in August, and then decreases from August to December to 4% again. Not only the number of high quality pixels is very low during winter, but also the spatial distribution is poor. The available high quality pixels in January are all situated in the most southern parts of the biome. Because using low quality and poorly distributed LAI values would not yield a PDF representative for the whole biome, only summer LAI values were used. For all years, the months June, July and August contain around 96% high quality pixels and have a relative homogeneous spatial distribution of these high quality LAI pixels.
Figure 4 shows for each month (June-August) the average PDF based on all available years (2003-2008). The peak around LAI = 2.2 (first peak) shifts slightly during the summer season reaching its highest LAI value in July. At LAI = 4.3 the PDF has a dip almost to P = 0 for all three months. This dip is followed by a second peak feature which is present around LAI = 5.2 and is most recognizable in July.

Figure 4: Monthly average PDF of LAI for boreal forest.

In order to compare differences between years, these summer months are combined into one single average PDF per year. Figure 5 shows these yearly summer average PDF's. Again, there is a dip at LAI = 4.3. Only the heights of the first and second peak vary.

Figure 5: Yearly summer average PDF of LAI for boreal forest.

Analyzing statistical measures like the mean, standard deviation, mode and median shows no indication that the PDF shifts towards higher or lower LAI values during the period 2003 to 2008. There are only minor fluctuations for which the mode and median show different patterns from the mean and standard deviation (Figure 6); between 2004 and 2005 the mode and the median increase, whereas the mean and the standard deviation decrease.
Figure 6: Statistical measures characterizing LAI values for boreal forest during summer.

All summer months can be averaged to one PDF which represents the average summer PDF for the period 2003-2008 for boreal forests. This PDF, presented in Figure 7, shows a double peaked feature, the first peak concentrating 93.6% of the probability.

Figure 7: Multiyear average PDF of LAI for boreal forests based on summer periods.

4.1.2. Temperate coniferous forest

As temperate forests are situated at lower latitudes than boreal forests, remote sensing systems have less difficulty to produce spectral reflectance data. For this reason, the MODIS LAI dataset contains homogeneously distributed high quality LAI data during the entire year for the temperate forest biomes. As the interannual difference is found to be minimal, a full year analysis is only done for the first complete year (2003, Figure 8) and the last complete year (2008, Figure 9) of the LAI dataset.
During the first months of 2003, the PDF shows a relatively high probability of pixels having LAI values less than 1 (Figure 8). As the year progresses, the probability-peak gets wider with the mode being at its maximum at LAI = 2.4 in July. Also during summer months there is a slightly higher probability for the occurrence of high LAI values (> 6), although a distinct second peak, which was the case for the boreal forest, is less pronounced. After August the probability-peak shifts back to lower LAI values again.

Figure 8: Probability distribution of LAI for temperate coniferous forest in 2003.

The general shapes of the 2008 PDF’s are similar to those of 2003 (Figure 9). Although in particular the heights and shape of the probability-peaks of the winter months vary slightly.

Figure 9: Probability distribution of LAI for temperate coniferous forest in 2008.
Because a visual comparison of these kinds of graphs is difficult to perform, the average PDF of the entire year 2003 is compared to that of 2008. The result is shown in Figure 10 by the black dashed line. The yearly average difference for LAI values below 1.0 is larger than for LAI values above 1.0. The largest differences occur during winter months (December – February), shown by the green line. However, the differences at high LAI values (>4.0), occur mostly during summer (June – August, red line).

Figure 10: Differences in PDF between 2008 and 2003 for temperate coniferous forest.

4.1.3. Temperate broadleaf and mixed forest

Also for the temperate broadleaf and mixed forest, a full year analysis is done for 2003 and 2008. The probability-peaks in the winter months are higher for these broadleaf and mixed

Figure 11: Probability distribution of LAI for temperate broadleaf and mixed forests in 2003.
forests (Figure 11 & Figure 12) than for the temperate coniferous forest. From January to June, the probability peak becomes wider and lower and shifts towards higher LAI values (Appendix 1 and Figure 11). Also, during summer, a distinct second peak is visible at LAI = 5.4. From June to September the PDF remains stable. After September, the first peak shifts back to lower LAI values again and the second peak disappears.

![Image](image-url)

Figure 12: Probability distribution of LAI for temperate broadleaf and mixed forests in 2008.

Just like for the temperate coniferous forest, the general shape of the 2003 PDF’s for temperate broadleaf and mixed forest are similar to those of the 2008 PDF’s (Figure 12). The second peak feature remains present during the summer period and, from September onwards, the first peak gradually shifts back to lower LAI values again and the second peak disappears.

The differences in average PDF between 2003 and 2008 (Figure 13) again shows fluctuations in the LAI < 1.0 region. However, in particular during winter (December – February) there are large differences in this region. For LAI > 1.0, the yearly average difference in probability between 2003 and 2008 fluctuates between P = -0.0005 and 0.0009. The differences for LAI > 4.0 mainly occur during summer (June – August).
4.2. Tree cover analysis

The influence of tree cover on the PDF was analyzed for all three biomes. For each biome, the effect of selecting different percentages of tree cover is described (sections 4.2.1 to 4.2.3) and visualized in a graph (Figure 14 to Figure 17).

4.2.1. Boreal coniferous forest

At low percentages of tree cover (0-10%) there is only one peak around LAI = 1.2 (Figure 14). When percentage of tree cover increases, gradually a second peak appears, having a maximum probability of $P=0.01$ at 60-70% tree cover. Using a higher percentage of tree cover
yields a lower second peak and thus a higher first peak. Although the height of the first peak changes from \( P=0.08 \) at 0-10% tree cover to a lower probability of \( P=0.04 \) at 40-50% tree cover and back again to a higher probability of \( P=0.06 \) at 80-90% tree cover, the position of this first peak continues to shift gradually towards higher LAI values (Figure 15). Figure 16 indicates that the first 7 classes (0-10% to 60-70% tree cover) contain each between 10% and 16% of the total amount of pixels. Because the 90-100% tree cover class contains only 487 pixels compared to the overall class average of 1 million pixels this 90-100% tree cover class was not included in the analysis.

4.2.2. Temperate coniferous forest

Compared to the boreal forests, the temperate coniferous forest shows a relatively constant PDF if different percentages of tree cover are selected (Figure 17). The second peak feature, which is not as smooth as in the PDF’s of the boreal forest, increases only slightly when higher percentages of tree cover are selected. The average number of pixels per class for the temperate coniferous forest (230,000 pixels) is lower than the average number of pixels per class in the temperate broadleaf and mixed forest and the boreal forest (both \( \pm 1 \) million pixels). Also, in this biome there are no LAI values with a percentage of tree cover of less than 30%, which results in a null PDF for these tree cover classes (Figure 17 & Figure 18).
The pie chart in Figure 19 indicates that the 30-40% class contains very little pixel values. However, from the 40-50% to 70-80% class, pixels are equally divided between classes, all covering between 20 to 30% of the pixels. The 90-100% class contains no pixels at all.

Figure 17: PDF’s for temperate coniferous forest in July 2003 using different percentages of tree cover.

Figure 18: Contour view of PDF’s for temperate coniferous forest in July 2003 using different percentages of tree cover.

Figure 19: Relative distribution of pixels among the 10 tree cover classes for temperate coniferous forest.
4.2.3. Temperate broadleaf and mixed forest

As the percentage of tree cover increases in the temperate broadleaf and mixed forest (Figure 20), the second peak at LAI = 5.3 increases from $P = 0$ at 0-10% tree cover to $P = 0.02$ at 50-60% tree cover. So the PDF of broadleaf and mixed forest changes from a single peak feature PDF to a double peak feature PDF when tree cover increases. Selecting more than 50-60% tree cover does not yield to large changes in the PDF, as the PDF stabilizes at high per-

Figure 20: PDF’s for temperate broadleaf and mixed forest in July 2003 using different percentages of tree cover.

Figure 21: Contour view of PDF’s for temperate broadleaf and mixed forest in July 2003 using different percentages of tree cover.

Figure 22: Relative distribution of pixels among the 10 tree cover classes for temperate broadleaf and mixed forest.
percentages of tree cover. Only the position of the first peak shifts back towards lower LAI values at high percentages of tree cover (Figure 21). The first class, 0-10% tree cover, contains more than 25% of the total number of selected LAI pixels (Figure 22). Except this 0-10% class and the classes above 80% tree cover, all classes contain between 8% and 13% of the LAI pixels. Very little LAI pixels (0.7%) are classified as having a tree cover of 80% or more.

4.3. LAI algorithm analysis

The two algorithms (radiative transfer algorithm and backup algorithm), upon which the MODIS LAI dataset is based, both produce independent LAI estimates. It was found that the PDF’s based on the output of the backup algorithm (Ba) are different from the PDF’s based on the radiative transfer algorithm (RT) output (Figure 23). The number of backup retrieved LAI values is relatively low (18,952) compared to the number of LAI values retrieved from the radiative transfer algorithm (6,552.527).

![Figure 23: PDF of LAI for boreal forest in July 2003 using different algorithms.](image)

4.4. Land cover analysis

For July 2003, the final selection of pixels from the biome specific temporal analyses was analyzed using the MODIS land cover dataset. Sections 4.4.1 to 4.4.3 present the results of this analysis. Each pixel within the selected subset is classified in the land cover dataset. For each biome, four different PDF’s were produced based on the four biggest land cover classes present in that biome.
4.4.1. Boreal coniferous forest

Of the 6.5 million used LAI pixels in the boreal forest, 4.8 million are classified as coniferous by the MODIS land cover product (Table 4). Together with savannah and shrubs, the coniferous and broadleaf classes are the four largest classes. For each of these classes a PDF was made (Figure 24). The coniferous pixels show a higher first peak and a lower second peak than the relative small amount of broadleaf pixels. Pixels that represent broadleaf vegetation show the highest second peak. However, all four PDF’s have a second peak feature and a dip at LAI = 4.3.

Table 4: Number of pixels per MODIS land cover class for boreal forest in July 2003.

<table>
<thead>
<tr>
<th>Class name</th>
<th>Number of pixels</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>91611</td>
<td>1.41%</td>
</tr>
<tr>
<td>Grasses/Cereal Crops</td>
<td>104193</td>
<td>1.60%</td>
</tr>
<tr>
<td>Shrubs</td>
<td>385572</td>
<td>5.93%</td>
</tr>
<tr>
<td>Broadleaf Crops</td>
<td>18506</td>
<td>0.28%</td>
</tr>
<tr>
<td>Savannah</td>
<td>853207</td>
<td>13.12%</td>
</tr>
<tr>
<td>Broadleaf Forest</td>
<td>199965</td>
<td>3.07%</td>
</tr>
<tr>
<td>Coniferous Forest</td>
<td>4848750</td>
<td>74.56%</td>
</tr>
<tr>
<td>Unvegetated</td>
<td>935</td>
<td>0.01%</td>
</tr>
<tr>
<td>Urban</td>
<td>794</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

Figure 24: PDF of LAI for boreal forest in July 2003 for the four largest MODIS land cover classes.

4.4.2. Temperate coniferous forest

Sixteen percent of the pixels of the temperate coniferous forest defined by the WWF are classified as broadleaf pixels in the MODIS land cover dataset (Table 5). Again, the PDF based on broadleaf pixels has the most pronounced second peak, although in this biome it has an irregular shape (Figure 25). This second probability-peak between LAI = 4.3 and LAI = 6.7 is skewed, having its maximum probability at LAI = 6.6, where after the probability drops to 0.0007 at LAI = 6.7. Ten percent of the selected temperate coniferous forest pixels
are classified as savannah in the land cover dataset. The first peak in the PDF of these savannah pixels shows much resemblance to the PDF of coniferous pixels. However, the second peak has the same irregular shape as the PDF’s based on broadleaf vegetation and grasses. Pixels classified as grasses show a first peak located at a lower LAI value than broadleaf, coniferous or savannah based PDF’s.

Table 5: Number of pixels per MODIS land cover class for temperate coniferous forest in July 2003.

<table>
<thead>
<tr>
<th>Class name</th>
<th>Number of pixels</th>
<th>% of total:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>6600</td>
<td>0.53%</td>
</tr>
<tr>
<td>Grasses/Cereal Crops</td>
<td>48645</td>
<td>3.89%</td>
</tr>
<tr>
<td>Shrubs</td>
<td>45550</td>
<td>3.64%</td>
</tr>
<tr>
<td>Broadleaf Crops</td>
<td>22921</td>
<td>1.83%</td>
</tr>
<tr>
<td>Savannah</td>
<td>128692</td>
<td>10.28%</td>
</tr>
<tr>
<td>Broadleaf Forest</td>
<td>198974</td>
<td>15.89%</td>
</tr>
<tr>
<td>Coniferous Forest</td>
<td>799472</td>
<td>63.86%</td>
</tr>
<tr>
<td>Unvegetated</td>
<td>821</td>
<td>0.07%</td>
</tr>
<tr>
<td>Urban</td>
<td>235</td>
<td>0.02%</td>
</tr>
</tbody>
</table>

Figure 25: PDF of LAI temperate coniferous forest in July 2003 for the four largest MODIS land cover classes.

4.4.3. Temperate broadleaf and mixed forest

As the name of the biome already suggests, the temperate broadleaf and mixed forest biome contains a considerable amount of both broadleaf pixels (2 million) and coniferous pixels (1.4 million) (Table 6). Like the boreal coniferous forest and the temperate coniferous forest, the pixels classified as broadleaf pixels in the MODIS land cover product have the highest second peak in the PDF (Figure 26). Also the mode of the first peak in the coniferous based PDF is higher than in the PDF based on the broadleaf pixels. Almost 25 percent of all
selected pixels are classified as non-forest by the land cover product (savannah and grasses being the largest non-forest classes). Also these classes produce PDF’s with a double peak feature and a dip in between. The dips in the PDF’s of the four largest classes are not located at exactly the same LAI value, but vary between LAI = 4.2 and LAI = 4.4. Also, all PDF’s show a sudden drop in probability at LAI values around LAI = 6.6 (Figure 26).

Table 6: Number of pixels per MODIS land cover class for temperate broadleaf and mixed forest in July 2003.

<table>
<thead>
<tr>
<th>Class name</th>
<th>Number of pixels</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>39946</td>
<td>0.88%</td>
</tr>
<tr>
<td>Grasses/Cereal Crops</td>
<td>414152</td>
<td>9.08%</td>
</tr>
<tr>
<td>Shrubs</td>
<td>71680</td>
<td>1.57%</td>
</tr>
<tr>
<td>Broadleaf Crops</td>
<td>193326</td>
<td>4.24%</td>
</tr>
<tr>
<td>Savannah</td>
<td>391858</td>
<td>8.59%</td>
</tr>
<tr>
<td>Broadleaf Forest</td>
<td>2032754</td>
<td>44.57%</td>
</tr>
<tr>
<td>Coniferous Forest</td>
<td>1405661</td>
<td>30.82%</td>
</tr>
<tr>
<td>Unvegetated</td>
<td>1119</td>
<td>0.02%</td>
</tr>
<tr>
<td>Urban</td>
<td>9815</td>
<td>0.22%</td>
</tr>
</tbody>
</table>

Figure 26: PDF of LAI temperate broadleaf and mixed forest in July 2003 for the four largest MODIS land cover classes.
5. DISCUSSION

This chapter gives an overview of some critical notes that should be taken into account when interpreting the final results. Section 5.1 describes some important accuracy aspects of the datasets, followed by section 5.2, which elaborates on the methodology. In section 5.3 an interpretation of the PDF’s is made.

5.1. Selected source data

Four different datasets were used to create the PDF’s of the LAI. The quality of these data products could have had an impact on the quality of the final PDF’s. Therefore, in the following paragraphs each dataset is described in terms of its quality and usability in this research.

**Biome dataset**

The biome dataset acted as a definition of the spatial extent of the study areas. Because in natural situations biomes do not have distinct, well-delineated borders, it is debatable where these biome borders are located. A biome dataset producer could easily shift the border between the boreal and temperate broadleaf and mixed forest biome with several kilometers without being necessarily inaccurate. It is a matter of definition. The WWF ecoregions dataset, which contain the biome extents, is mostly focused on forest preservation and wildlife habitats. Compared to a producer who is primarily focused on, for example, climatological aspects, this can affect the biome dataset in terms of spatial extent and number of biome classes. For this research the WWF dataset was chosen over others because it contains good metadata and background information. Choosing a different biome dataset could possibly affect the PDF’s slightly, but is not considered to be that relevant for this study because of the very high number of pixels used for building the PDF’s.

**LAI dataset**

Only one LAI dataset was found to be suitable for this research, the MODIS LAI dataset. This was mainly caused by accessibility and cost restrictions (see section 2.2). However, Garrigues et al. (2008) indicates that other LAI datasets (e.g. CYCLOPES) correspond better with ground truth data. This could mean that using the MODIS LAI dataset does not yield the best possible PDF’s. It should however be taken into account that comparing ground truth data with remote sensing outputs often implies that datasets with different spatial resolutions have to be compared. Ground truth data is often measured in plots of several meters (Chen et al., 1997; Fuentes et al., 2008), while the MODIS LAI data is acquired at a 1000-meter resolution. This induces upscaling difficulties when comparing the higher resolution ground truth data to the lower resolution remote sensing data.

Because of the limitations of passive remote sensing devices, the spatial coverage during winters has proven to be insufficient, especially in high latitudes. This has several causes: the inclination of solar radiation with the Earth surface in that region is very low, resulting in low amounts of solar radiation that can be reflected by vegetation. Also, during winter, cloud and snow coverage form an obstacle for the remote sensing instrument to measure spectral reflectance from the vegetation. Another limitation of using remote sensing
techniques is that the physical parameter itself is not measured directly but only estimated using an algorithm. This can cause problems like algorithm inaccuracies or failures. Many of these possible problems however, are described in a quality control layer, which reduces the risk of using erroneous data.

Results of the land cover analysis of the broadleaf and mixed forest emphasizes another limitation of using the RT-algorithm; The range of possible LAI values is restricted to a highest LAI value of LAI = 7.0. In Figure 26, a sudden drop in probability at LAI values around LAI = 6.6 is visible, while in-situ LAI values might show a higher range in LAI.

**Land cover dataset**

Although the RT-algorithm that calculates the LAI uses the collection 5 land cover dataset as input, the previous collection 4 land cover dataset was used to create land cover type specific PDF’s. This might have caused small errors in land cover type specific PDF’s, as a number of pixels are classified differently by the newest collection 5 version of the dataset. However, because it is not likely that on a global scale many forest pixels were classified differently in the two collections, it is assumed that using the collection 4 land cover dataset does not result in large errors in land cover class specific PDF’s.

Because the MODIS land cover dataset is used as input by the RT algorithm, it is important to keep in mind that any error in the classification propagates to the LAI product. Like any MODIS product, the land cover product is accompanied by a quality control layer. Although the information in this layer was not used in analyses in this research, the land cover product was assumed to be reasonably accurate. This is confirmed by the research of Friedl et al. (2010), although Lopes et al. (2003) report many misclassified pixels in a study area in Portugal. Lopes’ results indicate that evergreen broadleaf forest is the most confused land cover class, as it represents several in situ land cover types such as evergreen coniferous forest, mixed forest, woody savannah, and closed shrubs.

**Tree cover dataset**

Also the TC dataset is based on MODIS reflectance data. The quality control layer of this dataset was not used in analyses in this research. Although Montesano et al. (2009) report that using pixels classified as having less than 20% tree cover should be avoided in pixel by pixel analyses in boreal forests, the dataset in general was assumed to be reasonably accurate on the scale and extent it was used in this research. However, inaccuracies in the TC dataset may have affected the selection of pixels that were used to compute the PDF’s.

### 5.2. Methodological aspects

The reprojection procedure which had to be performed for the biome and tree cover product resulted in a 20 to 40 km shift in pixels between these datasets and the LAI dataset. This could be partly rectified by a georeferencing procedure, but a small shift of several kilometers remained. This could mean that LAI pixels were selected that belonged to another biome than was intended. However, because biome borders have natural gradual
transitions, this small shift is assumed to have no significant impact on the biome specific PDF’s.

One of the goals of the methodology was to select only pixels which represent forest areas in a specific biome. In order to realize that, the MODIS tree cover product was used. Selecting pixels containing more than 40% of tree cover should have eliminated most pixels containing low amounts of biomass (like grasses, crops etc.). However, because the producers of the VCF product use the terms ‘tree cover’ and ‘woody vegetation’ interchangeably, possibly the initial intention to select only pixels covered by trees was not completely realized, as also other ‘woody vegetation’ might have been selected. Choosing 40% as a lower limit was based on the tree cover analysis results, which indicated that selecting more than 40% tree cover did not yield big changes in the PDF’s. However, after completing the biome specific temporal analysis, an additional land cover differentiation was made for July 2003 by using the MODIS land cover product. Although only pixels with more than 40% tree cover were selected, in all biomes still many pixels classified in the land cover product as savannah, grasses, or shrubs were present. If the land cover product would have been used for selecting specifically forest pixels in the biome specific temporal analysis, the resulting PDF’s would have had an even higher probability that they would represent only forest area. However, this would negatively affect the number of pixels on which the PDF would be based upon, resulting in less spatial coverage. It is believed that depending on the practical implementation of the PDF either method yields usable PDF’s.

5.3. Interpretation of PDF’s

The temporal analyses indicate that for the two temperate forest biomes there is a typical seasonal behavior of the vegetation. For the boreal forest biome, only an interannual comparison could be made based on the summer period. Because most trees increase their biomass during the growing season in the form of an increased number of leaves, the LAI increases to its highest value in July. For all biomes there is little variation in LAI between years. The small variations that are present mostly occur in the LAI range from 0 to 1, which means that vegetation with low canopy coverage changes more than dense vegetation in the period 2003 – 2008.

5.3.1. Double peak feature

All three biomes show during summer two distinct peaks in their PDF, a feature that has not been described by the producers of the algorithm, nor can be found in literature from researchers at NASA and the collaborating University of Boston. Not only the double peak feature seems to be persistent in the summer PDF’s of all three biomes, but also the dip between the two peaks around LAI = 4.3 is present in all PDF’s. During winter months, the double peaked PDF changes back to only one peak at low LAI values. To explain these distinct PDF features, several possibilities were explored.

The two peaks might represent different types of vegetation (trees, grasses, crops etc.). But as the second peak is only present during summer, this would mean that the second vegetation type is completely absent during winter. Also because the double peak feature persists in all three examined forest biomes, having all their own distinct types of vegetation,
it is unlikely that each peak could entirely be explained by the differences in two types of vegetation. The theory became even more unlikely after the MODIS land cover dataset was analyzed: the outcomes of that analysis suggested that the double peak feature is also present when only pixels are selected belonging to one MODIS land cover class (e.g. only savannah). Although it is discussible whether these savannah pixels really represent savannah vegetation, as they contain more than 40% tree cover and are located in the boreal and temperate forest biomes.

Eriksson et al. (2006) suggested that the presence of understory vegetation impacts the LAI of forests by an average of 1.6 LAI-units. Although the second peak is on average 3 LAI-units higher than the first peak, this understory effect might also (partially) explain the two distinct LAI peaks for the summer period. To investigate this theory further it is important to have an idea of the spatial distribution of the LAI pixels belonging to the two distinct peaks. Figure 27 shows the spatial distribution of pixels belonging to the first peak and to the second peak of the PDF for temperate broadleaf and mixed forest in Europe in August 2003. There seems to be no distinct pattern in spatial distribution, although pixels belonging to the second peak are situated more in the southern part of this biome. If this theory would explain the second peak, however, this would mean that the understory effect only takes place during summer, while in that season one would expect the highest leaf coverage of a tree, covering most of the understory vegetation. So, if an understory effect takes place, one would expect it to be before or after the growing season. This however, is not the case.

Figure 27: Spatial distribution of temperate broadleaf & mixed forest LAI values for August 2003 in Europe. Green pixels are part of the first peak. Red pixels are part of the second peak.
A third explanation for the double peak might be found in the LAI algorithms themselves. Although the backup algorithm’s outputs were not used in any PDF, the results of the algorithm analysis that has been done, indicated that there are large differences in PDF’s when they are based on different algorithms. These differences indicated that the basic principles used by the algorithms (radiative transfer versus empirical relationships), can profoundly impact their outcome. Because within each biome different land cover classes occur, and because these land cover classes are one of the inputs for the RT model, the difference between class specific PDF’s was investigated. In all three biomes, all major land cover classes show a double peak feature with a dip between the peaks around LAI = 4.3. Therefore it could be concluded that not only forest PDF’s show this behavior, but also grasses, shrubs and savannah. It is more likely that these typical behaviors are a result of the properties of the RT algorithm itself than that all these vegetation types have almost exactly the same LAI dip around LAI = 4.3. Although the only way to check whether the algorithm itself is responsible for the double peak feature is to perform an extensive validation procedure, it is assumed that the double peak feature is caused by the RT algorithm.

5.3.2. Sensitivity of the PDF to tree cover

The results from the tree cover analysis indicated that there is no LAI pixel having a tree cover less than 30% for the temperate coniferous forest. This results in the absence of PDF’s for the first three tree cover classes. This means that of the 1.2 million km², there is no square kilometer having less than 30% woody vegetation, or in other words, more than 70% grasses or shrubs. This could be the case, but the land cover product indicates that there are also 128,000 pixels classified as savannah situated within the biome. One would expect that at least some of these savannah pixels should have more than 70% of other vegetation than trees. Probably either the tree cover product or the land cover product contains erroneous data for the temperate coniferous forest biome.

The PDF for the temperate coniferous forest biome remains relatively stable when increasing percentages of tree cover are selected. This is not the case for the temperate broadleaf and mixed forest and the boreal forest biomes. These two biomes show a shift of the first peak towards higher LAI values if a higher percentage of tree cover is selected and also an increase in the height of the second peak. It is interesting to see that if more than 60% tree cover is selected in the boreal forest, the height of the second peak decreases. This is not what one would expect if the forest density, and thus the number of leaves, increases. More tree cover should mean more leaves per m² and thus a higher LAI. Probably a saturation effect occurs when canopy coverage is more than 60 or 70%. After that point, an increase in the number of leaves doesn’t yield an increase in reflected radiation. A similar paradox can be found in the temperate broadleaf and mixed forest where the first peak shifts back towards lower LAI values. This might be caused by the decreased number of pixels for the highest tree cover classes.
6. CONCLUSIONS & RECOMMENDATIONS

In sections 6.1 and 6.2, an overview is given of respectively the main conclusions of this research and some recommendations that can form a basis for further research on this topic.

6.1. Conclusions

For three selected biomes a temporal analysis of the LAI has been done. For the temperate forests, results indicate during winter a single peaked PDF, having its peak around LAI = 1. During the growing season, this peak shifts towards higher LAI values around LAI = 2.4. Also a second peak appears during summer around LAI = 5.2. This second peak is most pronounced for the temperate broadleaf and mixed forest. Because for the boreal biome the MODIS LAI dataset doesn’t contain data for the whole year, PDF’s were only created for the summer months (June, July and August). Also in this boreal biome a second peak around LAI = 5.2 appears in the summer PDF’s. Land cover analyses indicate that this second peak feature persists in all detected land cover classes within each biome. The second peak is most pronounced in the temperate broadleaf and mixed forest because pixels classified by the land cover product as broadleaf result in the highest second probability-peak. The double peaked PDF with the dip at LAI = 4.3 is possibly an artifact caused by the RT algorithm. Further research is necessary to investigate this problem.

Between months, but also between biomes, the height of the two probability-peaks changes, while the position of the second peak remains at similar LAI values. Between years, all biomes show stable PDF’s with little variation. Although the full year analyses of the temperate forest biomes indicate that the biggest difference occurs in the LAI range from 0 to 1, it remains limited to a maximum difference in probability of P = 0.003.

The shape of the PDF’s is found to be influenced by the selected percentage of tree cover and the used algorithm. The RT algorithm’s output and backup algorithm’s output differ profoundly. The change in PDF related to the tree cover is biome dependent. The boreal forest biome and the temperate broadleaf and mixed forest biome show an increase in the height of the second probability-peak and a shift of the first peak towards higher LAI values when higher percentages of tree cover are selected. This is caused by the increased number of leaves in the pixels that have a high percentage of tree cover and thus a higher LAI value. However, if pixels with more than 60-70% of tree cover are selected, the temperate broadleaf and mixed forest biome shows a shift of the first peak back to lower LAI values again. Possibly a saturation effect occurs when canopy coverage is higher than 60 or 70%. After that point, an increase in the number of leaves doesn’t yield an increase in reflected radiation and thus in the computed LAI. In comparison with the other two biomes, the temperate coniferous forest has a relative stable PDF if higher percentages of tree cover are selected.

All results show that PDF’s contain important LAI information that can be of additional value for climate and land surface modelers. PDF’s provide, besides information about seasonal changes in LAI, also distribution information like outliers and modes. Because it is often assumed that LAI is normally distributed, enabling the use of mean and standard deviations as model input, inaccuracies can occur in the model’s output. The usage of PDF’s
as model input could be beneficial for climate and land surface modelers, as they describe temporal changes in more detail than mean values and standard deviations.

6.2. Recommendations

1. **Improve RT algorithm:** Although the PDF’s derived from the MODIS LAI product show the expected seasonal behavior, the algorithm that is used to generate LAI values could be further improved. Temporal analyses of the MODIS LAI data indicate that, during summer, the RT algorithm produces PDF’s with a curious double peaked feature. This bimodality is probably caused by the algorithm but has to be further investigated.

2. **Validate PDF’s:** In order to validate the MODIS biome specific PDF’s, they could be compared to PDF’s derived from other monthly global LAI datasets. This would improve the quality of the PDF’s and could also provide a decisive answer whether the double peak during summer is algorithm related.

3. **Adapt climate and land surface models:** In order to use all information embedded in the PDF’s, climate and land surface models should be made suitable to be able to handle more complex LAI inputs such as PDF’s.
REFERENCES


examination of the relationship between remotely sensed and field measurements of LAI. Functional Plant Biology, 35(10), 1070-1079.


Udvardy, M.D.F. (1975) A classification of the biogeographical provinces of the world. IUCN Morges (Switzerland).


Websites:
APPENDIX 1: Box plot of LAI for temperate broadleaf and mixed forests in 2003