GIMA

Geographical Information Management and Applications

Improving Leaf Area Index (LAI) estimation with satellite imagery

Msc.Thesis
Student Name:
Georgios Ovakoglou

Supervisors:
Dr. Arnold Bregt, Professor
Dr. Jan Clevers, Associate Professor

External Supervisor:
Dr.Thomas Alexandridis, Assistant Professor
Aristotle University of Thessaloniki (AUTH)

Universiteit Utrecht

TUDelft

UNIVERSITY OF TWENTE.

WAGENINGEN UR

Abstract

The Leaf Area Index (LAI) is an important parameter characterising vegetation and knowledge of LAI is crucial for describing the activities within an ecosystem. It is widely used as a basic input parameter in hydrological and bio-chemical models for the estimation of the water-cycle, agricultural primary production and other parameters. This study attempted to improve the spatial resolution of MODIS LAI product (1000m pixel size), through the regression analysis applied to MODIS EVI (1000m) and LAI data and the use of the estimated regression equations in a downscaling model using images from the EVI product of Landsat (30m) and several land-cover maps. Regression analysis was applied in 5 selected study sites around the world, which differ with respect to the climate conditions. Several scenarios were tested in order to find the important parameters affecting the LAI-EVI relationship (vegetation type, seasonality) and the data were found to be described best by a linear fit. During the downscaling process the estimated LAI-EVI equations were used to calculate LAI values and subsequently create LAI maps at the Landsat spatial resolution level (30m). The results of the study showed that vegetation type has the highest influence on the EVI-LAI relationship, as well as that the sensitivity of EVI to LAI is lower in periods of high biomass production. The created LAI maps showed visual similarity of high level in patterns of LAI value distribution, when compared to the corresponding lower resolution LAI maps (MODIS). The comparison of field data with the model estimated values of LAI showed high correlation especially during the dry period. The lowest correlation was observed during the rainy season when the availability of cloud free pixels in the LAI images was low. For most of the cases examined, the model gave statistically significant results (at 0.05 and 0.001 level) with the r coefficient values ranging from negative (-0.1783, one case) and relatively low (2 cases, 0.25 and 0.32) values, to moderate (3 cases, 0.4-0.7) and high (5 cases, 0.7-0.935). Limited samples per vegetation type on specific dates, the diversity of vegetation species within the same vegetation type, as well as saturated EVI values were evaluated to be the most possible factors affecting the regression analysis results. The model estimated LAI values were found to correlate better with the field measurements in study areas of the Southern Hemisphere.

Keywords: LAI, EVI, MODIS, Landsat, regression analysis, vegetation type, downscale model, high resolution LAI maps.

Table of Contents

List of figures

List of tables

1 Introduction

The Leaf Area Index (LAI) is an important parameter characterising vegetation and is defined as the one sided green leaf area per unit ground area in broadleaf canopies and as one half the total needle surface area per unit ground area in coniferous canopies (Myneni, 2012). Knowledge of LAI in a study area is crucial in order to describe the activities within an ecosystem. It is widely used as a basic input parameter in hydrological and bio-chemical models for the estimation of the water-cycle, agricultural primary production, exchange of carbon dioxide, desertification risk, etc.

Measuring LAI in-situ is in general considered to be a difficult and costly process. It can be achieved with direct methods that are ground-based (e.g. destructive sampling and litterfall collection), or indirect methods, such as photographing the canopy with a vertical hemispherical lens and processing the acquired photos with specialized software (Weiss et al., 2004; Alexandridis et al., 2013b).

Remote sensing is considered as an indirect method, which has the ability to provide LAI data at various spatiotemporal scales, due to the variety of spatial and temporal resolution of the available satellite data. Retrieval of LAI on the landscape scale can be achieved through different platforms, such as aerial, ground and space-based (Zheng and Moskal, 2009). The underlying hypothesis for estimating LAI with remote sensing is that increased vegetation biomass, which is related to high LAI values, reflects the near-infrared and absorbs the red wavelengths. Three methods are characteristic of the remote sensing approach:

- **the empirical based model**, relating vegetation indices with in-situ LAI estimations (Clevers, 1988a; Butson and Fernandes, 2004; Soudani et al., 2006);
- **the radiative transfer model**, which uses physical laws to describe with accuracy the spectral variation of canopy reflectance as a function of canopy, leaf and soil background (Goel, 1989; Kimes et al., 2000; Meroni et al., 2004);
- **the hybrid model**, which combines the advantages of both previous methods into an integrated LAI inverse model (Shen et al., 2014).

Colwell (1974) demonstrated that the near-infrared (NIR) and red reflected energy are closely related to the amount of plants (vegetation) present on an area. The photosynthetic process of any plant and especially the amount of chlorophyll present in its leaves during the stages of the plant's development, affects the amount of the reflected red energy, as this is decreased when the presence of chlorophyll is being increased. The exact opposite happens with the amount of NIR reflected energy, as it increases while the healthy plant leaves grow. In an ideal situation, the

two types of reflected energy could be quantified and used in order to interpret the amount of vegetation present in a location, the health state of the plants, or even their growth development. However, due to several characteristics such as the influence of the atmosphere between the satellite sensor and the object (plant) reflecting the energy, the canopy structure or background, etc., it is not feasible to quantify with accuracy the measured reflection value and use it to monitor the vegetation state. According to Huete et al. (1999), the solution to this problem is to combine two or more bands into a single equation, formulating this way the so-called vegetation indices (VI), which are widely used to describe, monitor and quantify the state of vegetation in an area, or to describe important vegetation characteristics.

Heterogeneity of vegetation causes under-estimation levels of LAI (Fensholt et al., 2004) and therefore it is wise to select for the research purposes the tools (VIs) that provide the maximum stability and accuracy possible. An index that is often used in estimating LAI is the Normalized Difference Vegetation Index (NDVI) (Wang et al., 2005a). Major disadvantage is that the NDVI is not sensitive at medium to large LAI values. From the free of cost available VI products, those of the Moderate Resolution Imaging Spectroradiometer (MODIS) were selected, because this sensor provides also a LAI product (which will be used for the present study purposes) in the same spatial resolution as the VI data and for the exact same time periods. The Enhanced Vegetation Index (EVI) was selected to be used for the research purposes, since according to relevant literature it is more stable and accurate compared to the NDVI, while in general these two indices have similar performance no matter what crops are analyzed (Colombo et al., 2003). Moreover, the EVI has improved sensitivity in high biomass regions and according to Wang et al. (2005b) the values of EVI are more useful than those of the NDVI in estimating LAI. Even though NDVI is considered to be more appropriate for estimating chlorophyll responsiveness of the vegetation, the EVI is more sensitive to the variations of canopy structure (canopy architecture, crown type, vegetation physiognomy and LAI) (Chen and Sun, 2010). An additional advantage is that EVI normalizes and adjusts the reflectance in the red band as a function of the reflectance in the blue band, minimizing this way the residual atmosphere influences (Huete et al., 1997).

As mentioned above, the LAI is a very important input parameter in hydrological models operating at river basin or sub-basin level (Schumann, 1993) and especially in the estimation of water related parameters, such as dynamics of the water flow (superficial or groundwater flow) (Pedro et al., 2011). For these models, LAI is a very common input as it defines the water loss by evapotranspiration, roughness for surface water flow and development stage of plants (Teixeira, 2010; Chambel-Leitão et al., 2011; Jassas et al., 2015). The models' requirement for LAI input is characterized by a relative flexibility in resolution and time steps, meaning that the ideal situation would be to acquire LAI data at as high resolution as possible,

and at frequent time steps to describe the development of vegetation (depending on the examined vegetation type).

This study was conducted within the project of the Research Committee of the Aristotle University of Thessaloniki (Greece) entitled "Improvement of the estimation of Leaf Area Index (LAI) at basin scale using satellite images" and capitalized on the results of the MyWater research project "Merging hydrological models with Earth observation data for reliable information on water" (http:\\mywater-fp7.eu), using part of the data acquired during the progress of the project for several test sites.

1.1 Problem Definition

Several products are available from FAO, ESA and NASA that provide LAI maps at regional or global scales, at high frequency (eight days interval), but at low spatial resolution (1000 meters, only recently upgraded to 500 meters). These products could satisfy the research needs of global and regional scale studies. However, at local scale studies such as the area covered by a catchment, the above mentioned LAI products cannot be considered representative due to the large level of generalization.

In relevant literature, efforts to estimate LAI at higher spatial resolution level are narrowed to only some kinds of crops (Walthall et al., 2004), certain types of forest vegetation (Eklundh et al., 2001), or in general for relatively limited study areas. They are also dependent on in-situ observations, which are costly and usually limited to a few sampled locations. For example, some researchers try to focus on the different characteristics of several biomes but without considering other characteristics affecting LAI, e.g. seasonality (Turner et al., 1999; Tian et al., 2000), while other (Myneni et al., 2002) take in consideration all the characteristics involved, with the research focusing on evaluating the existing LAI products. On the other hand, most cases are usually dedicated to certain types of crops or forests, taking in consideration all related landscape and climate characteristics (Chen and Cihlar, 1996; Fassnacht et al., 1997; Gupta et al., 2000; Johnson, 2003; Fensholt et al., 2004; Wang et al., 2005a), but their resulting methodology cannot find application in other vegetation types. In some cases (Colombo et al., 2003; Foody et al., 2003; Wang et al., 2005b) the researchers take in consideration all the characteristics related to seasonality, vegetation type and phenology stage of the examined areas, which are described with different equations per characteristic. Finally, there have also been approaches to study and estimate LAI at catchment area level, using satellite data of various spatial resolutions (Topaloglou et al., 2013; Silleos et al., 2014).

Considering all the above, there is the need to develop a methodology for estimating LAI at higher resolution levels, without being based on empirical equations and in-situ observations and moreover without focusing only to certain

vegetation types, but instead taking in consideration all vegetation species present in a landscape. Therefore, the present research focuses on a wide range of vegetation and seasonality characteristics in several study areas distributed worldwide. Since relevant literature for estimation of high resolution LAI only focuses at certain crop or vegetation types, or only for a certain period of the growing season, it would be interesting to examine whether there is a methodology that could find wider application (within the river basin and within the hydrological year). Moreover, if the downscaling method of MODIS LAI (presented in the next sub-chapter, 1.2) using MODIS EVI and Landsat EVI imagery proves to be successful, it will be feasible to develop a procedure which will provide remote sensing LAI data of high spatial and temporal accuracy, without the need of in-situ measurements.

1.2 Research objectives

The main goal of this research was the improvement of LAI estimation at the river basin scale, using LAI and EVI products of moderate and high spatial resolution, created from satellite imagery. As side goals were considered the spatio-temporal analysis of the relation between MODIS LAI and EVI, and the downscaling of LAI using higher resolution satellite images (Landsat). The research was applied in selected river basins scattered around the globe (Greece, Netherlands, Portugal, Mozambique and Brazil).

The results include the analysis and presentation of correlation of LAI with the EVI for the selected test sites, the formulation of a methodology for downscaling the estimation of LAI from lower to higher spatial resolution and finally, the creation of LAI maps with improved resolution for the test sites.

The above mentioned research objectives led to the three main research questions of this thesis:

1. What is the nature of the relation between LAI and EVI across the various climatic conditions (and thus vegetation types) of the study areas?
2. How is this relation affected by seasonality (different stages of vegetation growth) and land-cover (different types of vegetation)?

3. Can the relation (equations derived by regression analysis between EVI-LAI) be used to downscale LAI using higher resolution satellite images than MODIS?

1.3 Thesis Overview

The first chapter of the Thesis introduces the subject of the research and identifies the objectives examined in the following chapters. The second chapter is dedicated to define and analyze all the concepts and terms related to the theoretical background of the thesis, as well as the description of the algorithms behind the satellite products used. Description of the study areas and data used throughout the

research can be found in the third chapter, followed by the fourth chapter which is the description of the methodology used. Chapter five includes all the related results after the application of each methodological step described previously, along with methods and techniques to validate the methodological results, while chapter six includes discussion of the results. The final chapter is conclusions, where it is expressed whether the goals were partially or fully fulfilled and in which extent, followed by the list of references used throughout the research. Chapter nine is provided as an appendix.

2   Literature review

In this chapter the relevant literature is being analyzed and insight over the theoretical concepts behind relevant research and the products used throughout the research, as introduced in the first chapter, is provided.

2.1   Vegetation monitoring with VIs & LAI

In general, monitoring and analyzing the stages of vegetation growth during the repeated life-cycle of each type of plant, known as phenology, is considered to be a valuable tool in order to assess the impact of climatic changes. Several scientific fields, such as forestry and agriculture, are closely related and being affected by changes in climate, mainly because of the impact that such changes have on phenology, creating new conditions and affecting characteristics of the vegetation (van Vliet et al., 2003) in a manner similar to the effects of evolution on the human kind during the years. Therefore, it is crucial for the scientific community to have in place such tools or techniques that would simplify the procedure of phenology monitoring and at the same time minimize the effort and time needed to be spent during the relevant tasks.

The LAI, as well as the VIs, have been and are still considered to be some of the most valuable tools with a global use, in order to monitor and evaluate the progress during plants' growth. Several global studies have validated the MODIS LAI product (for details see section 2.2.1) using related fieldwork , for a wide range of vegetation types and time-periods, with the accuracy of the product estimated at 0.66 RMSE (Root-Mean-Square-Error) LAI units in case all types of biomes were taken into consideration, falling to a value of 0.5 RMSE units, when broadleaf forests were excluded (Myneni, 2012). Researchers warn that there has to be increased awareness using such data (MODIS), since the relation between LAI and VIs is not applicable everywhere and all the time, due to measurement geometry and spatial resolution of the examined plant canopies (Tian et al., 2000). Seasonality during the growth cycle of each vegetation type seems to be another crucial factor when regression analysis is applied between MODIS LAI and VI data. As Wang et al. (2005a) support, when regression analysis is performed between LAI and VI's of MODIS, R^2 values are constantly increasing during different growing stages, but they also warn that assumptions and conclusions should be made only per growth-cycle (over one year period), as year to other years comparison between LAI/VIs may be erroneous . Meteorological condition during different seasons is also a characteristic affecting the overall quality of such data, however by the 8-day image compositing method that MODIS uses, it is possible to eliminate the contamination from cloud cover (Knyazikhin et al., 1999).

Moreover, there are major advantages related to the use of MODIS data, since there is no need for additional processing of the datasets in order to correct any atmospheric effects present, as they are delivered already pre-corrected (Walker et al., 2012). Even though the spatial resolution of MODIS data is considered to range from moderate levels (250 m.) to relatively low (1000 m.), it has been proved that higher resolution data of Landsat can also be used together with MODIS, in order to improve the spatial resolution in the estimations. Masek et al. (2006) support in their research that comparison between MODIS and Landsat data shows very high consistency, but on the other hand, variations can still be found due to bandwidth and solar geometry differences (Gao et al., 2006). Although MODIS includes spectral bands with a number of differences compared with Landsat bands (in terms of band-passes and spectral response), they are generally considered to be equivalent (Masek et al., 2006). An overview of the similarities between the wavelength bands of Landsat (TM, ETM+, and OLI/TIRS) and MODIS sensors can be found in Table 1. From the presented wavelength values of the associated spectral bands, it is evident that the best agreement can be found between the data of Landsat 8 (OLI/TIRS sensors) and MODIS, but overall all the values for each band and the equivalent band of another sensor seem to be very close. Therefore, a priority has been given on choosing Landsat 8 data over Landsat 7 and 5 whenever this is possible, in order to retain errors, which might occur due to this inconsistency, to the minimum level possible.

Table 1: *Wavelengths (in micrometers) of blue, green, red and NIR bands for MODIS and Landsat sensors*

Sensor	Band 3	Band 4	Band 1	Band 2
MODIS	0.459 - 0.479	0.545 - 0.565	0.62 - 0.67	0.841 - 0.876
Landsat	**Band 1**	**Band 2**	**Band 3**	**Band 4**
(TM)	0.45 - 0.52	0.52 - 0.60	0.63 - 0.69	0.76 - 0.90
(ETM+)	**Band 1**	**Band 2**	**Band 3**	**Band 4**
	0.45 - 0.52	0.52 - 0.60	0.63 - 0.69	0.77 - 0.90
(OLI/TIRS)	**Band 2**	**Band 3**	**Band 4**	**Band 5**
	0.45 - 0.51	0.53 - 0.59	0.64 - 0.67	0.85 - 0.88

Considering the effect of the atmosphere on the quality of Landsat and MODIS data, it is supposed that atmospheric correction of MODIS images can be performed with a greater degree of accuracy than Landsat images due to the improved onboard capabilities of MODIS (Vermote et al., 1997; Vermote et al., 2002). Comparison between VIs (Landsat-NDVI, MODIS-NDVI) products of Landsat

and MODIS (Maiersperger et al., 2013) for a specified study area, showed that there is a very strong linear relation between the values of the two products, with the only exception being a dataset of MODIS used that was dominated by heavy cloud presence. The MODIS-NDVI values show a slight but systematic over-estimation compared to these of Landsat, with the R^2 value of the regression analysis between the datasets reaching a level of 0.8 (80%), the Root Mean Square Deviation (RMSD) value being 0.09 and the point values being plotted on the 1:1 trend-line showing a slope of 0.85.

EVI rather than other available VIs was selected for the Thesis purposes. Relevant research showed that for the period of the past 35 to 40 years, several attempts have been made to study the relation between LAI and other VIs (NDVI, EVI, Weighted Difference Vegetation Index (WDVI), Soil Adjusted Vegetation Index (SAVI) and other), analyzing each time the benefits or disadvantages behind these choices. For example, Clevers (1988b) stated that the most suitable reflectance factors that could be used in order to estimate LAI are those in the NIR wavelength, with only disadvantage being the fact that due to soil moisture, which shows variations during the growth stages of a canopy, problems might appear if a multi-temporal analysis is applied, since the soil reflectance is dramatically affected by it. The WDVI could be used to bypass this problem, as it is known for the improved correction for soil background when green vegetation is examined (Clevers, 1991). However, in seasons (periods) or in vegetation types where green vegetation is not the dominating characteristic of the examined canopy, problems in the LAI estimation might still occur. The NDVI, as mentioned in the introduction, is one of the most common indices used in relevant research for the LAI estimation, but still, for some regions (e.g. Western Europe) with agricultural conditions showing variations compared to other regions, the associated results are not of a satisfying level (Clevers, 1986). Cohen et al. (2003) support that coupling more than one VI of multiple dates in the regression analysis, could enhance the overall accuracy of the analysis results, but due to the fact that the present Thesis uses five different geographic regions as study areas and for a time period of a whole year, the related data and the work demanded would dramatically increase in volume.

Finally, Ganguly et al. (2012) have recently started developing (in a project funded by NASA) an algorithm, which will be able to estimate LAI at high resolution level through the direct use of Landsat data, and specifically by using the Bidirectional Reflectance Factor (BRF) as a function of spatial resolution and wavelength. However, this project is not finished yet and according to the scientific team responsible for it, several parameters such as the provided scale of the final product, the heterogeneity of vegetation as well as the validation stage of the results are factors that still need research.

Considering all the above, along with the basic advantages of the EVI presented in the introduction chapter, led to the choice of the data and techniques that was finally used in the present Thesis methodology.

2.2 MODIS products

The Moderate Resolution Imaging Spectroradiometer, also known as MODIS, is an instrument of the Terra satellite which was launched by NASA on 18th of December 1999. On 4th of May 2002, an additional similar instrument was launched on the Aqua satellite. Since then, the two instruments provide continuous daily data for uses related to the monitoring of ecosystems, oceans and the atmosphere, helping the scientific community in several fields to apply research on climate and earth surface changes (Savtchenko et al., 2004). 36 spectral bands are available through the MODIS instrument, covering a range from 415 nm to 14,235 nm, with a viewing swath of 2330 km and a spatial resolution which varies from 250 m to 500 m and 1000 m, depending on the characteristics of each product provided.

Several products have been created by the scientific teams of NASA since the initial launch of the instruments into orbit. These products are divided in five broader categories, namely MODIS Level1, Land, Ocean, Cryosphere and Atmosphere products. The LAI/FPAR and EVI products are part of the MODIS Land products category and are provided by the LPDAAC website already corrected for atmospheric gases, thin cirrus clouds and aerosols, based on the MOD09 product daily surface reflectance series (Fensholt et al., 2004).

MODIS data of LAI and EVI is provided with relevant information about its quality. Factors such as cloud or snow cover, dark soil type, gas emissions, sensor viewing angle, etc. can significantly affect the reliability and quality of the EVI and LAI data acquired by MODIS. Additionally, Myneni et al. (2002) point out that the algorithms behind the MODIS products are expected to fail more often over latitudes with frequent cloud cover, such as the tropics. For these reasons, developers of the products advise users to examine the per-pixel product quality information to screen out poor quality data before use in applications, science, or research (Huete et al., 1999). Subsequently, for each image downloaded, Quality Assurance (QA) data per pixel is also provided, in order to discard pixels that are characterized by low quality and thus low accuracy. This is considered as a common practice in relevant research (Ahl et al., 2006; Chen and Sun, 2010).

2.2.1 LAI algorithm description

The MODIS FPAR/LAI product is available for downloading through the Land Processes Distributed Active Archive Center (LPDAAC) internet platform of the U.S Geological Survey. The exact name of the LAI product used throughout the Thesis is MCD15 and apart from LAI data, it also includes data of the Fraction of Absorbed Photosynthetically Active Radiation (FPAR) that a plant canopy absorbs for photosynthesis and growth in the $0.4 - 0.7$ µm spectral range. MCD15 is

characterized by a spatial resolution of 1000 meters per pixel and recently a version of 500 meters per pixel has also been made available. Seven spectral bands (648 nm, 858 nm, 470 nm, 555 nm, 1240 nm, 1640 nm, and 2130 nm) participate in the development of the final product, along with land-cover data (from MODIS MODAGAGG and MOD12Q1 land products), since the algorithm and the final estimation of LAI values per pixel depends heavily on the type of vegetation present in each pixel. Vegetation is described by these products in seven general classes, dividing all possible types of plants in seven biomes (grasslands, crops, shrubs, broad-leaf crops, savannahs, broad-leaf forests and needle-leaf forests).

The philosophy behind formulation of the MODIS LAI algorithm is mainly based on the three-dimensional radiative transfer theory. Atmospherically corrected surface spectral bi-directional reflectance is used as input in order to make an estimation of the most probable LAI values in the areas of interest (Myneni et al., 2002). In order to tackle the problem of inverse LAI retrieval from Bidirectional Reflectance Distribution Function (BRDF), a three-dimensional approach is used by the developers of the product, adopting several techniques such as Green's function and ad-joint formulation. According to Myneni et al. (2002), the initial function of the algorithm is to try and estimate through the use of characteristics such as the sun position and view directions, using BRDF values and any uncertainties present, the observed values of LAI and FPAR. The values found are then compared with LAI and FPAR values from models representing similar natural conditions, in order to evaluate which of these are the acceptable values (or the so-called solutions). The dispersion magnitude of the estimated values is used as input to evaluate the reliability of the corresponding LAI values and it is not possible to improve the accuracy in the retrievals if any additional information is not available. Taking in consideration the natural variability of vegetation canopies, it is then possible to divide the three-dimensional transfer of radiation in canopies to two sub-problems; one considering dark surfaces and the other related to anisotropic sources at the canopy bottom (Knyazikhin and Marshak, 2000). Red and near-infrared bands are used to examine whether the canopy leaves are completely absorptive or if they fully reflect/transmit the radiation, with the in between conditions being characterized as super-positions (Myneni et al., 2002). It is then feasible to proceed with more accuracy in the specification of changes in structural characteristics of the canopy. All the previously mentioned characteristics are finally stored in Look-Up-Tables (LUT), which can be used as reference data to model various canopy structures and soil types. Overall, as explained by Wang et al. (2001), reliability and accuracy of the algorithm is heavily dependent on the quality and quantity of spectral information used as input.

After all these steps, if the algorithm still fails to make a valid estimation due to the effects of the atmosphere, a backup algorithm using the values of Vegetation Indices (VIs) is used instead (Myneni et al., 2002). However, the data created through

the backup algorithm is provided by LPDAAC with a quality flag and is labeled as "marginal", indicating that researchers should use such data with caution.

Wang et al. (2005b) also note that cloud state should always be a quality check flag and this was the reasoning behind the choice of the MODIS research team to compose 8-day period images with the LAI values, in order to avoid the effect of cloud coverage in the output data (Knyazikhin et al., 1999). Relevant literature points out that the minimum period over which significant changes might occur to the LAI values of any vegetation type in real life conditions is 9 days (one-year cycle agricultural crops) (Alexandridis et al., 2008), and therefore the composition period of 8 days was evaluated as sufficient and safe to be adopted for the Thesis purposes.

2.2.2 EVI algorithm description

The Vegetation Indices (VIs) provided by MODIS are level 2 products and consider two different indices: the Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI). They are available at various spatial resolution levels, with the majority of the related products characterized by a spatial resolution of 500 and 1000 m, while limited products can also be found in 250m (Huete et al., 2002). Unlike the MODIS LAI/FPAR product, the estimated values of the VI products are computed in the same way in time and space, which actually means that characteristics related to the type of land-cover and soil in the areas of interest are not taken into consideration. In general, the two products are characterized by a 16-day time-step and for their creation a MODIS-specific compositing method is used, which is based on the Quality Assurance (QA) product, in order to remove low quality pixels. After the exclusion of low quality and low accuracy pixels due to extreme atmospheric (clouds, etc.) or land (snow, inland water, etc.) conditions, a constrained view angle (<30 degrees) approach is applied on the remaining good quality VI values, which selects a pixel to represent the compositing period. The philosophy behind the choice of the final value is that from the two highest values available, the one characterizing the pixel located closer to nadir is being used (Huete et al., 1999).

The overall accuracy of the MODIS EVI product is estimated up to ± 0.025 units (Huete et al., 2002). Apart from the atmospheric and viewing angle parameters that might influence the EVI value estimation accuracy, factors such as extreme brightness or darkness of a surface, snow coverage and presence of desert areas or inland water bodies might also affect the accuracy of the estimations. This is mainly due to the atmospheric corrections applied to the blue, red and NIR spectral bands data, which participate in the equation used to calculate the EVI values, and follows below (Huete et al., 1999):

$$EVI = G \times \frac{(NIR - RED)}{(NIR + C1 \times RED - C2 \times Blue + L)}$$ where G is the
Gain factor, near infra-red (NIR), red and blue are atmospherically corrected values
of surface reflectance, L is the factor for the canopy background adjustment and C1 -
C2 are the coefficients of the aerosol resistance term used for the blue band, in order
to correct aerosol influences in the red band.

The MODIS EVI-algorithm uses the values of G=2.5, L=1, C1 = 6, and C2 = 7.5.
Huete et al. (1999) add that the overall accuracy of the estimated MODIS EVI values
shows a high level of agreement with satellite data of other sources (such as Landsat
EVI and ASTER EVI), as well as with field means (such as FLUXNET Towers), no matter
the area of interest (worldwide character) and without being affected by seasonality.
Agreement between the different sensors used for the EVI estimation is even higher
in areas with high levels of biomass.

2.3 Landsat products

Landsat, which is a play of words between Land and Satellite, is the longest
program running related to satellite imagery data since 23rd of July 1972, which was
the date the Earth Resources Technology Satellite was launched. Since then, 7 more
satellites have been sent into orbit, with one of them (Landsat 6, 5th of October 1993)
failing to reach the orbit during launch. The most recent of them, Landsat 8, was
launched on 11th of February 2013. The Landsat satellites and their sensors are
considered to be major components of the NASA EOS program, supplying the
scientific community with continuous high resolution earth surface data used in the
fields of environment, agriculture, biology, surveying, cartography and many more.
The US Geological Survey is responsible for the distribution of Landsat data through
the EarthExplorer platform available on their website. Recently, pre-processed higher
level data of Landsat (level 2 and higher) has also been made available for ordering,
through the EROS_ESPA platform of USGS
(https://espa.cr.usgs.gov/login?next=http%3A%2F%2Fespa.cr.usgs.gov%2F)

Through the years, the sensors on board of each satellite have been evolving,
with 5 different types being available until the launch of Landsat 8: the Multi-Spectral
Scanner (MSS, Landsat 1-3), the Thematic Mapper (TM, Landsat 4-5), the Enhanced
Thematic Mapper Plus (ETM+, Landsat 7), the Operational Land Imager (OLI, Landsat
8) and the Thermal Infrared Sensor (TIRS, Landsat 8). Through the available sensors,
it was feasible and still is for the Landsat program to supply visible, infrared, thermal
and panchromatic imagery. The spatial resolution characterizing the Landsat data is
considered to be high, with the related products having a pixel resolution of 15, 30,
60 or 120m, depending on which spectral band the data comes from. The viewing
swath of each image covers an area of 185x185 km and the sensors operate in 11
spectral bands (7-8 bands until the launch of Landsat 8), covering a range between

0.43μm to 12.51μm. The temporal resolution of Landsat data is characterized by a 16-day period time-step.

2.3.1 EVI algorithm description

The equation used to estimate the EVI values on the Landsat level remains the same as in case of MODIS data, since the spectral bands used by the sensors of the two satellites operate in similar wavelengths and usually even the same co-efficient values are being used. Until recently, it was feasible to acquire EVI Landsat data by downloading the calibrated Landsat level 1 data of the needed period (EarthExplorer website platform) and follow a process to convert the data in EVI values per pixel. During this process the values of the bands (NIR-Red-Blue) participating in the MODIS EVI algorithm presented previously, have to be initially converted to 8-bit digital numbers ranging from 0-255 values (DN values), and then the surface reflectance (SR) values are being used in the EVI equation (using the same coefficients as in the MODIS equation) in order to calculate the final values of EVI.

Application of atmospheric correction is suggested by the developers in-between these steps, in order to ensure that the data is accurate, with several Remote Sensing software programs developing models to standardize the procedure, in order to reduce the effort and time needed to pre-process such data. Since the introduction of the EROS-ESPA website platform, it is possible to order already pre-processed and atmospherically corrected products of Landsat which include not only level 1 data, but also ready products related to VIs, and in general products that are feasible to be generated through the use of Landsat data. The additional benefit besides the cost related to human-hours and the time spent until the final VI or other level 2 or higher product is ready, is the fact that all the ordered data is ensured to be of high quality and no human related mistakes are involved, since the routines applied on the data until it reaches its final format, are being developed and applied by the scientific team of Landsat in a standardized procedure. Recently (January 2016), the developers of the data introduced in their EVI algorithm a Gain Factor (G) of 2.5 (which was not included in their previous versions), in order for the data to be of the same magnitude as those of the MODIS EVI. The rest of the coefficients participating in the equation had the exact same values as those used in the MODIS equation since the beginning of the product distribution.

2.3.2 Landsat atmospheric correction methods

The atmospheric correction routine used by the Landsat team, the Landsat Calibration, Reflectance, Atmospheric Correction Preprocessing Code or better known as LEDAPS, was developed with a philosophy similar to the one followed by MODIS team for the atmospheric correction of their products, using the same radiative transfer model in the algorithm and thus, decreasing the potential source of errors if the two datasets were compared (Masek et al., 2006). It was initially created

in 2006 by NASA, Goddard Space Flight Center (GSFC) and the University of Maryland and was intended to be used for the atmospheric correction of Landsat 4, 5 (TM) and 7 (ETM+) data. An advanced version of LEDAPS was released in 2011, with the contribution of the USGS Landsat Program as well.

The LEDAPS algorithm uses as input Landsat data along with data about digital elevation, aerosol thickness, water vapor, geo-potential height, and ozone, and through the use of radiative transfer models it provides information in the form of raster files for several characteristics; top of atmosphere (TOA) reflectance, surface reflectance and brightness temperature. Additionally, masks for the exclusion of low quality pixels (contaminated with the presence of clouds, cloud shadows or adjacent clouds) are available, as well as a mask to distinguish and isolate areas covered only by land or water. The techniques applied through the LEDAPS routine are divided in 6 modules, which are based on three more general techniques:

- The conversion of DN values to TOA reflectance.
- The detection of cloud contaminated pixels in the TOA reflectance.
- The conversion from TOA reflectance values and any auxiliary datasets to surface reflectance (SR) values.

The accuracy, precision and overall quality of the outcome of the atmospheric correction is ensured to be high, if some basic preconditions are followed; the Landsat data intended to be used as well as the metadata coming with it must be properly formatted and should be of TM or ETM+ sensor origin, must be geo-referenced and in image (geotiff, .img, etc.) file format, and must be accompanied by precise data about water vapor, air temperature, ozone and digital elevation (DEM) (Schmidt et al., 2013).

For the recently available data coming from Landsat 8, an additional algorithm has been developed, called the Landsat Surface Reflectance Code (LaSRC). This algorithm uses a unique radiative transfer model, along with auxiliary climate data from MODIS and coastal aerosol data in order to perform aerosol inversion tests. The most notable innovation implemented, compared to the LEDAPS algorithm, is the fact that LaSRC also takes advantage and uses data about the view zenith and the solar zenith angles as part of the atmospheric correction calculation.

3   Study areas and data

This chapter includes a description of the geography and basic characteristics (meteorological conditions, soil characteristics, vegetation presence, etc.) of the locations selected as study areas of the Thesis. Additionally, technical information about the quality and quantity of data used throughout the research is provided.

3.1   Study areas

Study sites of the research were selected to be catchment areas of rivers located in Greece, Portugal, Netherlands, Mozambique and Brazil, including the cross-border areas (Figure 1). The selected areas provided a relatively high diversity of the included types of natural vegetation and agricultural systems, in order for the research to be feasible to express conclusions considering a wide range of possible cases that could be found in real life conditions. Moreover, the geographic distribution of the study sites offered the opportunity to examine the relation between LAI and EVI across different vegetation types. This was feasible because of the different climate conditions existing between the selected locations, which subsequently are related to different vegetation conditions in each area. Nestos has a Mediterranean climate, Queimados has dry Tropical, Rijnland has temperate Oceanian, Tamega has warm Temperate - Mediterranean and Umbeluzi has a warm Tropical climate.
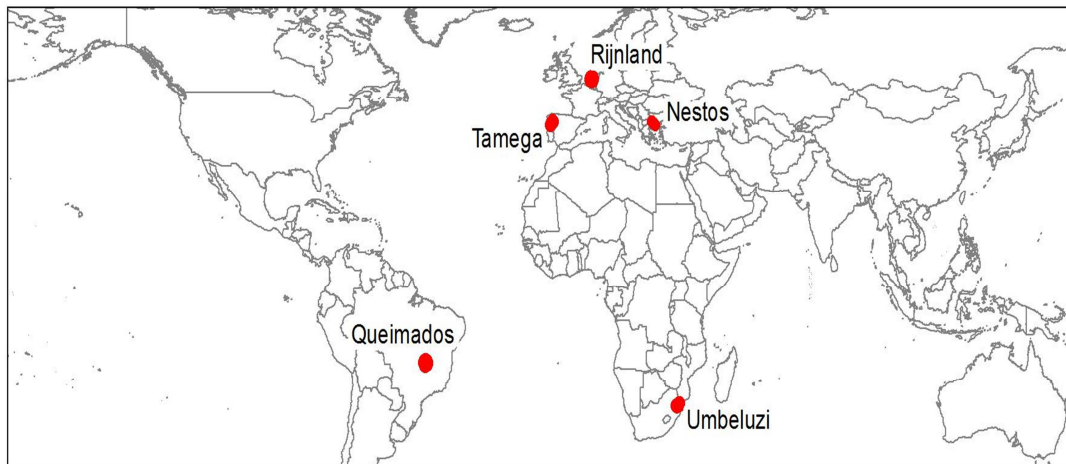
Figure 1: *Geographic distribution of study sites.*

The methodology was applied on all vegetation types considered to be characteristic of the selected catchment areas, such as agricultural areas, forest areas, grassland areas, etc.

3.1.1   Nestos

The study area of Nestos (Figure 2) includes Nestos River, which is located at a transboundary area between Bulgaria and Greece and discharges into the north

Aegean Sea. Its flow starts from the Rila and Pirin mountains of Bulgaria with a southeast direction. The total catchment area including it is around 6400 km² and its length is 234 km. A network of tributaries can be found on the northern part of the basin with most important of them being Diavolorema (356.4 km²), Arkoudorema (283.3 km²) and Despati (118.9 km²). The river basin morphology of Nestos is mainly characterized as mountainous or semi-mountainous, with only exception being the delta floodplain area before the discharge into the sea. Three dams are located in the Greek part of the river having a significant impact on hydrology. One of them is located at Toxotes village and is used for irrigation purposes, covering the needs of an area of approximately 25,000 ha. The other two (Thissavros dam, 1998 and Platanovrisi dam, 1999) are hydroelectric dams which were constructed for energy production, flood control and storage of irrigation water, with the Thissavros Dam considered to be as the highest and largest earthen dam in volume constructed in Greece (Boskidis et al., 2011). After the construction of the dams and through the years, several local organizations have been complaining about the influence these have to the river outflow, with water scarcity being their main concern. A shallow groundwater level in the floodplain is considered to be relatively important.
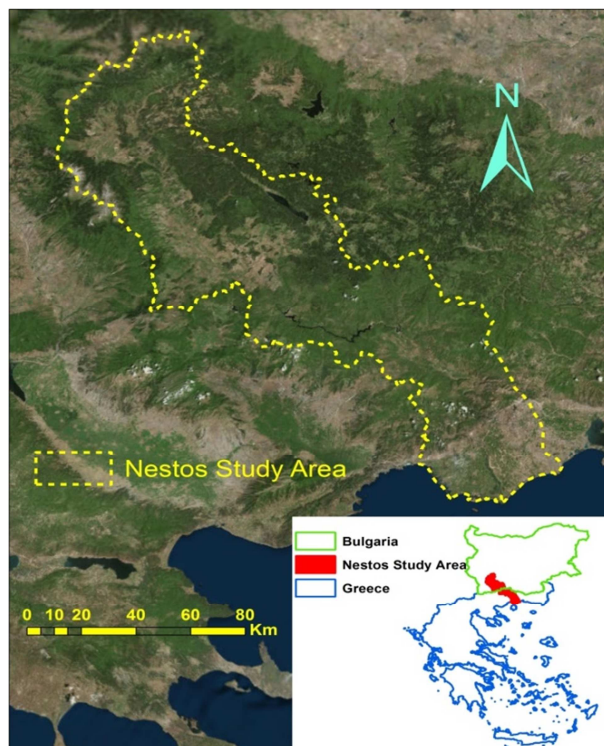
Figure 2: *Overview of Nestos study area. (source: Bing Maps)*

The average rainfall received by the Nestos area is about 600-1200 mm. The annual water flow is estimated to be 1830 x 106 m³. The climate in the mountainous area is characterized by cold winters and hot summers and is different from the

climate of the floodplain area which is characterized by hot and dry summers and mild winters, indicating a Mediterranean climate.

The hydrogeology of the river basin can be described in two parts: the northern part, characterized by the presence of carbonate rocks and the southern part consisting of alluvial deposits (Boskidis et al., 2011). Typical sandy soils can be found in the Nestos watershed, with the vegetation present at the area characterized mainly as natural vegetation, with a coverage rate close to 60% of the total area and the rest characterized as agricultural land. The natural vegetation can be described as typical Mediterranean climatic zone phenology (dry grass in the arid summer), while the agricultural vegetation mainly consists of single crop cycle plants. Coniferous and broadleaved forests, as well as shrubs are the main types of natural vegetation present in the mountainous parts. The river's delta hosts many habitats of rare and wild flora and fauna. This was the main reason to characterize the area as a Ramsar Wetland of International Importance and to include it in the Natura 2000 network.

An overview of the most important land-cover types found per study area according to GlobCover 2009 data, as well as the total area covered by each one of them, is presented in Table 2.

Table 2: *Land-cover types per study area.*

Land Cover type	Area covered (km²) and percentage % per study area									
	Nestos		Rijnland		Tamega		Queimados		Umbeluzi	
Artificial areas	157.91	2.45%	170.88	14.21%	180.14	4.5%	30.3	0.8%	53.4	1%
Herbaceous rainfed crops	152.69	2.37%	6678.82	56.43%	393.59	9.8%	1328.9	36.6%	169.4	3%
Irrigated herbaceous crops	285.36	4.44%			96.66	2.4%	235.1	6.5%	171.7	3.1%
Permanent crops	35.27	0.55%	34.74	2.89%	305.26	7.6%			21.6	0.4%
Broad-leaved forests	1685.17	26.19%	28.71	2.39%	238.62	5.9%	460.9	12.7%	931.7	17.1%
Needle-leave forests	1508.19	23.44%	31.69	2.63%	719.65	17.9%			5.9	0.1%
Shrubs	1311.71	20.39%	97.70	8.12%	1844,67	45.9%	522.4	14.4%	2021.6	37%
Grasslands	968.88	15.06%	90.59	7.53%	104.41	2.6%	941.5	25.9%	1397.8	25.6%
Recently burned areas					37.39	0.9%				
Barren	269.37	4.19%	29.37	2.44%	88.63	2.2%	64.5	1.8%	646.2	11.8%
Water bodies	59.41	0.92%	40.42	3.36%	7.45	0.2%	45.5	1.3%	47.3	0.9%

Source: MyWater (http:\\mywater-fp7.eu)

3.1.2 Rijnland

The Rijnland area (Figure 3) is located in the western part of The Netherlands covering approximately 1075 km². A percentage of 72% of the total area is characterized by areas under sea level, 15% consists of free draining areas and 8% of dune areas. The area is dominated by the presence of water canals mainly used for irrigation and leisure purposes, and combined with the low altitude mentioned

above leads to high risk of flooding during periods of heavy rainfall. However, the presence of several dikes, along with pump stations used to discharge excess water, eliminates this risk to minimum levels. A storage basin system of canals, lakes and pumping stations serve drainage and irrigation purposes to and from the main water system in connection with the North Sea.
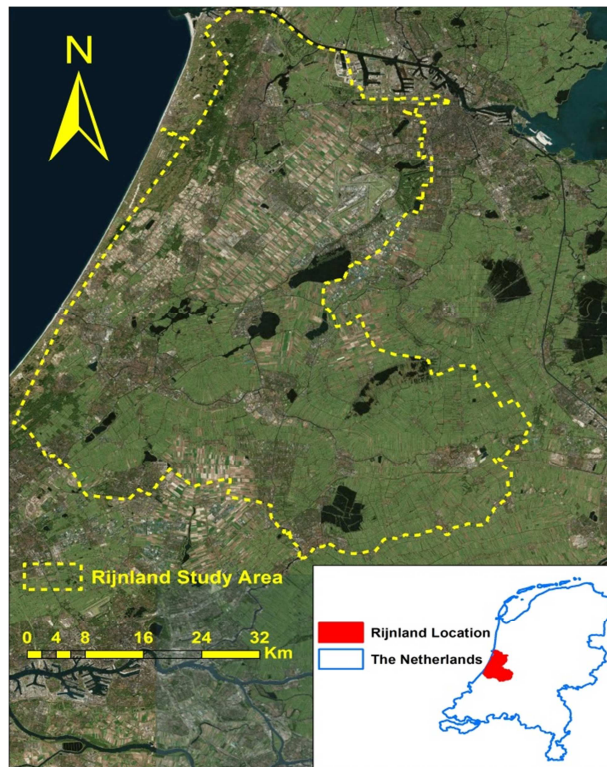
Figure 3: *Overview of Rijnland study area. (source: Bing Maps)*

In general, there is a very high percentage of vegetation coverage in the Rijnland area. The natural vegetation is fully covering the ground, with green vegetation present even during the summer period (with the exception of the coastal and sand dune areas), due to the presence of the water canals. Natural vegetation in the Rijnland area occurs in small patches, and mainly consists of broadleaved and a few coniferous trees. Natural herbaceous vegetation can be found in the coastal sand dunes and small and large settlements are also located in the area including lots of green space. Considering the farming areas, the fields are all rainfed with the most frequent crop being pasture for grazing purposes. Several other crops with a single cropping cycle can also be found, as well as greenhouses.

3.1.3 Tamega

The Tamega study area (Figure 4) includes Tamega River, which is a tributary of the Douro River, located in the transboundary area between Spain and Portugal. It is characterized by a North East - South West orientation and flows over a distance of

approximately 164.5 km, from which 140.5 km are located in Portugal and 24 km in Spain. The Tamega basin covers an area of 2,646 km² and the river's flow starts near the Serra of S. Mamede in Spain and enters in Portugal at an elevation of approximately 375m. The basin area is mainly characterized by high annual rainfall levels and steep slopes, which along with the reduced to moderate soil's erosive potential often leads to floods in the neighboring areas.
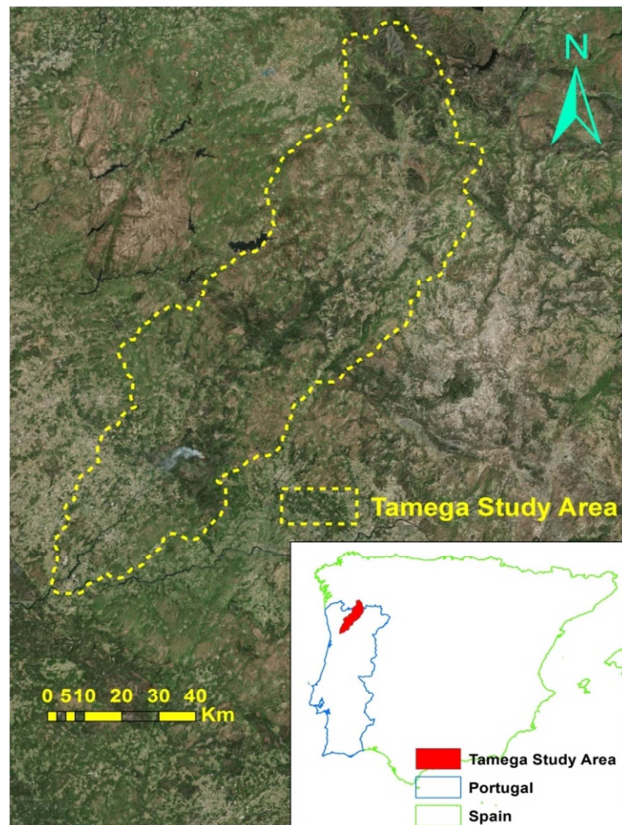
Figure 4: *Overview of Tamega study area. (source: Bing Maps)*

The vegetation in the area is a mixture of natural and agricultural land, with some of the fields located in the lower parts of the basin being irrigated. There is high vegetation coverage in the natural areas of the site, with forest fire occurrence being frequent in the area, mainly in managed forest plantations. The natural vegetation could be characterized as the typical Mediterranean climatic zone phenology, with dry grass presence during the summer period. The agricultural activities in the area mainly include fields with a single crop cycle.

3.1.4 Queimados

The watershed of Queimados is located between the cities of Brasília and Unaí and covers an area of 3629 km². Several rivers cross the sub-basin, with Rio Sao Marcos (466.7 km length, 870 m altitude) and Rio Preto being the largest and most important for the region, mainly because a reservoir used for agricultural (irrigation)

purposes, as well as a hydro-electric power plant built in 2004 (with 105 MW capacity, 1060 m dam length, 38946 ha³ volume) are located there. The dry-Tropical climate of the area gives an average annual temperature of 27 °C, with the maximum temperature being 40 °C and the minimum 12 °C. Occasionally, floods occur in the surroundings of Unaí city, but with the risk of such incidents characterized as relatively low.
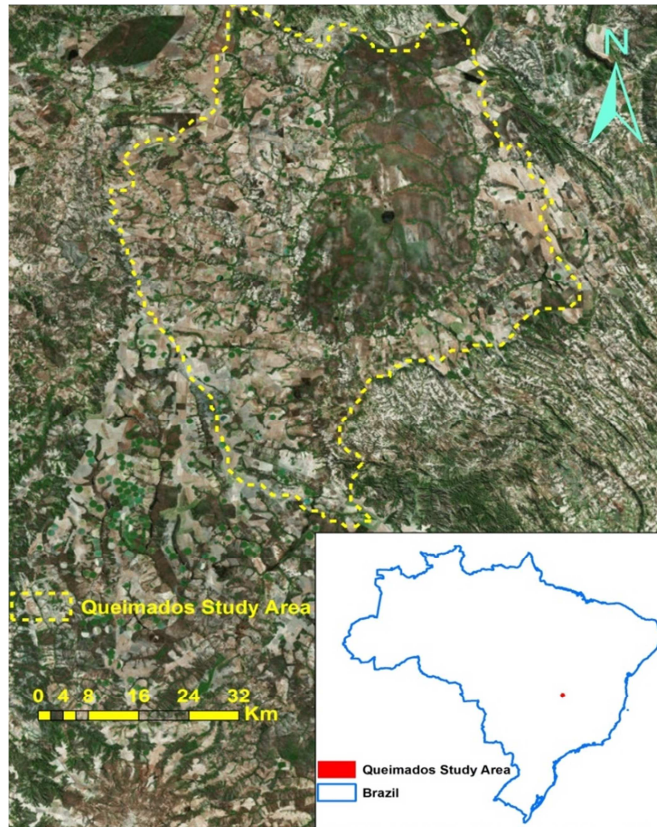
Figure 5: *Overview of Queimados study area. (source: Bing Maps)*

The study area is mainly characterized as agricultural. Characteristic of the local climate is the fact that agricultural activities might have more than one cropping cycle per year. Main crops cultivated in the area are grains, soybeans, rice, corn, wheat and beans. The predominant irrigation method is pivot irrigation, so the shape of each field in the study area can help to identify whether it is irrigated or not; for example rectangular shaped fields tend to be non-irrigated. Considering the natural vegetation presence in the area, it is characterized by a high degree of coverage and is mainly found along river corridors. Small patches of the "Brazilian cerrado", which is considered to be vulnerable to forest fires during the dry season (winter), usually can be found between the cultivated lands. In the central part of the study area, a military zone with no access to civilians exists. Stretches of grass for grazing are also distributed within the study site.

3.1.5 Umbeluzi

The study area of Umbeluzi (Figure 6) includes River Umbeluzi, which is a transboundary river between Mozambique, Swaziland and South Africa. The total area covered by its sub-basin is approximately 5400 km², with 2160 km² of these located in Mozambique, 3132 km² in Swaziland and 108 km² in South Africa. Nearby rivers are characterized as rather small in the study area, with the largest water bodies being the reservoirs of the dams located there.
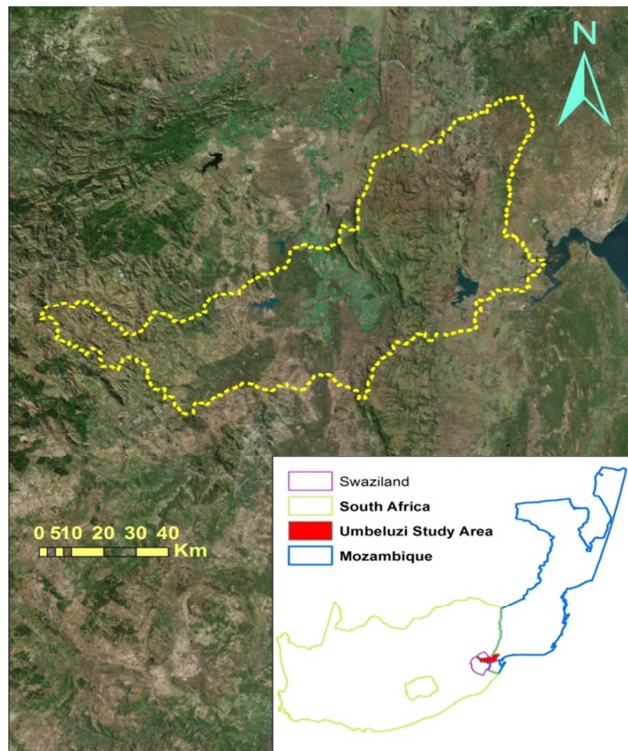
Figure 6: *Overview of Umbeluzi study area. (source: Bing Maps)*

As noticed also in the Queimados study area, due to the climate conditions present in the Umbeluzi sub-basin, it is possible that more than one cropping cycle per year might occur for some of the crops cultivated. Apart from only a few but very large sized irrigated plantations, there are also some small-scale irrigated fields. The rest of the agricultural land is rainfed and is consisting of small fields with mixed natural vegetation (grasses and bushes), and with a few fruit trees. Barren land presence is rare, while grasses and weeds can be found growing all over the study site, even in fields shortly after harvesting. Natural vegetation can be divided in two main categories: grasses and bushes. However, the term bush in this location does not define only small sized trees (not exceeding 1.5 meters), but mainly characterizes whole areas with savanna-like vegetation, sparse or thick trees and bushes. Tall trees or needle-leaved forests (with the common meaning used for these two words worldwide) are totally absent from the landscape.

3.2 Description of data

In this sub-chapter, information is provided about the quantity, quality and accuracy of data used for the Thesis, as well as any related errors that might be present due to known issues, which are being analyzed.

3.2.1 MODIS data

MODIS images are available for downloading through the LPDAAC website in .hdf format. The data downloaded included the annual time series for both EVI and LAI MODIS products (1000 m spatial resolution), along with their associated Quality Control (QC) datasets.

All pixels of each 8-day composite image of MODIS LAI and EVI acquired have been used in subsequent analyses, excluding those of low quality, and those falling on or close to the borders of different land cover types (mixed pixels) according to the GLOBCOVER land cover map, a process that is explained later in the Methodology chapter (chapter 4). The datasets were directly downloaded from the associated internet sources (LPDAAC).

3.2.1.1 LAI data

The selected MODIS LAI product (MCD15A2) represents 8-day composites retrieved from daily, atmosphere-corrected, bidirectional surface reflectance. However, a 16-day time-step was used for the Thesis purposes, because this is the frequency of image availability for the related EVI MODIS images that has been used along with the LAI data. A series of Terra/MODIS satellite images covering a period of a whole hydrological year (October 2012 – September 2013) was downloaded, and from the 46 available composite images, only 23 were finally used per study area, for the same periods EVI data existed. Examples of LAI MODIS images for the study area of Nestos can be found in Figure 7, where low LAI values are presented with red colour and high LAI values with green colour.
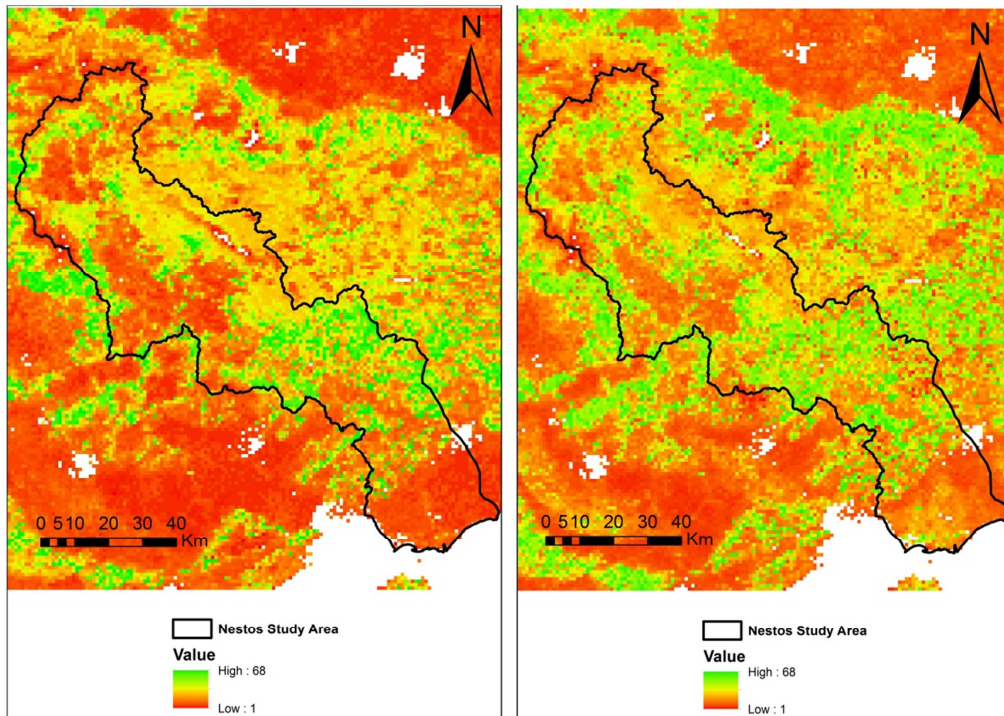
Figure 7: *Examples of LAI MODIS images for DOY 2013-273 (left) and DOY 2013-177 (right) in the study area of Nestos.*

3.2.1.2 EVI data

The used product is MOD13A2 "Vegetation Indices" generated at 16-day intervals. The same procedure as for the LAI data was followed to acquire the EVI data, resulting into a time-series of 23 images per study area for the same period (October 2012 – September 2013) and for the exact same 8-day periods as for the MODIS LAI images. MODIS uses the starting date of each 8-day composite to label the acquired images, which was also adopted for the Thesis purposes. Therefore, whenever throughout the Thesis the acquisition Day Of Year (DOY) for a MODIS image is used, it refers to the whole 8 day period, with the name of the image given after the starting date of the composite. Examples of MODIS EVI images can be found in Figure 8 for the study area of Tamega, in which high EVI values are represented with green color and low with red color.
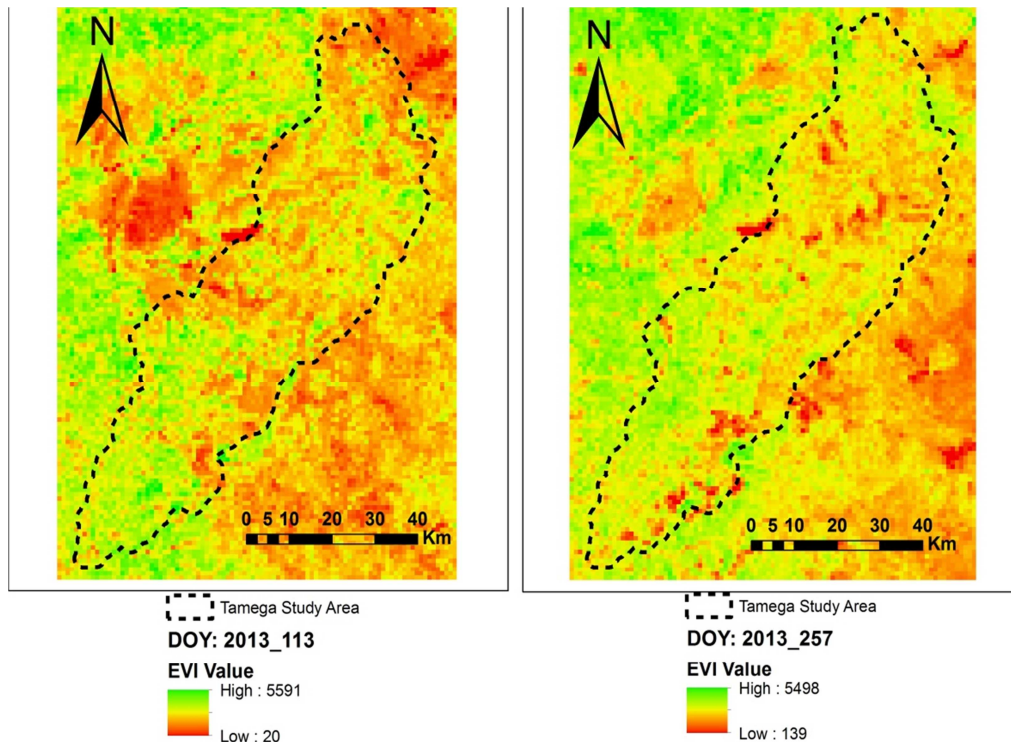
Figure 8: *Examples of EVI MODIS for DOY 2013-113 (left) and DOY 2013-257 (right) images in the study area of Tamega.*

3.2.1.3 Quality Control (QC) data

As mentioned in chapter two, the downloaded EVI and LAI MODIS datasets included extra layers of information, indicating the quality of each pixel. This Quality Assurance (QA) data provided information over each pixel's quality in the associated images, ensuring this way that pixels characterized by low precision in the relevant estimations would not be taken into consideration. After the necessary pre-processing of the QA data, which is explained in detail in the fourth chapter (Methodology), it was feasible to apply filters to the related data tables of each area, in order to isolate the so-called "marginal" (labeled with "1" value) or only "good" data (labeled with value "0"), for which "good" data the overall accuracy is supposed to be the maximum possible as supported by the product developers. In relevant research, when data of marginal quality was also used, the authors warned to use it with caution and only when there is a large data gap (Wang et al., 2005b). Due to the large size of a pixel (1000m) in the present research, it was decided to only use data of "good" quality, excluding any additional factor that could affect the overall accuracy negatively. An additional advantage from the exclusion of the "marginal" data is the fact that values of LAI are not estimated with the product's backup algorithm. This practically means that data from VI products were not used to estimate the LAI values, and therefore it would be wise to avoid using dependent data, in a research that intends to apply regression analysis using this data.

3.2.2 Landsat EVI data

Several atmospherically corrected satellite images from Landsat 7 and 8 within the same hydrological year as the previous mentioned MODIS images have been downloaded for each study area, covering the main phenological stages of vegetation. From these images, a selection was done of only those of a date close to the DOY of field measurements, where the MyWater project took place per study area. This was mainly decided in order to be able to validate the accuracy of the created higher resolution LAI maps, by comparing the LAI values per pixel of interest with the LAI values of high accuracy measured during the MyWater project at the same locations. Additionally, from the remaining images those covered extensively with clouds were also excluded, especially in the areas of interest (vegetation land-cover), since the data could not be used due to low accuracy. An overview of the images finally used can be found in Table 3. The images had also a proper UTM/WGS84 projection system pre-assigned, relative to the geographic location of each study area (UTM zones).

Table 3: *Final selection of the Landsat EVI images used.*

Study area	Satellite type	DOY (out of 365)	Cloud coverage
	Landsat 7	096/2013	50-60%
Nestos	Landsat 8	104/2013	75-80%
	Landsat 8	184/2013	40-45%
Rijnland	Landsat 7	265/2013	70-75%
	Landsat 8	273/2013	10-15%
	Landsat 7	147/2013	70-75%
Tamega	Landsat 7	243/2013	5-10%
	Landsat 8	251/2013	5-10%
	Landsat 7	309/2012	40-50%
Umbeluzi	Landsat 8	175/2013	<5%
	Landsat 8	191/2013	55-60%

An additional problem that needed to be tackled with the Landsat EVI data was the known problem of the Landsat 7 scan-line corrector, which creates line-gaps of no data in the images. However, in this case such a problem was evaluated as not affecting the associated research results, since this data was intended to be used only in comparison with existing field measurements in specified sample points. Application of interpolation techniques that could fill-in the gaps was therefore evaluated as inappropriate for the Thesis research purposes. This was decided in order to retain accuracy of the results at the highest possible level and avoid inserting errors (Alexandridis et al., 2013a). Examples of unprocessed (raw) Landsat EVI images for the study area of Umbeluzi can be found in Figure 9, where apart from the valid range of EVI values for healthy vegetation (value range 100 – 10000), values

representing non-vegetative conditions or low quality data are also included (values <100 and >10000).
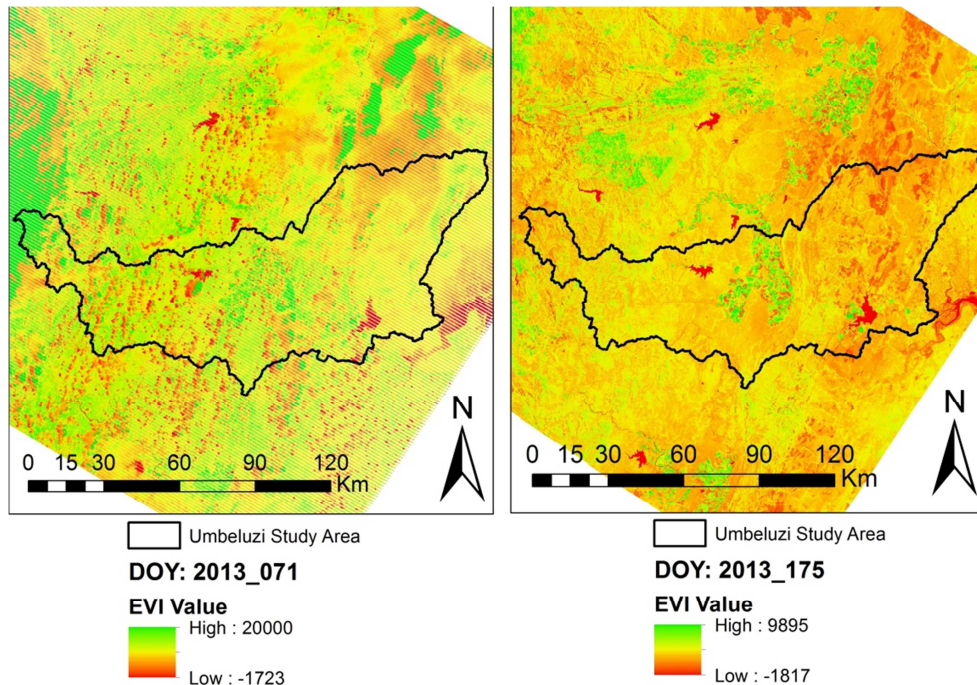
Figure 9: *Examples of EVI Landsat 7 (DOY 2013-071, left) and Landsat 8 (DOY 2013-175, right) images in the study area of Umbeluzi.*

3.2.3   Land-cover data

Information on land-cover at various scales is considered to be critical for a range of ecological, socioeconomic and policy questions. Land-cover maps can be used as input in order to produce LAI related data, data about evapotranspiration or prediction of soil erosion. Moreover, land-cover maps can be used for the detection of environmental changes, if landscape metrics of the areas of interest are also available for use.

Since the research intended to downscale the MODIS LAI maps' resolution of 1000m to Landsat resolution of 30m, a land-cover dataset (source: Landsat) corresponding to the new pixel size was also used, as well as land-cover maps of lower resolution (GlobCover, 300m resolution).

The land-use/land-cover (LULC) data used for the Thesis purposes has been created during the progress of the MyWater project for all the study areas and at a regional (1:1,000,000, GlobCover) and local (1:50,000, Landsat) scale. The Land Cover Classification System (LCCS) developed by FAO was used to produce the MyWater land-cover classes, describing the LULC in 12 classes in the regional and local maps. LCCS is a system classifying data a-priori and for this reason all the related classes intended to be used should be defined before the collection and classification of any data. The 12 classes defined were further reduced to 6 more general classes for the Thesis purposes, including types of biomes similar to those used by the MODIS land-

cover products, in order for the data to be comparable. These 6 classes representing vegetation types are: irrigated crops, rainfed crops, broadleaved forests, needle-leaved forests, shrublands and grasslands. The regional maps were produced by modifying the GlobCover product from the MEdium Resolution Imaging Spectrometer (MERIS) satellite, into the MyWater nomenclature. The local maps were produced by using spectral classification techniques on digitally enhanced Landsat multispectral images and field surveyed data. The maps were provided in common raster format (.img and .tif).

3.2.3.1 GlobCover data

The GlobCover project provides global land cover maps, using as input data the MERIS sensor on board of the ENVISAT satellite mission, characterized by a spatial resolution of 300 m. The composites were created through the pre-processing module of GlobCover, including corrections related to the detection of clouds and the effects of the atmosphere, geo-localisation and re-mapping. The GlobCover product is considered to be the highest resolution (300m) Global Land Cover product ever produced.
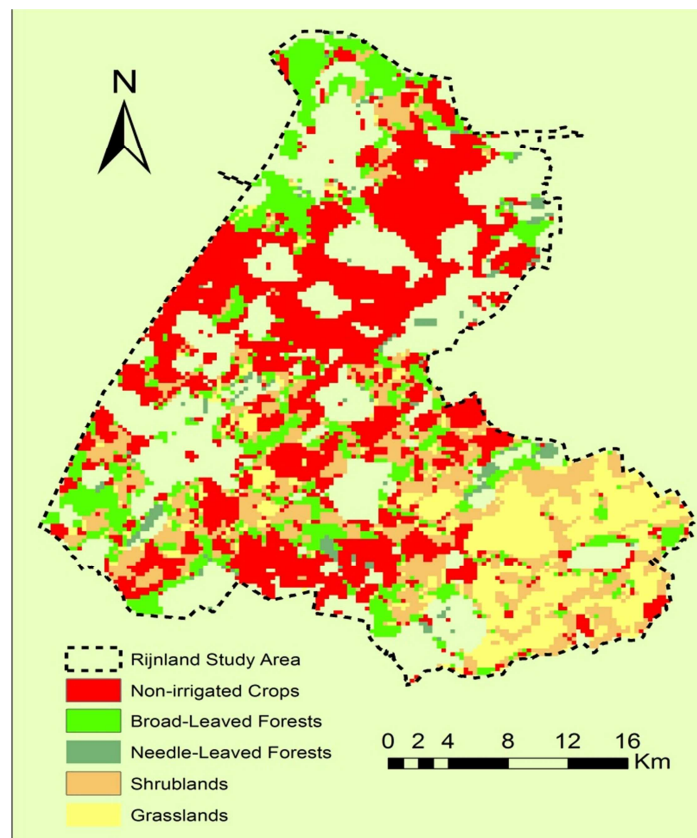
Figure 10: *GlobCover land-cover map of the Rijnland study area.*

A global image of GlobCover was acquired during the MyWater project for further processing, in order to create land-cover maps at a 1:1,000,000 scale

(regional level). The image was then clipped to each study area extent and a proper UTM/WGS '84 projection system was assigned to each image, according to the UTM zone each study area is located. Pre-processing of the image included geometric correction, cloud masking, atmospheric correction, mosaicking, and classification based on 4 steps; per-pixel classification algorithm, per-cluster temporal characterization, per-cluster classification algorithm, and labeling-rule based procedure. An example for the GlobCover map created for the Rijnland study area is demonstrated in Figure 10.

3.2.3.2 Landsat data

Landsat land-use/land-cover data has been created through the acquisition of Level 1 data of the Landsat 5 sensor at a local scale (1:50,000) during the progress of the MyWater project. The pre-processing of the downloaded Landsat 5 data included five methodological steps in order for the data to be prepared for the creation of the high-resolution land-cover maps. The first step was to ortho-correct the images according to the UTM/WGS 84 projection system, in order to be compatible with the other datasets used throughout the research and according to the specific UTM geographical zone of each study area for each relevant image. Then, pixel values (DN) were converted to at-sensor radiance using information from the image metadata. Atmospheric correction was then applied to scaled surface radiance after using a dark object subtraction (DOS) approach. The DOS approach is a simple and efficient method for atmospheric correction as shown in relevant studies (Song et al., 2001). Conversion to surface reflectance was implemented using two topographic corrections; the first was to use the improved cosine correction and the second to use the Lambert surface assumption. As a final step followed for the pre-processing of the land-cover images, the mosaicking of the images into a single image was performed, in case several images were needed to cover the selected study areas.

After pre-processing, the Landsat 5 TM images were ready to be used for the creation of the land-cover maps. The basic criteria addressed for the final choice of the images intended to be used, were the total percentage of cloud contamination on each image (as low as possible) and the period (season) the image was taken, in order to take into account as many phenological stages of vegetation growth as possible. Due to the frequent cloud presence, spectral and textural features of the available Landsat 5 images were extracted, whenever these were available and cloud free, and merged with the initial satellite images, resulting into various composite outputs.
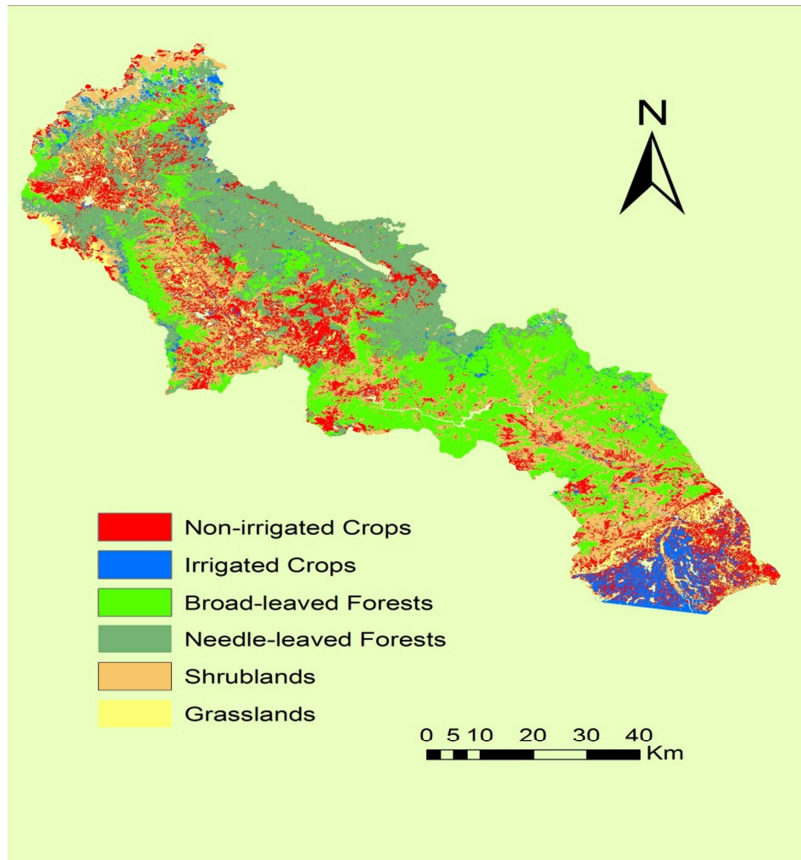
Figure 11: *Landsat 5 land-cover map of the Nestos study area*

As a next step, it had to be ensured that the classification accuracy of the to-be-created land-cover maps, would be the best possible. A pixel-by-pixel sampling method has been tailored for this purpose, resulting into 8,500 spots being applied to the previously created composite images, producing a spectral signature for each class. A sum of eight intermediate thematic maps was created using the spectral signatures. Maximum-likelihood classifier was used in a supervised classification, achieving fusion of the thematic maps and improving the classification accuracy. The thematic maps were then merged into a single composite image and a majority filter of 3 by 3 pixels was applied on the classified image. This process was repeated for each image available by mixing the used seasonal dates (e.g. summer-autumn), and by doing so, it was feasible to take the changes existing in vegetation phenology into consideration.

An additional classification technique was used during the process of polygon sampling which was explained above, called the Multi-Stage classification. This is also a classification method applied per-pixel in which the pixels are being automatically assigned by the classifier to the class they resemble the most. In order to follow this approach, the values representing information classes, e.g. a land-cover type, had to be identified by a human operator in each image.

In order to specify training and testing samples, several techniques have been used. Each training site was assigned an amount of at least 30 x waveband samples, created either by using a pixel-by-pixel sampling technique, or, by selecting sample polygons with a multitude of pixel units. The reasoning behind these choices was the intention to represent intra-class variability, as well as to inhibit problems related to dimensionality, which might occur when using parametric classifiers. Corine Land Cover (CLC) maps of each study area were used as the strata for the selection of samples, with the focus on choosing within each study area the center of polygons characterized as homogenous land cover areas.

Considering the methodology used to collect the training observations, there were two main techniques used. Random sampling design was the choice when sampling polygons were used, while a subset technique was used for the pixel-by-pixel technique, with 85% of the sample used as training sample and 15% used as validation sample. For the random sampling design, two hundred observations per area of interest were selected in order to attain the desired precision, with the acceptable error (ε) rate found to be 6.89% at the 95% confidence interval.

In order to improve the accuracy of the training and validation results, a procedure where the trainer assigned the observation in two land cover related fields was used. With this method followed, the ambiguous areas that might fall in two categories of land cover – always according to the visual interpretation of the tester – could also be taken into account, if one of the two land cover characterizations was the same as the one indicated with the classification process result.

The validation samples were used as the means to evaluate the accuracy of each intermediate map created. A threshold of 80% was set by the MyWater scientific team to be the limit below which a map was rejected from further processing.

An example of a land-cover map created at Landsat spatial resolution for the study area of Nestos can be found in Figure 11.

3.2.3.3 Corine Land Cover Data

CORINE (Coordination of Information on the Environment) Land Cover (CLC) is a map representing the European environmental landscape and its creation was based on the visual interpretation of satellite images (example of a CLC map for the Rijnland study area is demonstrated in Figure 12). The data provided by CLC is mainly in the format of digital land-cover maps, covering almost all the European continent. The European Environment Agency (EEA) and the Joint Research Centre (JRC) are responsible for the management of the CLC. The project was initiated in 1990 and since then several updated versions have been made available. The most significant data source for the creation of the CLC maps until the version of 2006 were ortho-rectified images acquired through the Landsat 7 satellite, with a spatial resolution that reaches 12.5 m. Since the version of 2006 and for the versions that followed,

SPOT satellite data started to be used instead of Landsat. The CLC provided scale is 1:100,000, where the minimum mapping unit is 25ha, with all land cover changes larger than 5ha being mapped. The data of CLC are freely available on the following website:
 http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version.

The CLC product is using a 3 - level hierarchical nomenclature, with 5 classes participating on the first level, 15 classes on the second and 44 classes on the third level.
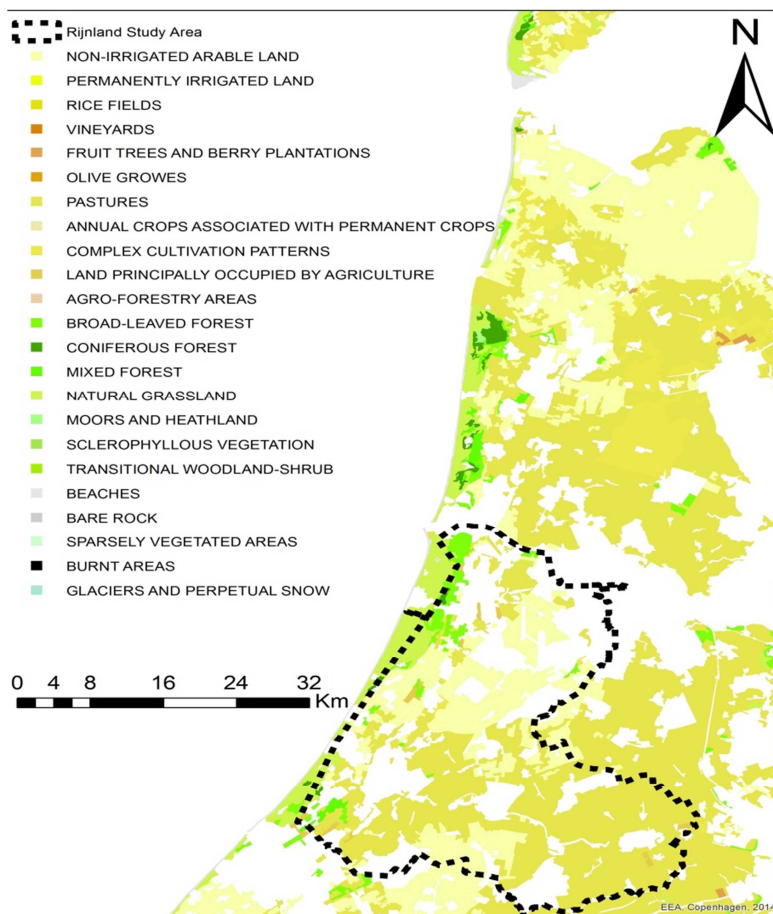
Figure 12: *Corine Land-Cover map for the study area of Rijnland.*

During the progress of the MyWater project, CLC maps of 2006 were used for the study areas such data was available (unfortunately, Greece was not included in these areas), in order to fill-in data gaps of GlobCover and Landsat data. The classification of ambiguous areas and the validation of the classification process, when this was needed, were the additional tasks in which CLC data had been used.

3.2.4 Field data

During the progress of the MyWater project, field measured LAI data was acquired for each study area in time-periods characteristic of the vegetation phenological stages and within the time-series of the MODIS satellite images downloaded. An overview of the dates field measurements took place can be found in Table 4. In the present Thesis, these measurements were intended to be used for the validation of the LAI maps created at the Landsat level (pixel resolution 30 m), since they are considered to be of very high accuracy.

Table 4*: Available field measurements per date and study area.*

Study area	Date	Measured locations
Nestos	3-5/11/2012	27
	9-10/04/2013	27
	2-3/07/2013	27
Rijnland	26-29/06/2013	25
	23-26/09/2013	25
Queimados	6/06/2012*	31*
Tamega	20-23/05/2013	39
	2-5/09/2013	39
Umbeluzi	7-9/01/2012	25
	12-14/03/2013	22
	1-3/07/2013	24

* Date not included within the research time series.

The data acquired during the field survey consider Land Cover Land Use data (LULC), and Leaf Area Index (LAI) data. Representative sites in each study area were selected by the scientific team responsible for the MyWater project, covering all possible types of vegetation that could be present in each study area.

For the sampling design, field surveys were performed in order to collect data for the training, calibration and validation of the products and models developed during the MyWater project. Therefore, Land Cover Land Use (LULC) data, as well as LAI data was collected in representative locations of the study areas. For the selection of a suitable sample design strategy for the research purposes, four key characteristics as identified by Brogaard and Ólafsdóttir (1997) were used:

- The number of sample sites required within each study area.
- The spatial distribution of the sample sites within each study area.
- The required size of each sample site selected.
- The number of subplots required within each sample site.

Considering the spatial distribution of the sample sites selected, the stratified random sampling approach was preferred rather than a random sampling approach,

in order to maintain a high level of accuracy and additionally reduce the amount of samples. The design approach followed for the strata in all study areas ensured the reduction of variability within the strata, by identifying the factors defining variability for each parameter examined and overlaying them with GIS analysis. As a next step, non-vegetative classes and their associated areas were discarded from the resulting maps, along with areas evaluated as too small to be considered as a sample. Then, the vegetative classes with similar characteristics were merged into a more general class, with the final result being the actual strata map.

The representation of ground data at the scale of the image, in order to ensure an adequate range of data for accuracy testing, has been described by Curran and Williamson (1986) as a characteristic of significant importance for studies related to remote sensing data. The final amount of sites selected as samples is highly dependent to the size of the study area, as well as to its spatial variability. Classical sampling theory indicates that the amount of sample sites required, in order to reach the desired precision, can be found by the equation

$$n = σt / ε,$$

 with n being the size of sample, σ the standard deviation, t the user's t value with n-1 degrees of freedom at a certain confidence level and ε being the acceptable error. A two-sided test was used with a 95% confidence level, while the spatial variability of the study area was expressed by the standard deviation of a number of spatially random measurements, extracted from a priori data of the study areas that was collected during past studies.

Following the guidelines of Townshend and Justice (1988) for the estimation of the required size of each sample site, the pixel size and the geometric accuracy of the satellite images used were taken into consideration, by using the formula:

$$A = (P (1 + 2G))^2,$$

where A indicates the sampling area, P the pixel size and G the geometric accuracy of the image (pixel amount).

As for the number of subplots required within each study site, which is an important parameter as mentioned above, it was evaluated as of high importance to ensure the spatial compatibility between ground data and pixel resolution. Therefore, a series of measurements (ranging from 3 to 15) was suggested to form a grid, a transect, a cross or a triangle within each sample site. The process followed was to measure the corners (vertices) of a triangle with dimensions equal to 2/3 of the satellite image's pixel size and then average the three results estimated. In cases of sparse vegetation or plantation, where great variability was expected, an approach of taking measurements on the corners and the center of each sample site was followed.

Considering the measured LAI data, a protocol for measuring LAI using hemispheric camera photographs was used in order to ensure that the quality of the

measurements would be as high as possible (Alexandridis et al., 2013b). The technique used to measure LAI was to estimate the canopy thickness by measuring the light transmission through the canopy. In order to ensure that LAI estimates would be comparable in time and space, it was considered essential to compensate for the variable sun irradiance throughout the day and year, as well as under variable tree canopies. Therefore, the exposure of the photographic camera used was calibrated relatively to the sky, and this exposure was increased by two stops during image acquisition under the canopy (Welles and Norman, 1991). Underexposure in near horizontal directions was avoided by using a zenith angle range between 0 to 75° (van Wijk and Williams, 2005). In cases of low stature vegetation, where distortions might be present to the LAI estimated values (van Wijk and Williams, 2005), the hemispherical photographs were taken from above in a downwards direction (Bréda, 2003). Figure 13 shows the camera used during the field measurements during the process of taking an upward photo (left part of the figure), as also a downward photograph taken in the study area of Queimados (right part).

Figure 13: *Measuring LAI in the study area of Queimados.*

This process resulted into several hemispherical photographs of upward or downward direction, which were later processed with specialized software, but with a different approach for each of the two cases. The upward direction photographs were processed based on gap-fraction identification, separating the canopy (dark) from the sky (light) pixels. The photos of downward direction were processed by using reflectance ratios, in order to identify the features on the photographs based on the spectral behavior of the examined wavelengths. The software used for downward photograph processing was CAN-EYE, freely available for downloading at: https://www4.paca.inra.fr/can-eye/. This is image analysis software used for the extraction of characteristics related to the structure of plant canopies, with the use of multiple digital RGB or binary images. The analysis of the downward direction photographs was performed by grouping each series of the available photographs based on the date each photograph was taken and by the light conditions present. An automatic class pre-selection was then performed, with the user identifying or correcting after this step the final sample of vegetation and non-vegetation pixels on a hemispherical photograph. Then, this sample was used for all the photographs of

each group, with an image classification process following, in order to acquire photographs indicating two or three possible states: vegetative state, non-vegetative state and mixed state. When this step was finished, the software automatically calculated LAI values in multiple photographs according to the grouping performed previously, with the final results provided in a spreadsheet.

Figure 14: *Hemispherical photograph in RGB and processing with specialized software.*

Considering the upward direction photographs, the HEMISFER software was used (Figure 14), which is also a hemispherical photograph processing specialized software:
http://www.wsl.ch/dienstleistungen/produkte/software/hemisfer/index_EN.
In this case, the pixels were first classified as sky (white) or canopy (black) by the application of a brightness threshold to each picture analyzed. The optimal threshold can automatically be calculated by the HEMISPHER software using several methods, but for the MyWater project the LAI-2000 method (Stenberg, 1996) was chosen to be the one used, as it was reported to be closer to a standard reference. For the next step, the light transmission (T) was calculated as a proportion of the white pixels present in the hemispherical photograph. Then followed the calculation of the average number of times that a light ray would touch the canopy when travelling a distance equal to the thickness of the canopy, providing this way the contact number (K), which is described by the equation: K = -cos Θ ln T. Finally, the K values were integrated over the rings in order to estimate the LAI values.

A list of the materials used during the LAI field measurements and the processing of the acquired field data, in order to estimate the LAI values in MyWater project study areas, can be found below:

- A Sigma 4.5 mm F2.8 fish-eye hemispherical lens
  http://www.sigmaphoto.com/shop/45mm-f28-ex-dc-hsm-circular-fisheye-sigma
- A Canon EOS 1000D photographic camera. The highest image resolution and quality available, according to the camera settings and capability, were used.

- A standard tripod allowing the camera to take vertical upwards and downwards images.
- A standard clinometer to measure the average slope at each sample location.
- A standard compass to measure the direction of the maximum slope (aspect) at each sample location.
- A standard handheld GPS receiver to measure the geographic coordinates at each sample location, or to find the pre-determined sample locations when such existed.
- Specialized software for processing the hemispherical photos, such as HEMISFER or CAN-EYE. According to HEMISFER, a high-quality JPEG file gives practically the same results as an uncompressed file (RAW format) and therefore high quality JPEG was the format chosen for the acquired hemispherical photographs.
- Surveying sheets for keeping hardcopy notes of the visited locations.

4   Methodology

Literature analysis and data identification and acquisition were the first steps for the formulation of a methodology plan, in order to address the suitable techniques and come up with an appropriate plan-of-attack. The study examined the relation between LAI and EVI with respect to the Day Of Year (DOY, actually representing a period of 8 (LAI) or 16 (EVI) days, composed to a single image) and the vegetation type of each satellite image acquired. Regression analysis was the means to examine the LAI-EVI relation. The equations estimated during the regression analysis of MODIS LAI-EVI, were used in a model along with Landsat EVI images, in order to produce LAI maps at a Landsat spatial resolution level. Spread sheet software was used to create and organize the databases on which statistical analysis was applied, using appropriate software (JMP7). The regression equations were applied on the Landsat data using GIS software (ArcGIS Desktop).

4.1   Data pre-processing

The data acquired for the Thesis purposes was not initially (when downloaded) in a format that could make direct use possible. This was because geographical information (such as a projection system) was missing in some datasets, while in other the format of the related files was not compatible with the used software or with the rest of the data. Therefore, pre-processing of the data was judged as a necessary step and was applied in order to successfully use the data and acquire results of as high accuracy as possible.

4.1.1   MODIS LAI-EVI images pre-processing

Pre-processing of MODIS data included the conversion of all the downloaded LAI and EVI images of each study area, to a format compatible with GIS software used (ArcGIS), the assignment of a proper projection system according to the related geographic zones each study site is located (UTM zones, WGS '84 datum), as well as the definition of areas of interest (river basins) in each study area and the trimming of images to these limits. Software provided by the LPDAAC platform (MRT4, https://lpdaac.usgs.gov/tools/modis_reprojection_tool) was used to automatically assign a proper UTM projection system to each associated image, depending on which study area it represented, as well as to automatically convert the images from their initial .hdf format to .tiff format. The downloaded MODIS datasets included more than one layer for each associated image, and therefore MRT4 was also used to extract only the needed layers for each image dataset. In this case, layers representing EVI and LAI data, as well as layers indicating quality characteristics – such as Quality Assurance (QA) and pixel reliability – were extracted for each image downloaded. Finally, vector files that were created to specify each study area using

ArcGIS software were used as masks, in order to clip each MODIS image to the appropriate extent.

4.1.2 Creation of MODIS sample points

In order to apply the regression analysis between the MODIS images of EVI and LAI data, a systematic procedure was defined, in which a file for each study area representing sample points was used as the means to extract LAI and EVI values per image pixel. Therefore, for each one of the study sites a point shapefile was created, using the center of the pixels in each image used (a single MODIS image was used per study area due to common geo-reference) as the location where the points should be placed (1000m vertical and horizontal distance per point, same as MODIS pixel resolution). However, because the MODIS images did not include only types of land-cover that were intended to be used in the regression analysis (such as any vegetation type), but also areas covered by water or urban areas, a further exclusion of pixels representing non-vegetative states was applied.

For the above reason, the point shapefiles created were used along with the GlobCover land-cover maps, in order to eliminate the sample points that would not participate in the regression analysis. GlobCover data uses a spatial resolution approximately 3 times finer than the MODIS pixel (300 m instead of 1000 m), so it was also feasible to determine the effect of mixed vegetation pixels and further exclude any sample points representing such a state. As a next step, the raster data was converted to vector format (polygons) grouped according to their vegetation classes, and then further converted to line format, in order to be used as buffer to exclude the MODIS sample points located closer than a distance of 300m (one GlobCover pixel) from any line present. By this exclusion, it was feasible to erase any sample points that might fall between different vegetation classes, and with the spatial resolution used, which was approximately 1/3 of the MODIS pixel resolution, it was feasible to retain the amount of excluded pixels to the minimum possible level.

4.1.3 Quality Control application

According to Myneni (2012), since the MODIS LAI algorithm is executed irrespective to the quality of input data for the MCD15A2 product, the Quality Assurance (QA) layer (FparLai_QC) should always be used to ensure per-pixel accuracy of LAI provided values. In the associated QA layer of each MODIS LAI image downloaded, the actual information about each pixel's quality is included in the SCF_QC bit field, which is used as the key indicator characterizing LAI quality, and in which field the bit patterns can be parsed from right to left, while individual bits within a bit-word can be read from left to right. Through the use of LDOPE software provided by the LPDAAC platform, it was feasible to develop and run a script (see appendix A for a detailed description of the software and the actual script code developed), which converts the information of SCF_QC bit field to actual pixel values, indicating the quality of the associated LAI values. For the study purposes, only pixels of a "good" quality were isolated for further use, excluding data of "marginal" and

"bad" quality, in order to ensure the highest quality possible in the associated results. From the remaining pixel values in the associated datasets, extreme values for LAI (<0, >20) and for EVI (<0.1) were excluded, as they clearly did not represent any possible vegetation state.

Testing of how linear regression analysis performs in each case of different quality datasets had to be applied on datasets of each study area, in order to ensure that the choice of using data of only "good" quality and thus, excluding in some cases a significant amount of samples (pixels), was justified and indeed ensured the high level of accuracy needed in the associated results, rather than just making datasets of certain 8-day images unusable, due to the small amount of samples remaining after the exclusion. Therefore, datasets representing only "good" and "good" and "marginal" quality data were tested through regression analysis between the LAI and EVI values, in order to indicate the influence of low and marginal pixels in the associated results of R^2 values.

4.2 MODIS LAI-EVI regression analysis

The methodology followed used regression analysis between the LAI and the EVI products of MODIS, in order to define the regression equation type representing the actual data used for each study area. Since relevant literature, as presented in the second chapter, was controversial over the type of equation better describing the relation between LAI and VI data, it had to be examined whether the data fit a linear, square, square root, or polynomial type of equation. The nature of the LAI-EVI relation was studied for the five study areas by evaluating the R^2 values, the p probability values, the mean difference between the examined values, the amount of pixels participating in each case of a regression analysis between EVI and LAI datasets, as well as by analyzing the results with the creation of density plots indicating the distribution of pairs of LAI and EVI values on the x and y axes. The datasets used for this purpose represented each vegetation type found in the associated areas, and moreover the selection of the datasets was done according to the influence of seasonality, by selecting 16-day images representing different vegetation states throughout the period of a year.

When the final choice of the type of equation that would be used for the EVI-LAI regression analysis was decided, the LAI and EVI pixel values from each pair of images were exported in tabular form per test site, including information over vegetation type and the 16-day period date it came from. This process led to the creation of the final tables that were going to be used for the regression analysis, in order to examine the relation between LAI and EVI, and derive the relevant regression equations with respect to vegetation type per study area. Moreover, the relevant regression equations were estimated per DOY to identify any seasonal effect, as well as per DOY and vegetation type together, in order to examine the overall accuracy of the associated results in each case examined and decide which

approach is the most precise. The phenological phases of vegetation types were used to explain the temporal variation of the relation between LAI and EVI.

4.2.1 Total regression for whole year

The tables with the EVI and LAI values within each study area were merged into two tables, one representing the EVI data and the other representing LAI data for the whole year of the research period. To achieve this, beside the fields indicating the sample point ID, the LULC, and the LAI or EVI value, an extra field was added indicating the DOY each dataset represented. Although relevant literature suggests that each vegetation class should be examined individually, it was evaluated as interesting to compare the results of this case with the results of the other cases explained in the following sub-chapters of the Thesis. Plots of the EVI and LAI values on x,y axes and the fit of each dataset by a linear equation were the output of this step, along with statistics over the mean difference, R^2, f probability and Root Mean Square Error (RMSE) values of the regression analysis results.

4.2.2 Regression per DOY

In this case, regression analysis was applied between LAI and EVI datasets of each study area without taking into consideration the different types of vegetation. However, this time the datasets were used with respect to the DOY each dataset was associated with. The anticipated results were similar to the total year regression scenario presented above, but this time, the results could be used to evaluate the effect of seasonality in each study area's data, and moreover, to get a clearer picture of which periods during the different phases of vegetation (in each study area) were more probable to produce erroneous or inaccurate data.

4.2.3 Regression per vegetation type

The effect of vegetation type on the regression equations is pointed out by relevant literature as the most important parameter to be taken into consideration, when the criteria of regression analysis between LAI and EVI data are decided. In this case, the datasets were classified according to the vegetation class they belong to and then regression analysis using a linear equation was performed between the datasets of LAI and EVI. The same parameters and statistical values were examined as in previous cases.

4.2.4 Regression per vegetation type & DOY

In this step, the relation of EVI and LAI was decided to be examined by taking into account both the characteristics of seasonality (DOY) and vegetation type, in order to examine the level of improvement in the associated regression analysis results and estimate the equations that would later be used for the downscaling methodology of LAI MODIS maps to the Landsat spatial resolution level. Therefore, the datasets were once again classified according to these characteristics and regression analysis was applied for each combination of LAI and EVI datasets

available, per DOY, vegetation type and study area. The results of this case were in the same form as in previous cases in order to be comparable.

4.3 Comparison of Landsat and MODIS spectral bands

Several studies have demonstrated the similarities between the MODIS and Landsat band wavelengths as presented in chapter 2, showing that the data from these two sources is directly comparable. However, since there are minor differences between the band values of the two sensors (Table 1), it was evaluated as necessary to proceed in a comparison between the level 1 products of the 2 satellites. Therefore, for two of the study areas (Umbeluzi and Tamega) -in order to ensure that the results would not be valid just for one case-, images of the same period were downloaded representing pixel values of red, blue and infra-red bands for both Landsat and MODIS satellites. The choice of these particular bands was done according to the EVI equation, which uses values of these bands for the calculation of EVI values. Unfortunately, the EarthExplorer platform did not provide any MODIS surface reflectance data for the needed bands in a spatial resolution of 1000 meters, and for this reason images with a pixel of 500 meters that were available were downloaded. The pre-processing needed for the MODIS images was the same as with the rest of the MODIS images used throughout the Thesis (assignment of UTM projection system/datum, clip to the study area limits through ArcGIS software), while for the Landsat data downloaded, the needed layers representing red, blue and near-infrared bands were extracted as single images from the compressed file downloaded and then the same procedure as with the MODIS pre-processing was followed once again.

The overall goal of this step was to prove that the pixel values between each equivalent image (MODIS red compared with Landsat red band image, blue with the equivalent blue, etc.) show a high level of correlation in their corresponding pixel values. Correlation analysis was the method used to prove the above hypothesis. However, in this case the data of the two satellites was not directly comparable, due to the different spatial resolution each downloaded product used (500m pixel for MODIS and 30m for Landsat). Moreover, since this study uses a pixel of 1000m for the MODIS LAI and EVI data, the MODIS red, blue and infra-red data with a spatial resolution of 500m should also be converted to the same spatial resolution level with the rest of the data. It was decided to compare both the 500m (initial format) and 1000m versions of MODIS images with the Landsat images, in order to also evaluate the scaling effect between the data versions. Therefore, a methodology was developed during which:

   i.   The clipped Landsat images were used as mask to clip the MODIS data.
   ii.   One MODIS image of 500m and one of 1000m were used to create point shapefile layers (1 point per pixel) for both study areas examined.

iii. One MODIS image of 500m and one of 1000m were used to create fishnet-shaped layers (1 square polygon per pixel) for both study areas examined.

iv. The cloud and low pixel reliability mask (CF_Mask) of Landsat, was used per image of interest (both MODIS and Landsat images) to extract ambiguous areas from the images. Additionally, areas covered by non-vegetative land types (water bodies, rocks, etc.) were also extracted.

v. The fishnet-shaped layers of 500m and 1000m were used as masks to calculate zonal statistics (ArcGIS) for all the images (MODIS and Landsat) examined. By doing this, each image was converted to a spatial resolution relative to the fishnet layer used each time, with the resulting pixels representing the mean value of the pixels included in each fishnet block (pixel). Moreover by following this procedure, all the resulting images were characterized by the exact same spatial geometry, making the pixels directly comparable with each other.

vi. The CF_mask layer of Landsat was used once again to extract ambiguous and non-vegetative areas per image and study area of interest, in the resulting images.

vii. The point shapefiles created during the second step of this methodology, were used as masks to extract values from each corresponding image, in order to create data tables with the values of each pixel in the associated images.

After the application of the above mentioned steps, it was feasible to apply correlation analysis between each associated data table created (MODIS and Landsat 500m, MODIS and Landsat 1000m). As an additional step, it was decided to apply a similar methodological routine as described above between the EVI images of Landsat and MODIS, in order to compare directly the downloaded data used for the Thesis and analyze their correlation levels. However in this case, since the EVI data of MODIS was available for downloading directly in a 1000m spatial revolution, the in-between steps used for the 500m spatial resolution were excluded from the process.

4.4 Downscaling of MODIS LAI maps

In order to downscale MODIS LAI maps to the spatial resolution of Landsat (1000m to 30m), the regression equations estimated for MODIS EVI and LAI in chapter 4.2, were used on Landsat EVI images, in order to estimate LAI values per pixel. The application of the MODIS regression equations on the Landsat EVI maps was done per vegetation category and per date of interest, as these were the characteristics decided to be used after testing the results for several parameters (chapter 4.2). Landsat derived land-cover maps (30 meters pixel, available from MyWater project) were used as the means to indicate the land-cover type of each pixel per image and study area. The result was the creation of LAI maps at Landsat

spatial resolution level, for the dates Landsat EVI images were available (seasonal coverage).

4.4.1 MODIS LAI downscaling model development

The methodology used was quite simple (Figure 15); a model was created in ArcGIS software per area of interest, which used as initial input the available Landsat EVI images, one per each run of the model. For each image of interest, its corresponding CF_Mask layer was initially used to exclude pixels of low quality and not representing a vegetation state, and then using the Landsat land cover maps, each pixel was assigned an additional value indicating the vegetation class it falls in. As a next step, the pixels of each image were divided into several images, with the amount of images depending on the vegetation classes present in each study area. Then, the equations calculated per vegetation class and date of interest during the MODIS LAI-EVI regression analysis were used as formulas to estimate the LAI values per pixel, using the corresponding EVI values along with the rest of the factors participating in each equation.
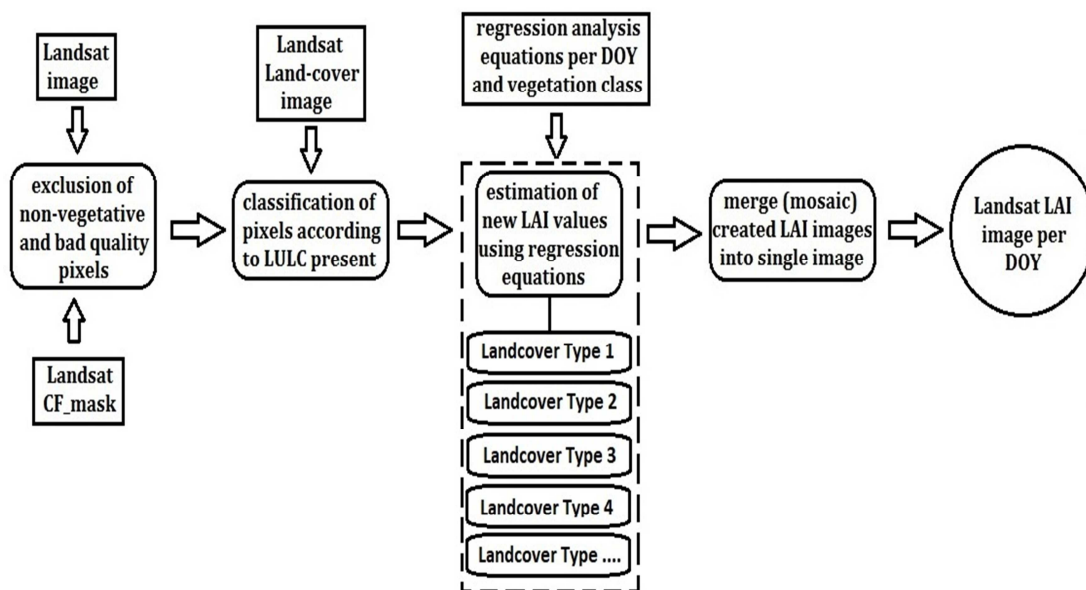
Figure 15: *Flowchart of MODIS LAI downscaling model*

The resulting images were then mosaicked to a single image, which represented the LAI values in the Landsat spatial resolution level, for the exact same areas as in the Landsat EVI images.

4.5 Comparison of Landsat LAI with in-situ measured LAI

The final step of the methodology was the comparison of the LAI values estimated at Landsat spatial resolution level with the actual LAI values measured in-situ during the progress of MyWater project (described in chapter 3.2.4), for dates as

close as possible to the period the field measurements were acquired. Therefore, point shapefiles per date that field measurements took place and per study area were created, in order to be used as masks and extract the associated Landsat LAI values per point of interest. With the resulting data tables indicating in-situ measured and Landsat estimated LAI values, as well as the land cover type present per point of interest according to in-situ observations and according to the Landsat land cover maps, it was feasible to apply correlation analysis and evaluate the results of the downscaling methodology. The results were in the form of plots indicating the level of correlation between the LAI values, as well as in the form of statistics indicating the r-value, the mean difference, the p probability and the RMSE.

5 Results

5.1 Pre-processed data

Pre-processing of the images and in general the data intended to be used in GIS software environment, led to the geo-referencing process of each dataset used, according to the study area located. The datum assigned to all datasets was WGS 84, while the projection system assigned to each dataset was the Universal Transverse Mercator (UTM) system. Depending on the location of each study area (latitude-longitude), the UTM projection system is divided in zones (North and South of the equator) and when it is assigned to a dataset, a choice has to be done which zone includes the area of interest. Therefore, all the data related to the Nestos study area were assigned a 34N (North) UTM system, a 31N was assigned to Rijnland data, a 29N to Tamega data, a 29S (South) to Umbeluzi data and a 23S to Queimados data.

5.1.1 MODIS LAI and EVI pre-processed images

All MODIS LAI and EVI datasets, including their quality control data, were clipped to the same extent depending on the study area each dataset was located. The result of the preprocessing of the images was 23 LAI and 23 EVI images representing 16 – day periods for each study area, and equivalent images indicating pixel quality characteristics for each image available. The total amount of the images pre-processed was 230 for the EVI-LAI data and 230 images for the Quality control data.

5.1.2 Sampled MODIS pixels

The methodology followed for the creation of the MODIS sample points, resulted into 5 point shapefiles (one per study area), which represented the locations that would be used as samples for the regression analysis with points located in the center of each associated pixel. Figure 16 displays the point distribution in the Tamega and Nestos study areas. In this figure, the density of sample points in vegetation areas characterized by heterogeneity or mixed vegetation pixels (e.g. Southern part of Tamega study area) was limited compared with homogenous areas characterized by a single type of vegetation (e.g. central and southern parts of Nestos study area).
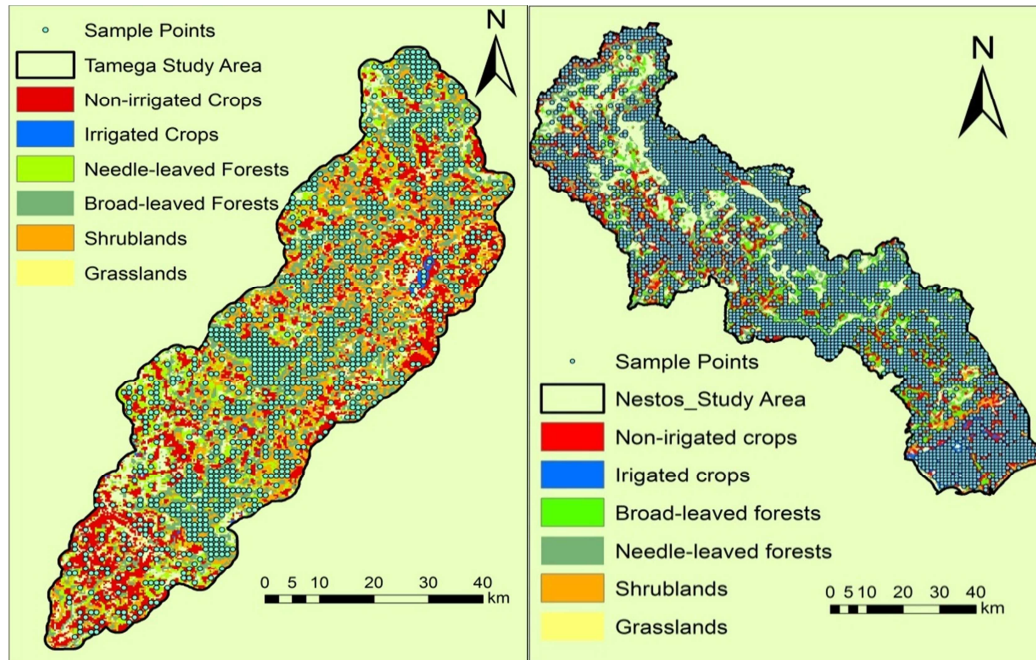
Figure 16: *Examples of MODIS sampled pixels in Tamega (left) and Nestos (right) study areas.*

These point shapefiles were used per study area as mask layers, in order to retrieve EVI, LAI and LULC values through ArcGIS software, from all the associated images. The geometric matching between MODIS LAI and EVI data was identical after processing the images with the MRT4 software, and with the technique used for the exclusion of mixed pixels in the GlobCover land-cover maps (chapter 3.2.3), it was ensured that the dominant vegetation type per sampled pixel would be selected. The data were then stored in 5 tables of .xls format (one per study area) classified per pixel Identification Number (ID) and also including the DOY of the image each value was retrieved from. The total amount of the MODIS sampled pixels per study area can be found in Table 5. There is a variable representation of samples following the distribution of vegetation types. Although some classes have more than 100 samples in some study areas (e.g. rainfed (non-irrigated) crops, broadleaved forests, shrubs), other are not well represented (e.g. irrigated crops, grasslands).

Table 5: *MODIS sampled pixels per study area and vegetation type.*

Site	Rainfed	Irrigated	Broadleaved	Needleleaved	Shrubs	Grass
Nestos	572	143	808	704	5	7
Queimados	164	0	31	14	28	49
Rijnland	170	6	85	675	215	4
Tamega	1673	33	263	-	378	-
Umbeluzi	112	-	709	-	311	360

5.1.3 Impact of Quality Control

Testing of how linear regression analysis between MODIS EVI and LAI performs in each case of different quality datasets (all available data selection and only good quality data selection) was applied and showed that in most cases there was a slight improvement in the R^2 (coefficient of determination) values, ranging from 0.5% – 5%. In case of the Tamega study area the R^2 value showed maximum improvement with the percentage reaching almost 10%, as presented in Table 6.

Table 6: *Comparison of regression analysis results for all study sites between all data and only good data, according to QA criteria. N is the total amount of pixels used and p is the probability value.*

Site	All data			Good data		
	R^2	N	p	R^2	N	p
Nestos	0.477914	53736	<.0001	0.46158	38900	<.0001
Queimados	0.501308	56328	<.0001	0.570605	45398	0
Rijnland	0.269199	6840	<.0001	0.264657	2210	<.0001
Tamega	0.26711	27720	<.0001	0.356649	22669	<.0001
Umbeluzi	0.643699	48672	0	0.670594	39508	0

Judging from the excluded pixels in the Tamega study area (Figure 17), it was evident why there was such an improvement in the R^2 values, since a large amount of the excluded pixels were characterized by low LAI values (<2) and relatively high EVI values (>0.3), resulting into an improvement of the linear character of the equation.

There were also two cases (Rijnland and Nestos) where on the contrary to other study sites the application of Quality Control process decreased the overall value of R^2 for a percentage of 0.5% - 1.5%.

In case of Rijnland (Figure 17), the amount of available pixels participating in the regression analysis was limited, compared to other study areas (exclusion rate reached 60%), due to the small sized study area. Therefore, it was judged that this case was not the most appropriate to express conclusions safely, even though the overall difference in R^2 values was negligible. Finally, the application of Quality Control process eliminated the already limited sample points in grasslands for the Umbeluzi study area.
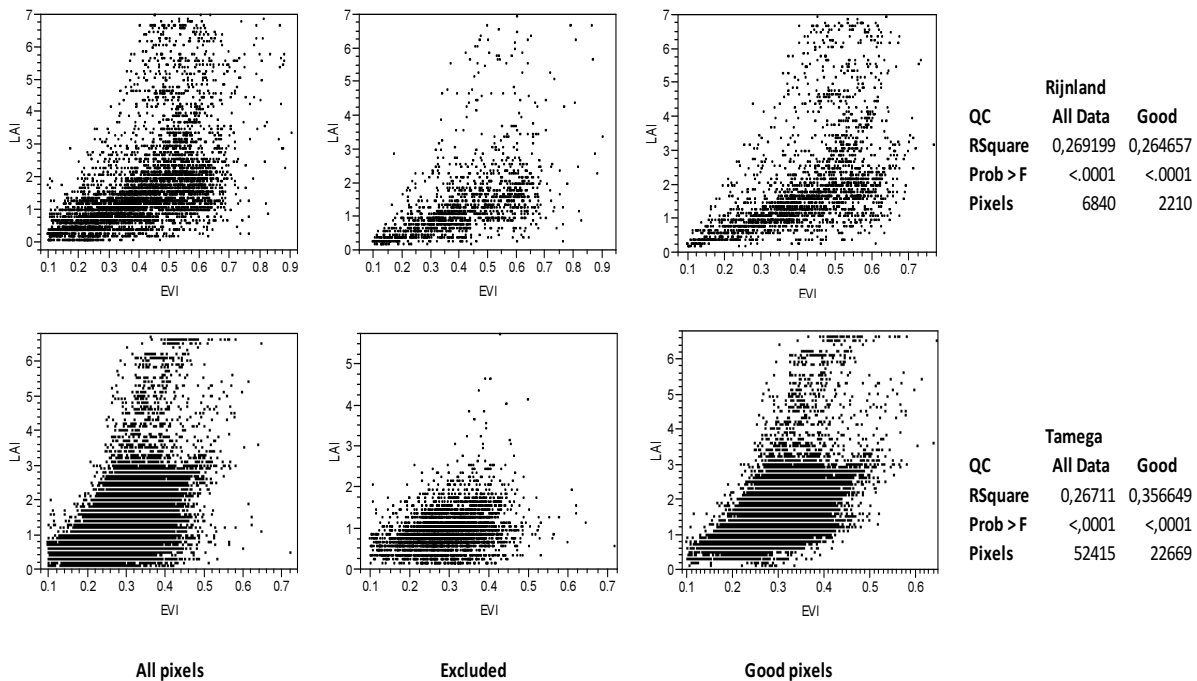
Figure 17: Example of a*pplication of Quality Control (QC) to sample points of Rijnland (1st row) and Tamega (2nd row) study areas.*

Although the Quality Control process did not provide significant improvement of the regression analysis results (best case was the Tamega study area with an improvement of approximately 10% for the R² result), it was decided to apply the QC, since in the majority of cases examined there was a slight improvement. Moreover, this process (according to relevant literature as demonstrated in chapter 2) was expected to increase the confidence level and the accuracy of the associated results.

5.2 MODIS LAI-EVI relation type

Regression analysis (y=LAI, x=EVI) was applied for all good data of each study area's time series of images. The results showed that the relation was best described by a linear type equation for all study areas, rather than other types of equations (an overview of the results per study area and per equation type tested can be found in (Table 7).

Linear equations returned the best values of R² in the majority of study areas, except from the cases where a polynomial equation was used, which showed slightly improved results in all cases compared to other regression types. Considering the regression results when a power or square root equation fit was used, it was judged that since they showed lower R² values and higher RMSE values for the majority of the examined cases, they should be rejected for further use in the Thesis methodology.

Table 7: *Regression analysis results for all tested equation types per study area.*

	Equation Type							
	Linear		**Square**		**Square root**		**Polynomial**	
Study Area	**R²**	**RMSE**	**R²**	**RMSE**	**R²**	**RMSE**	**R²**	**RMSE**
Nestos	0.4616	1.1412	0.4395	1.1644	0.4609	1.1419	0.4678	1.1392
Queimados	0.5706	0.5633	0.5281	0.5905	0.5664	0.5661	0.5720	0.5624
Rijnland	0.2647	1.3010	0.2497	1.3141	0.2637	1.3019	0.2658	1.3003
Tamega	0.3566	0.6610	0.3568	0.6609	0.3488	0.6650	0.3591	0.6597
Umbeluzi	0.6706	0.5866	0.6827	0.5756	0.6440	0.6098	0.6842	0.5743

**Probability (p) value is <.0001 or 0 in all cases and study areas

The polynomial trend lines had a similar fit to the data as the linear one, as presented in Figure 18, also evident from the distribution of the points in density plots (Figure 19). Since the overall improvement of the R^2 values was only marginal (best improvement shown in Umbeluzi with a rate of 1.4% or 0.014 in absolute values), it was judged that there was no need to choose such a complex type of equation, compared to the simpler linear equation. The RMSE in all study areas remained at similar values in both cases (linear and polynomial).
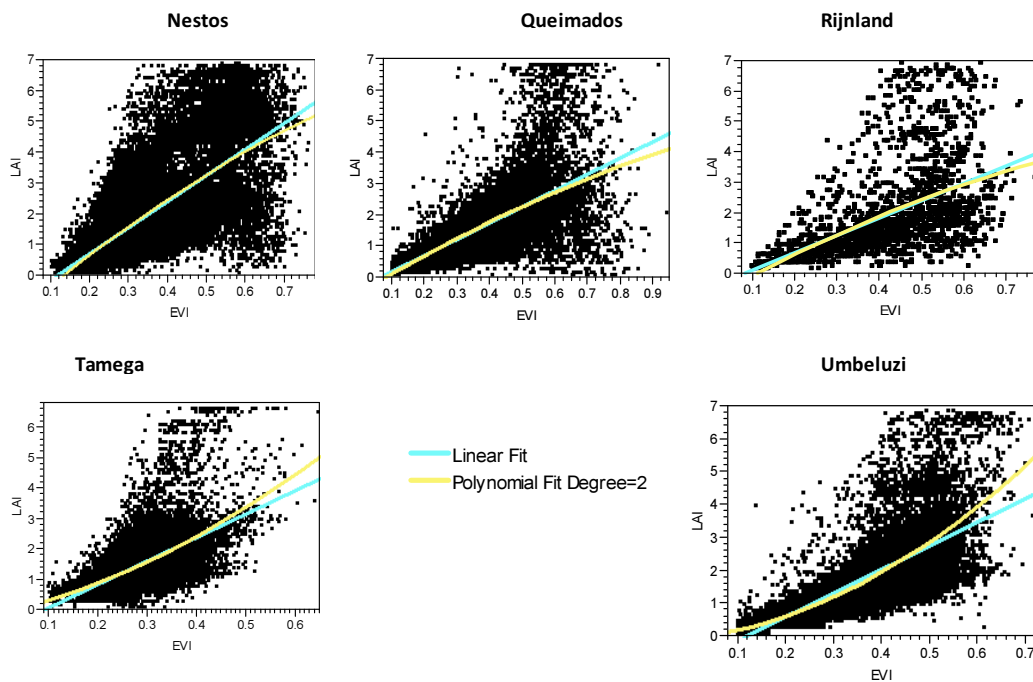
Figure 18: *Overview of linear equation (light blue color) and polynomial equation (yellow color) fits for all good LAI(y) – EVI(x) data per study area.*

The density plots visualizing the distribution and density of the sample points on the x and y axes (Figure 19) showed that indeed the distribution of sample points was characterized by a relative linearity, especially in the study areas of Umbeluzi, and Queimados.
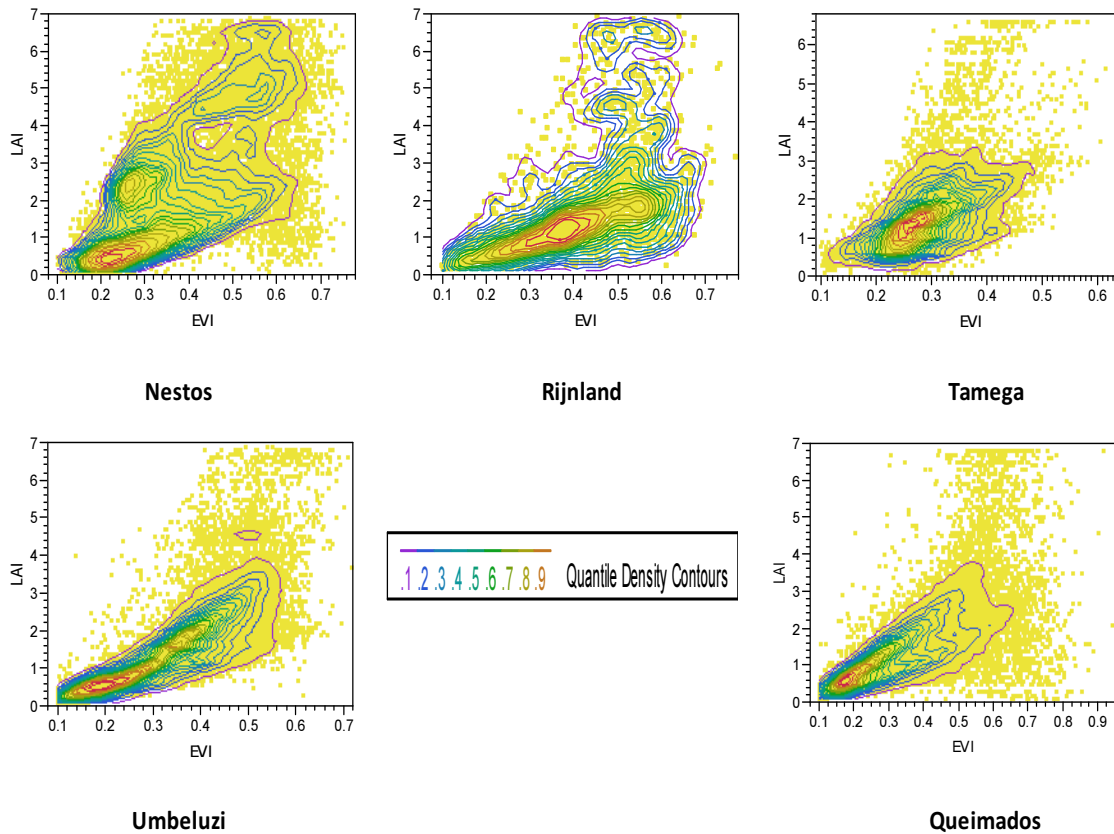
Figure 19: *Density plots of point distribution for all LAI(y) –EVI(x) data per study area.*

5.2.1 Relation for whole year

A linear regression equation was applied to all good data pixels per study area (Table 8). The coefficient of determination showed a fluctuation from relatively high values in Umbeluzi and Queimados, to medium values in Nestos, when in Rijnland and Tamega only a small percentage of the variability was accounted for by the model. All results were statistically significant, as shown by the probability (p) value. The slope of the equations was higher in Nestos, Tamega and Umbeluzi, demonstrating a higher sensitivity of LAI to EVI changes.

Table 8: *Linear regression equations per study area for all good data. N is the amount of sample points, R² is coefficient of determination, Intercept is the expected mean value of Y when all X=0, slope shows the steepness of the linear model and p stands for the probability value.*

Site	N	R²	Intercept	Slope	p
Nestos	38900	0.4615	-0.9416	8.4353	0
Queimados	45398	0.5706	-0.2975	5.1678	0
Rijnland	2210	0.2647	-0.4886	5.7599	<.0001
Tamega	22669	0.3566	-0.6849	7.6922	0
Umbeluzi	39508	0.6706	-0.8298	7.1751	0

Analysis of the density plots showed that some points (not the majority) were drifting away from the regression line, towards high LAI, while the EVI values remained at lower levels from the expected according to the linear fit, as demonstrated in the example for the Queimados study area (Figure 20).
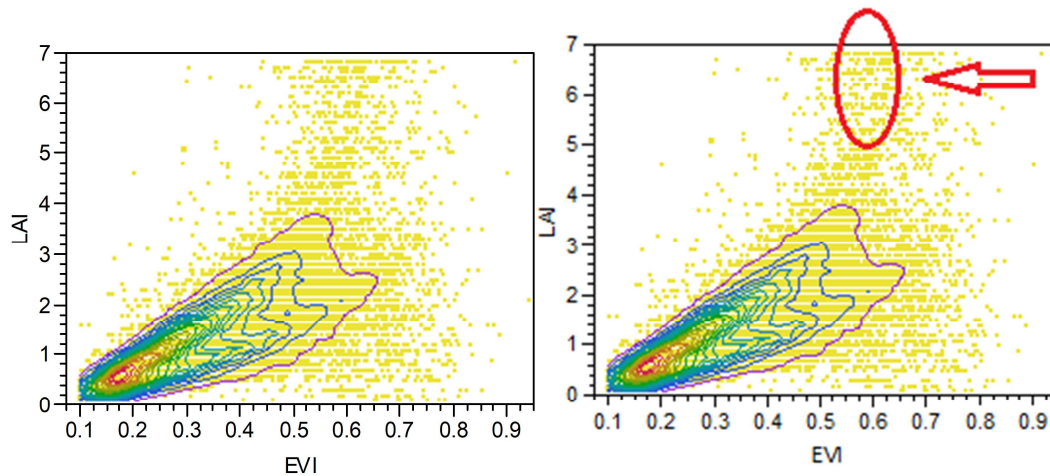
Queimados

Figure 20: *Density plot of all good EVI(x) and LAI(y) data for the Queimados study area and possible signs of saturated values indicated with red color on the plot.*

5.2.2 Relation across time

While examining the EVI-LAI relation per DOY, significant changes in the results of the regression analysis were revealed, mainly during periods of extreme meteorological conditions (i.e., heavy cloud/snow cover), when the excluded data rate reached a level of 100% in extreme cases (e.g. Rijnland from the end of October until February- Figure 21, when some parts of the Rijnland line is missing). When the

amount of excluded sample points did not exceed a rate of approximately 50-60% per DOY, the regression analysis results remained at similar levels or were improved per study area, compared to the regression analysis for the whole year (section 5.2.1).

For the dates that the exclusion rate of pixels was >60%, the majority of R^2 values when examined across time were lower than the values when examined undivided for the whole year. In most cases (Nestos, Tamega, Umbeluzi) the difference of the values fluctuated between 20-30% and never dropped under 10%. In most areas, R^2 values during November, February and March were characterized as low (Nestos, Rijnland, Tamega - Northern Hemisphere), however, there were also random dates during the rest of the year (e.g. Rijnland early summer) when low values were also present (Figure 21). Since Rijnland is characterized by a relatively flat landscape and shallow groundwater level, it is highly possible that extreme fluctuations could be also caused by water inundation that could occur after heavy rainfalls and in areas with poor drainage, which could affect the accuracy of LAI and EVI values. Agricultural areas and grasslands are more likely to be affected by such conditions. Moreover, the relatively small field size and the long and narrow shape that often characterizes the agricultural areas located in Rijnland, could also lead to erroneous identification of the vegetation type of such areas as mixed vegetation.
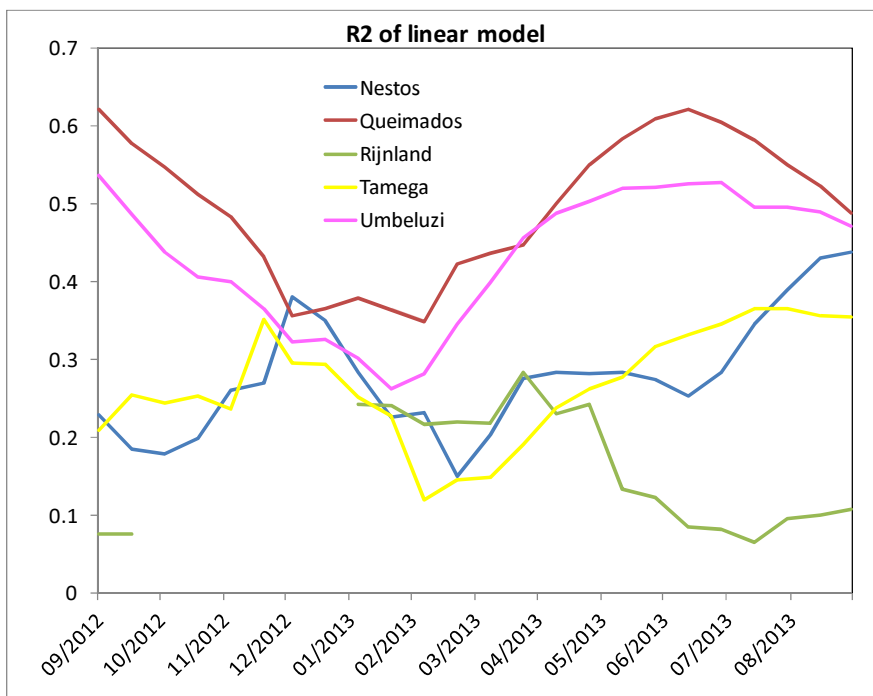
Figure 21: $R^2(y)$ per DOY(x) and study area
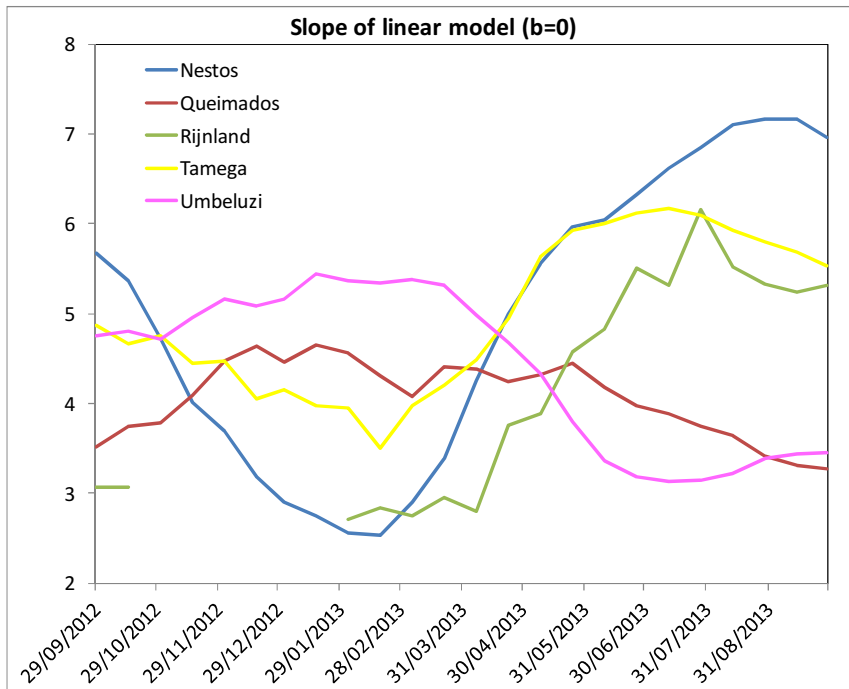
Figure 22: *Plot of slope(y) when intercept =0, indicating the steepness of the linear model per DOY(x) and study area.*

The analysis of the slope of the model (Figure 22) showed that sensitivity of LAI to EVI reached a maximum level between November and February for the study areas located in the Southern Hemisphere, and between June and August for those located in the Northern Hemisphere.

During consecutive DOY in certain periods for each study area, there were notable changes in the patterns of the EVI and LAI scatter plots, which means that the EVI-LAI relation could dramatically change from a 16-days period to the next (Figure 23).
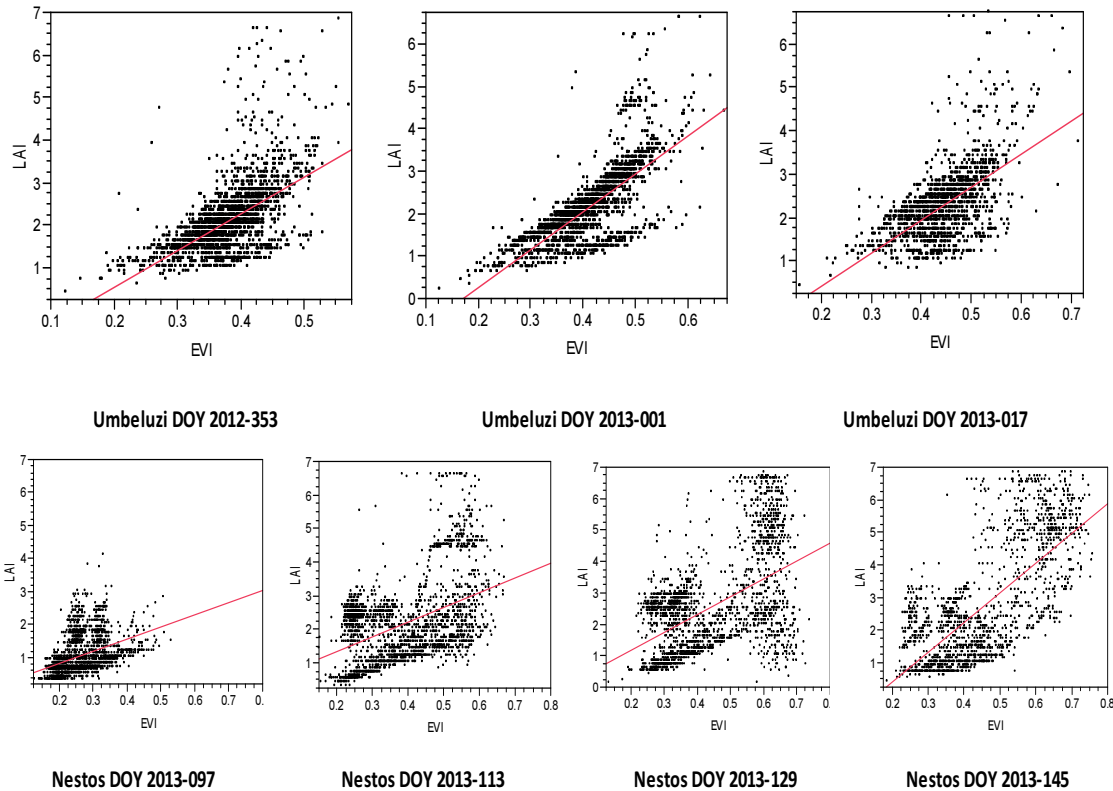
Figure 23: *Plots of EVI(x) – LAI(y) data for 3 consecutive DOY for the Umbeluzi study area (1st row) and for 4 consecutive DOY for the Nestos study area (2nd row).*

5.2.3 Relation per vegetation type

When the regression analysis was applied per vegetation type there was a strong increase in R^2 values compared to the whole year results (chapter 5.2.1, Table 8), as demonstrated in Figure 25. The coefficient of determination (R^2) was high and statistically significant (ranging from 0.41 to 0.83 in approximately 90% of the examined cases). There was no evident pattern characterizing certain vegetation types by higher or lower R^2 values when the study areas were examined as a total, however, there was a tendency in each study area -for different vegetation types- to have significantly lower R^2 values compared to the rest of the existing vegetation types. For example, in areas with needle-leaved vegetation the EVI related to LAI with lower R^2 values for the study area of Nestos, but not for Rijnland and Tamega. The same situation was observed for shrublands of Nestos ($R^2=0.44$) and Rijnland ($R^2=0.23$), but in Umbeluzi for example, the value of R^2 was close to 0.8 for this vegetation type. Irrigated crops overall showed stability in the EVI-LAI relation, with high R^2 values in all study areas where such data was available.

There were study areas where the majority of the sampled pixels per vegetation type were concentrated in certain areas of the plot showing similar values for EVI and LAI (e.g. Nestos study area - all vegetation types except the broadleaved

forests, Figure 24). On the contrary, there were study areas where the EVI and LAI pairs of values of all vegetation types were distributed in random patterns (e.g. Rijnland study area, Figure 24) all over the plot, without being easily identified as groups of values concentrated per vegetation type in certain parts of the plot.

LULC	R²
Linear Fit LULC==21	0.673
Linear Fit LULC==22	0.704
Linear Fit LULC==31	0.564
Linear Fit LULC==32	0.17
Linear Fit LULC==34	0.444
Linear Fit LULC==35	0.706

Vegetation Type
21 = Non-irrigated
22 = Irrigated
31 = Broadleaved
32 = Needleleaved
34 = Shrubs
35 = Grass

LULC	R²
Linear Fit LULC==21	0.483
Linear Fit LULC==31	0.453
Linear Fit LULC==32	0.488
Linear Fit LULC==34	0.232
Linear Fit LULC==35	0.117

Figure 24: *Examples of linear type of regression per vegetation type available for all good LAI(y) – EVI(x) data and R² results per case for the study areas of Nestos (up) and Rijnland (down).*

An overview of the associated regression analysis results for the coefficient of determination estimated per study area and vegetation type can be found in Figure 25.
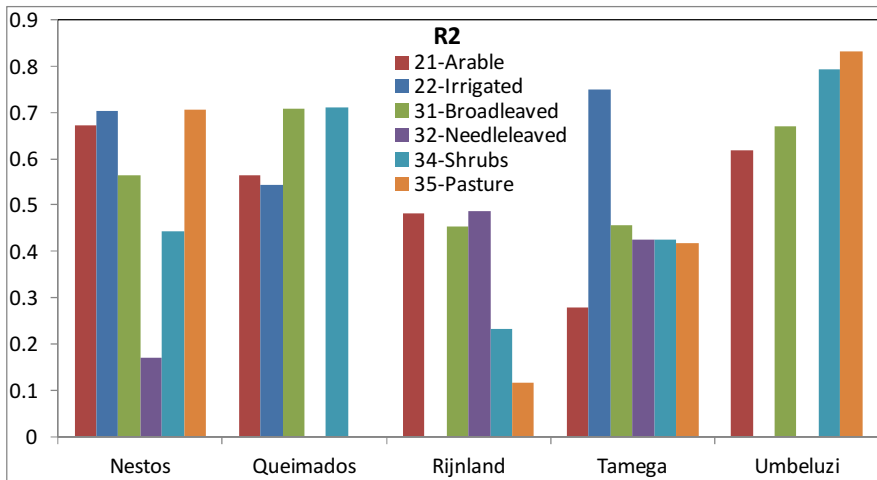
Figure 25: *R²(y) per vegetation type and study area.*
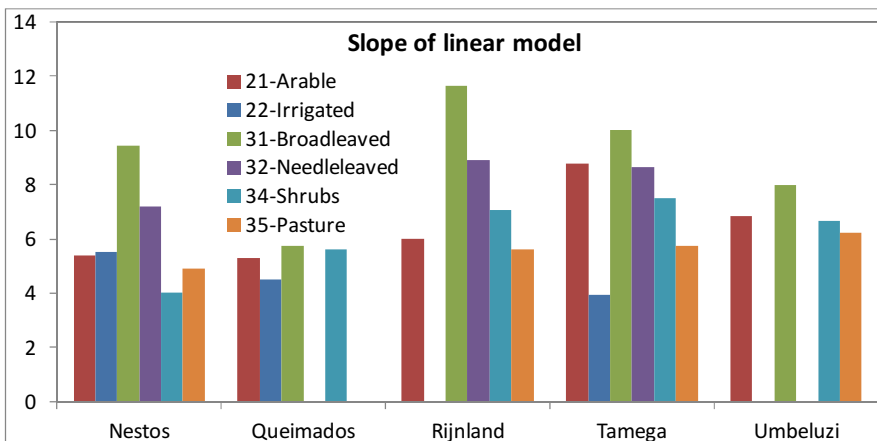
Figure 26: *Slope(y) describing the linear model per vegetation type and study area.*

Judging from the slope of the linear model, as demonstrated in Figure 26, Tamega and Rijnland had the highest slope values in general for the majority of the examined cases, with Nestos following and Umbeluzi and Queimados being characterized by the lowest values.

5.2.4 Relation per vegetation type and DOY

Regression analysis was performed once again, with the EVI and LAI datasets divided per vegetation type and DOY simultaneously (good quality pixels only). Approximately 700 equations were calculated for all data available in all study areas. Once again, the results tended to be more inaccurate and with a lower confidence level during periods of extreme meteorological conditions (rain/snow fall, cloud presence) as seen in previous cases. During such periods, there were vegetation types that were totally excluded by the QC process, or remained with a limited amount of sample points, as it can be seen in Table 9 for the vegetation classes of

shrubs (LULC 34, 7 pixels remained) and grass (LULC 35, 4 pixels remained) in the Nestos study area.

An example of the regression analysis results estimated between LAI and EVI according to this case can be found in Table 9, where the values of R^2, amount of sample points (N) and probability value (p) are presented for 2 dates and for 2 study areas (Nestos, Umbeluzi). In this example, it was observed that for the study area of Nestos on DOY 2013-097, the a factor in the equation: y = a + b * x, was always negative in cases where a high value of R^2 was found, when the two cases of low R^2 values were characterized by a positive (a) factor. Moreover, the p value was very high in two cases (shrubs and grass), due to the limited amount of pixels representing these vegetation types. The remaining vegetation types were represented by a large amount of pixels and had p values close to 0. The results for Umbeluzi on DOY 2013-177 (Table 9) were consistent for all the related regression analysis results.

Table 9: *Examples of regression analysis results for Nestos and Umbeluzi study areas per vegetation type and DOY.*

Study Area	DOY	LULC	R^2	N	p
Nestos	2013_097	21	0.49147	489	<.0001
Nestos	2013_097	22	0.657395	139	<.0001
Nestos	2013_097	31	0.083514	773	<.0001
Nestos	2013_097	32	0.014546	553	0.0045
Nestos	2013_097	34	0.569685	4	0.2452
Nestos	2013_097	35	0.411727	7	0.1203
Umbeluzi	**2013_177**	21	0.507396	448	<.0001
Umbeluzi	2013_177	31	0.4284	236	<.0001
Umbeluzi	2013_177	34	0.400739	171	<.0001
Umbeluzi	2013_177	35	0.598282	181	<.0001

Random patterns of regression analysis results were present in all DOY and vegetation types and for all study areas. In all study areas, there was always a tendency for certain vegetation types to be related with relatively lower R^2 values compared to the rest of the vegetation types present, during most of the DOY examined through the year. For example, in the plots per vegetation type of the Nestos study area for DOY 2013-097 (Figure 27), the non-irrigated and irrigated crops seem to be well fit and well distributed around the regression line, while the grass and the shrubs types also show linearity but with a small amount of pixels participating in the regression. Broadleaved forests show the expected LAI and EVI values for that season, but clearly are divided into two groups of pixels on the plot, indicating two different main species of plants present in the sampled pixels. In the needle-leaved forests the R^2 value found was <0.015, which is evaluated as very low.

The LAI and EVI values do not seem to be representing the state of vegetation expected during that time of the season (early spring) for the needle-leaved forests. The amount of sampled pixels participating in the regression is rather large (553) so, it is likely that this is a statistically significant result (at 0.0045 level) and the low value of R^2 are not associated to limited amount of samples (such as in cases of grasslands and shrublands). The low R^2 value of needle-leaved forests could be explained from the slope of the line as demonstrated in Figure 27 which is also low, and practically means that the changes of EVI do not justify the very low changes of LAI values (the majority of the sample points had a tendency for LAI values <1, while EVI had a fair distribution of values), which could happen due to the presence of herbaceous vegetation at the ground level or in-between such a vegetation type (needle-leaved forests). In case of the broadleaved forests, the point samples were distributed almost evenly around two parts of the diagram, one with values of EVI ranging from 0.2 to 0.3 and values of LAI from 1 to 2.5, and the second with EVI values ranging from 0.3 to approximately 0.4 and the LAI values characterized by the same range as in the first case (LAI = 1 - 2.5).
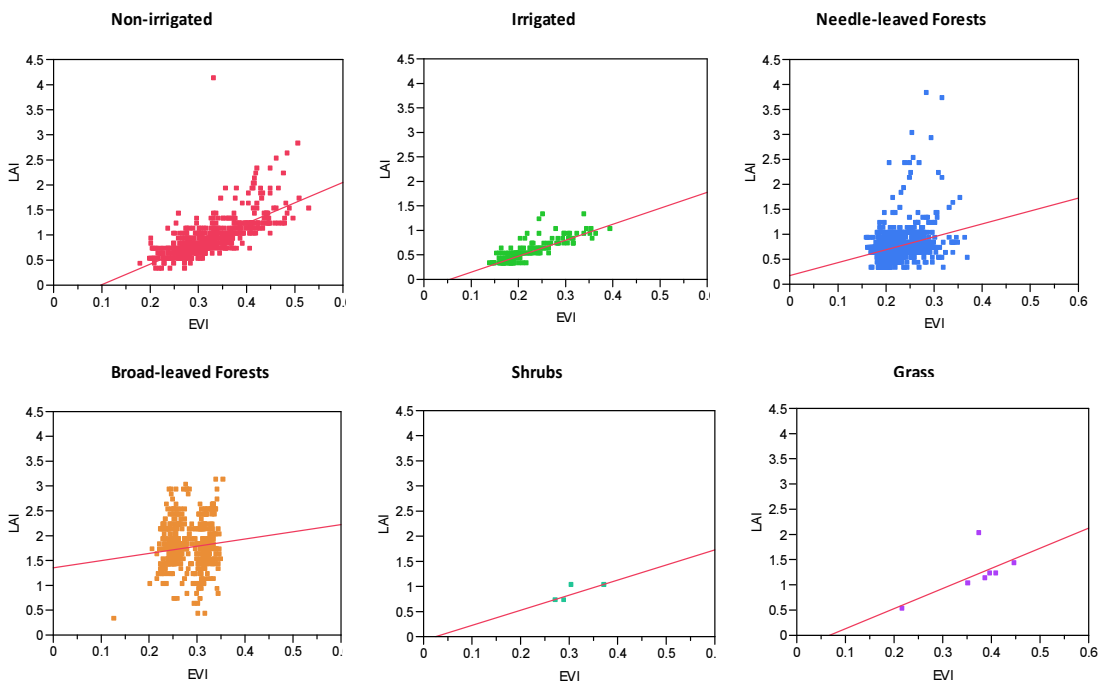
Figure 27: *Scatter plots and linear models for LAI(y) and EVI(x) on DOY 2013-097 for the Nestos study area per vegetation type available.*

As it was expected, dates for which low R^2 values had been observed using the undivided dataset per vegetation type (results per DOY chapter 5.2.2), displayed also low values of R^2 when examined by DOY and vegetation type together. By dividing the datasets into vegetation classes, it was feasible to distinguish the

vegetation types (and the associated sampled pixel data) that affected negatively the regression analysis results in each pair of LAI-EVI image available.

Because of the large amount of the estimated regression results, it was decided to give an overview of the 700 equations calculated for all study areas. Therefore, descriptive statistics are presented in Table 10 per study area and vegetation type, with R^2 values representing the average value for all DOYs of the time-series. In general, the data representing the irrigated crop type tended to have higher R^2 values, for the DOYs and study areas that such data was available, with the average of the R^2 values for the whole year estimated as a value of 0.52. However, in the Nestos area, the R^2 value for this vegetation type was approximately 0.25, which is very low compared to the other 2 study areas that such data was available. This is mainly due to the diversity of irrigated crop types in the Nestos area with different phenological cycle, including rice, maize, alfalfa, vegetables, and some fruit orchards. The other 2 areas with higher R^2 values for the irrigated crops (0.77 for Tamega and 0.54 for Queimados), have rather uniform irrigated vegetation types dominated by maize in Tamega, and soya, rice and sorghum in Queimados. The vegetation types of shrubs, grass and non-irrigated crops have a value of R^2 around 0.40, with Rijnland having the lowest value in the grass type (0.0214 - lowest value average of all cases, inundation phenomena might have affected the results). The lowest R^2 values can be found in the vegetation types of broad-leaved and needle-leaved forests, with only exception the area of Queimados who had a value of over 0.60.

Table 10: *Average of R^2 values per study area and vegetation type for the per DOY and vegetation type regression analysis, and average of R^2 values for the whole year.*

Study area	Average annual R^2 per vegetation type					
	Non-rrigated	Irrigated	Broad-leaved	Needle-leave	Shrubs	Grass
Nestos	0.4650	0.2575	0.079	0.2477	0.3768	0.6592
Queimados	0.4756	0.5452	0.6186		0.6108	
Rijnland	0.4173		0.424	0.1897	0.1867	0.0214
Tamega	0.1736	0.7747	0.1942	0.3535	0.2958	0.3214
Umbeluzi	0.4088		0.3018		0.4123	0.5875
Total	**0.3881**	**0.5258**	**0.2472**	**0.2636**	**0.3765**	**0.3974**

Queimados showed the highest R^2 values throughout the whole period of the year, with Umbeluzi following. These 2 areas are both located in the Southern Hemisphere. On the contrary, the areas in the Northern Hemisphere show moderate regression analysis results. Of these, Rijnland showed the lowest regression analysis results, with best values found from the end of February until the end of May, and gaps of data showing before and after this period. Figure 28 and Figure 29 give a good view of the mentioned situation, with the Queimados area (Figure 29) showing low results only in one case of a 16-day DOY representation (mid-January until

February) and for only 2 of the 4 vegetation types present (non-irrigated and irrigated crops).
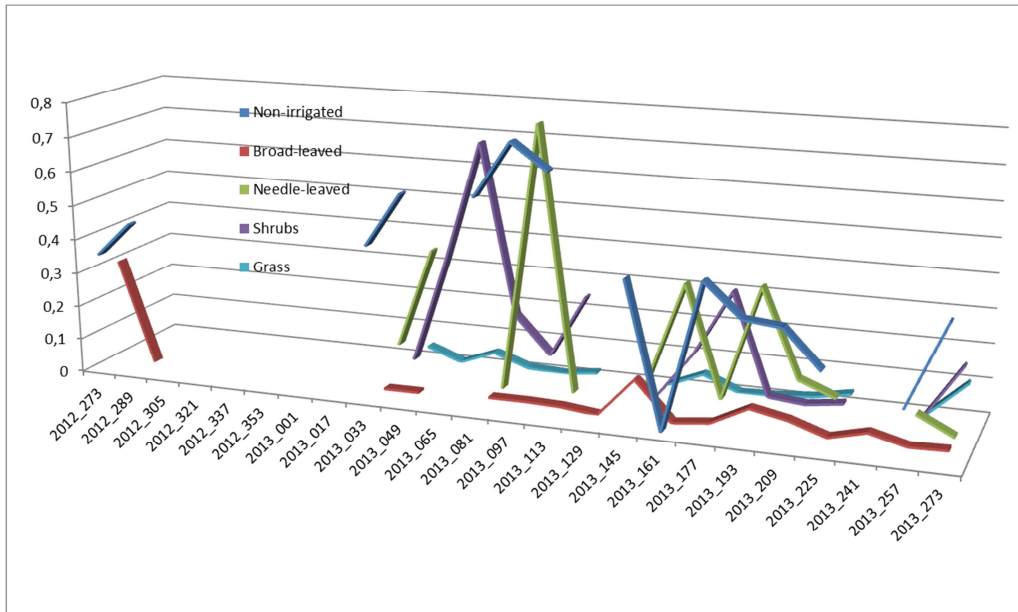
Figure 28: *R²(y) per DOY(x) available for all vegetation types in the Rijnland study area.*
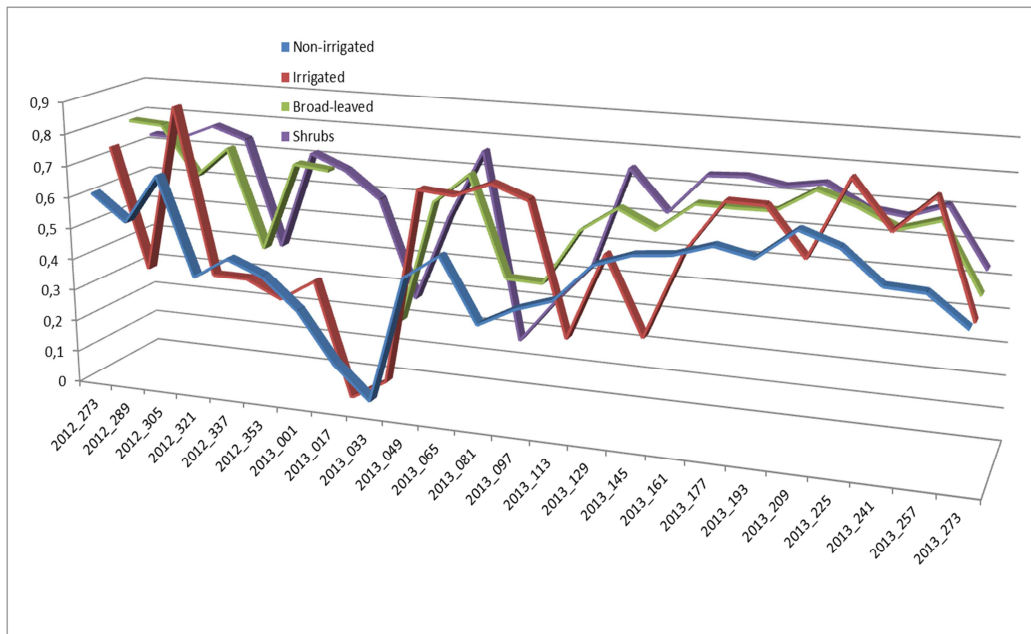
Figure 29: *R²(y) per DOY(x) available for all vegetation types in the Queimados study area.*

5.3 Relation between Landsat and MODIS reflectance

The comparison between the MODIS and Landsat Red, Blue, and Near Infrared bands for the study areas of Tamega and Umbeluzi showed high R² values between the corresponding pixel values of the datasets. The values (Figure 30, Figure 31, Figure 32 and Figure 33) were calculated to be approximately between 0.925 and 0.995, with the p probability value being highly significant for all cases and the mean difference estimated by the paired t-test applied was not exceeding an absolute value of 0.012 in all cases examined.

Obviously, the differences between the data of MODIS and Landsat were evaluated as minor in both datasets tested (500m. and 1000m. pixel resolution) as well as in both areas examined (Umbeluzi and Tamega) and thus the data was evaluated as directly comparable.

Red band. 500 meters		NIR band. 500 meters		Blue band. 500 meters	
RSquare	0.982846	RSquare	0.926595	RSquare	0.974836
Root Mean Square Error	0.002693	Root Mean Square Error	0.010547	Root Mean Square Error	0.001449
Prob > F	0	Prob > F	0	Prob > F	0
Mean Difference	0.000723	Mean Difference	-0.01125	Mean Difference	-0.00271
Observations (or Sum Wgts)	135781	Observations (or Sum Wgts)	135781	Observations (or Sum Wgts)	135781

Figure 30: *Statistical analysis and comparison of Red, NIR and Blue bands between MODIS(y) and Landsat(x) surface reflectance images at a 500 m resolution for Umbeluzi study area (MODIS DOY 2013-175).*

Red band. 1000 meters		NIR band. 1000 meters		Blue band. 1000 meters	
RSquare	0.993597	RSquare	0.957469	RSquare	0.988278
Root Mean Square Error	0.001528	Root Mean Square Error	0.007253	Root Mean Square Error	0.000925
Prob > F	0	Prob > F	0	Prob > F	0
Mean Difference	0.000722	Mean Difference	-0.01131	Mean Difference	-0.00271
Observations (or Sum Wgts)	34402	Observations (or Sum Wgts)	34402	Observations (or Sum Wgts)	34402

Figure 31: *Statistical analysis and comparison of Red, NIR and Blue bands between MODIS(y) and Landsat(x) surface reflectance images at a 1000 m resolution for Umbeluzi study area (MODIS DOY 2013-175).*

Red band. 500 meters		NIR band. 500 meters		Blue band. 500 meters	
RSquare	0.965977	RSquare	0.939475	RSquare	0.951987
Root Mean Square Error	0.004988	Root Mean Square Error	0.011824	Root Mean Square Error	0.002836
Prob > F	0	Prob > F	0	Prob > F	0
Mean Difference	0.003033	Mean Difference	0.003524	Mean Difference	-0.00025
Observations (or Sum Wgts)	122673	Observations (or Sum Wgts)	122673	Observations (or Sum Wgts)	122673

Figure 32: *Statistical analysis and comparison of Red, NIR and Blue bands between MODIS(x) and Landsat(y) surface reflectance images at a 500 m resolution for Tamega study area (MODIS DOY 2013-251).*
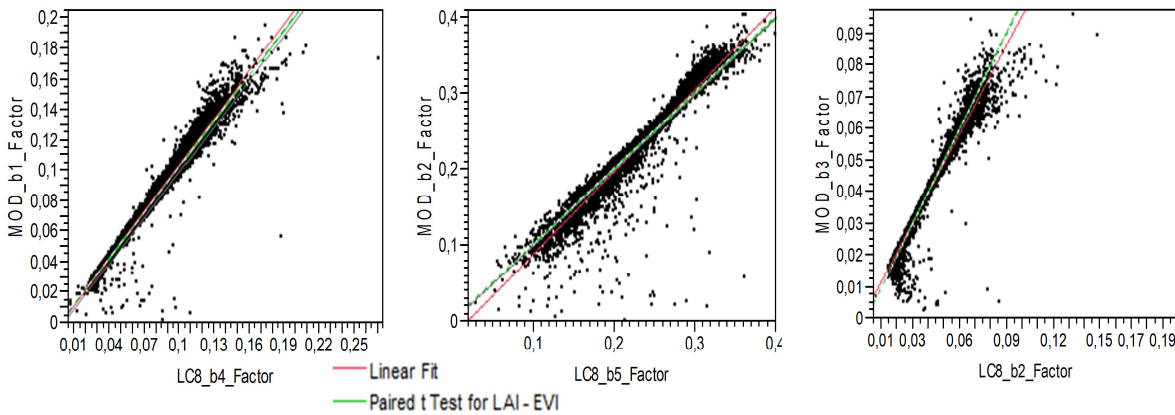
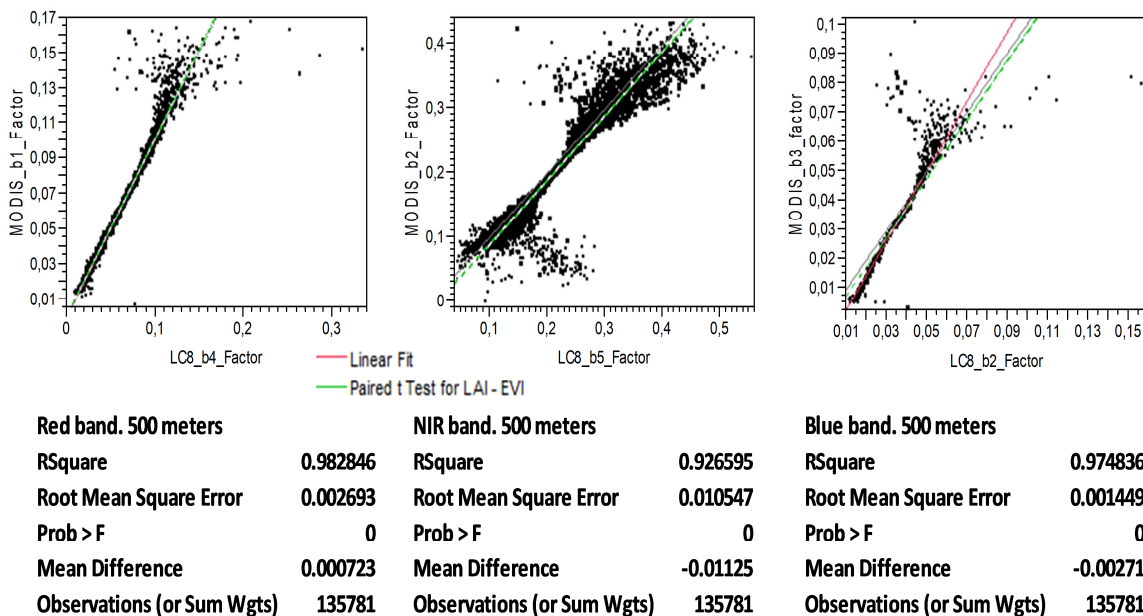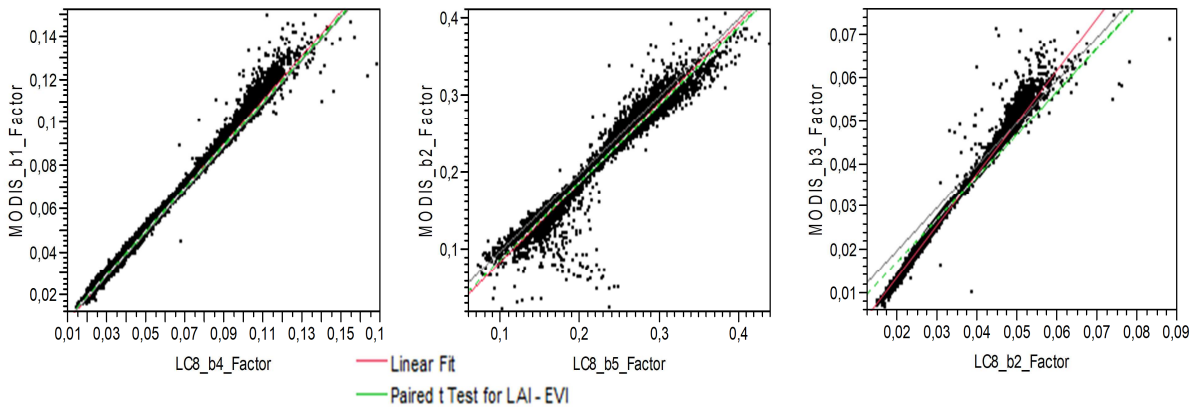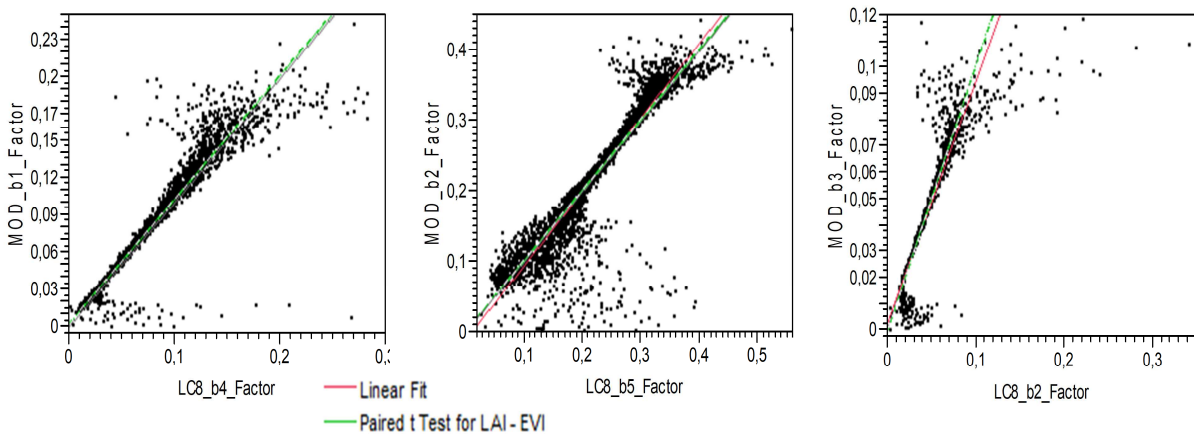| Red band. 1000 meters		NIR band. 1000 meters		Blue band. 1000 meters	
RSquare	0.982357	RSquare	0.959271	RSquare	0.972471
Root Mean Square Error	0.003316	Root Mean Square Error	0.008961	Root Mean Square Error	0.001969
Prob > F	0	Prob > F	0	Prob > F	0
Mean Difference	0.003001	Mean Difference	0.003262	Mean Difference	-0.00028
Observations (or Sum Wgts)	31644	Observations (or Sum Wgts)	31644	Observations (or Sum Wgts)	31644

Figure 33: *Statistical analysis and comparison of Red, NIR and Blue bands between MODIS(x) and Landsat(y) surface reflectance images at a 1000 m resolution for Tamega study area (MODIS DOY 2013-251).*

The additional check between the datasets of MODIS and Landsat EVI showed slightly lower R^2 values compared to surface reflectance data correlation results –but overall still high as presented in Figure 34, with a value of approximately 0.89 for the coefficient of determination in the Tamega study area and 0.87 for Umbeluzi. The mean difference was approximately 0.031 in absolute value. Once again, the p probability value was 0 for both cases. An additional finding was the fact that the surface reflectance images comparison between MODIS and Landsat showed slightly higher values of R^2 for the Umbeluzi study area when compared to these of the equivalent cases in Tamega study area, but when the EVI values of MODIS and Landsat were compared, the opposite happened; the R^2 value result estimated for the Tamega study area was slightly improved compared to this of Umbeluzi but still the difference was not significant as in the previous cases examined.
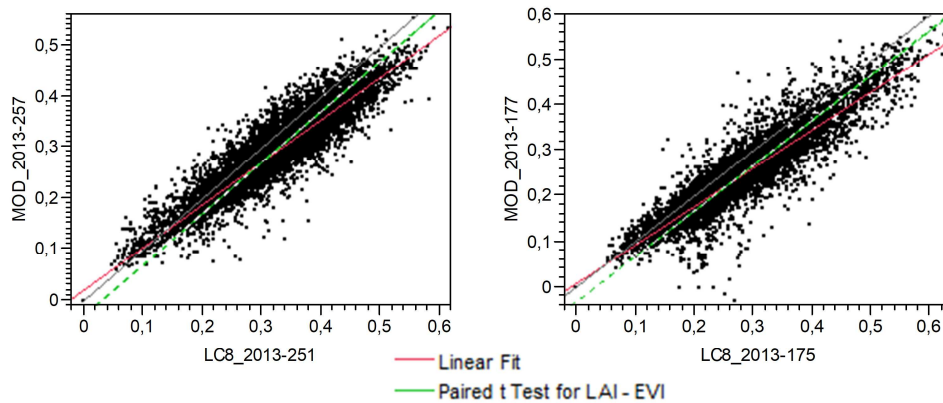
Tamega study area		Umbeluzi study area	
RSquare	0.889898	**RSquare**	0.86636
Root Mean Square Error	0.026586	**Root Mean Square Error**	0.025333
Prob > F	0	**Prob > F**	0
Mean Difference	-0.03095	**Mean Difference**	-0.03142
Observations (or Sum Wgts)	12445	**Observations (or Sum Wgts)**	16986

Figure 34: *Regression analysis of EVI values between MODIS(x) and Landsat(y) data for the study areas of Tamega (left, MODIS DOY 2013-257) and Umbeluzi (right, MODIS DOY 2013-177).*

5.4 High resolution LAI maps

The MODIS estimated regression equations per DOY and vegetation type were used in the model of downscaling MODIS LAI to Landsat resolution level for the days (DOY) close to the dates a Landsat EVI image was available. Despite the fact that this methodology is applicable for all available Landsat images during the study period (at least for those without heavy cloud presence), it was applied for all dates that LAI field measurements were available for its validation. The field measurements of the Queimados study area included only one dataset (LAI values were measured in the field only for one period during the year) and moreover, these measurements concerned a period of approximately 6 months before the starting date of the regression analysis data used for the Thesis purposes (chapter 5.3). Therefore, it was decided not to include the Queimados study area in the results of this chapter.

An example of the tables created per date and image of interest can be found in Table 11 for the Nestos and Umbeluzi study areas. For cases where for a MODIS 16-day image, two images of Landsat EVI were available (e.g. Nestos study area, Table 11) for dates close to the MODIS DOY (one image from Landsat 7 and one from Landsat 8), the regression analysis equations remained the same per vegetation type when used in the model.

Table 11: *Examples of the tables created showing the equation used in the LAI downscaling model per date and area of interest.*

Study Area	Sensor	DOY Landsat	DOY MODIS	LULC	Equation
Nestos	**LC7**	**2013_096**	**2013_097**	**21**	**LAI = -0.404429 + 4.0996533*EVI**
Nestos	LC7	2013_096	2013_097	22	LAI = -0.182653 + 3.2657184*EVI
Nestos	LC7	2013_096	2013_097	31	LAI = 0.1656716 + 2.5926736*EVI
Nestos	LC7	2013_096	2013_097	32	LAI = 1.3574629 + 1.4591396*EVI
Nestos	LC7	2013_096	2013_097	34	LAI = -0.073846 + 2.9878605*EVI
Nestos	LC7	2013_096	2013_097	35	LAI = -0.264893 + 3.9639147*EVI
Nestos	**LC8**	**2013_104**	**2013_097**	**21**	**LAI = -0.404429 + 4.0996533*EVI**
Nestos	LC8	2013_104	2013_097	22	LAI = -0.182653 + 3.2657184*EVI
Nestos	LC8	2013_104	2013_097	31	LAI = 0.1656716 + 2.5926736*EVI
Nestos	LC8	2013_104	2013_097	32	LAI = 1.3574629 + 1.4591396*EVI
Nestos	LC8	2013_104	2013_097	34	LAI = -0.073846 + 2.9878605*EVI
Nestos	LC8	2013_104	2013_097	35	LAI = -0.264893 + 3.9639147*EVI
Umbeluzi	**LC8**	**2013_175**	**2013_177**	**21**	**LAI = -0.092231 + 2.9176324*EVI**
Umbeluzi	LC8	2013_175	2013_177	31	LAI = -0.20038 + 3.6218802*EVI
Umbeluzi	LC8	2013_175	2013_177	34	LAI = -0.144638 + 3.2141674*EVI
Umbeluzi	LC8	2013_175	2013_177	35	LAI = -0.271659 + 3.9360174*EVI

Several LAI maps were created at the Landsat spatial resolution level and are presented per date of interest and per study area.

Nestos study area

For the study area of Nestos, even though there were 3 datasets of field measurements available during the year, one was unusable due to cloudy conditions. Thus, LAI maps at Landsat resolution level were created only for 2 periods (April and July) in Nestos.

For the field measurements of April (exact dates for all available field measurement datasets per study area can be found in Table 4 of chapter 3.2.4), 2 EVI Landsat images were available to be used in the model, one from Landsat 7 (DOY 2013-096, Figure 35) and one from Landsat 8 (DOY 2013-104, Figure 36). The model excluded the areas with low quality data per date of interest (cloud, water, rocks, etc.) and for the remaining pixels, LAI values were calculated. For both dates, only some of the sites where field measurements were applied were included in the created LAI maps.

For the Landsat 7 EVI image of DOY 2013-096, the model created a LAI map including approximately 65% of the study area, leaving however only 4 field measurements available for comparison with the estimated LAI values (central part of the study area). The amount of points could be up to 8 for this LAI map, but the data gaps of Landsat 7 due to the sensor problem (SLC-off) explained in chapter 3.2.2 made this impossible, since some of the points were located in-between such data

gaps. The highest LAI values (3.33691, green color on the map) can be found in areas in the central and northern parts of the map, where needle-leaved forests are located, as well as in some areas where broad-leaved forests are present, which is logical considering the season of the year (early spring). On the other hand, crops show relatively low values at this season.
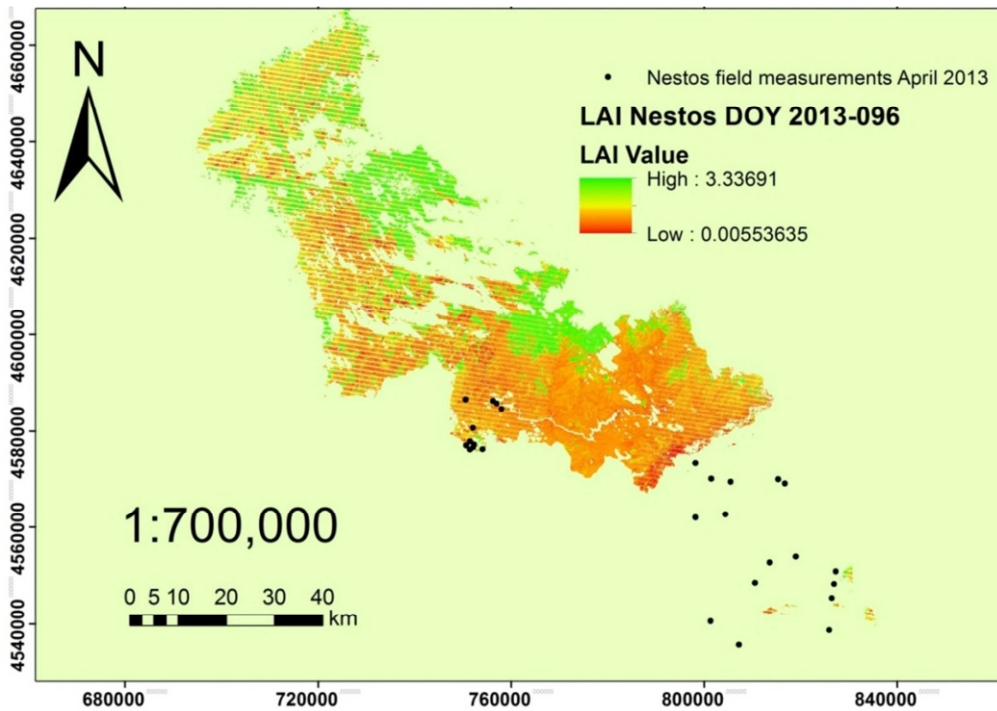
Figure 35: *LAI map at Landsat resolution (30m) for the Nestos study area (DOY 2013-096).*

For the Landsat 8 EVI image of DOY 2013-104, the same regression equation was used as with the DOY 2013-096 Landsat 7 image, in order to create the Landsat resolution LAI map. This time the excluded areas due to low quality data reached approximately 90% of the total study area. However, 5 field measurement points were available for the parts of the study area that LAI values have been estimated. The highest LAI value present is again close to the one found in the map of DOY 2013-096 (3.4857) and is characterizing mainly areas with needle-leaved forests in the northern part of the map. However, this time there are also areas with broad-leaved forests in the southern part of the map having a high LAI value.
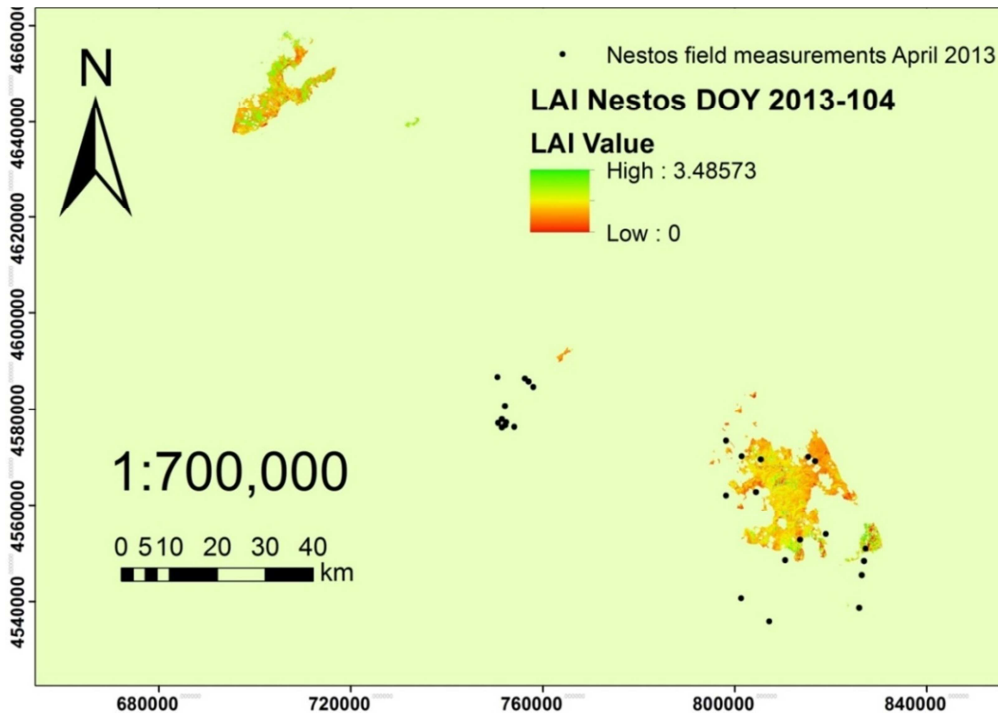
Figure 36: *LAI map at Landsat resolution (30m) for the Nestos study area (DOY 2013-104).*

An additional LAI map was created for the period of July that field measurements were available (DOY 2013-184, Figure 37). This time, almost 80% of the study area was included in the estimated LAI values of the map, since the summer season in the local climate (June - August) is generally characterized by good weather conditions and consequently almost zero cloud contamination in the acquired Landsat images. The highest LAI value found at this time of season is close to 12 (characterizing areas covered with broad-leaved and needle-leaved forests in the central and northern part of the map), but on this occasion there are also areas with agricultural activity in the southern part of the map (irrigated crops) showing LAI values close to 6 and even higher, since at this time of the year several species of crops are in the stage of full growth. On the other hand, the lowest values of LAI close to 0 can be found in areas with arable (non-irrigated) crops.
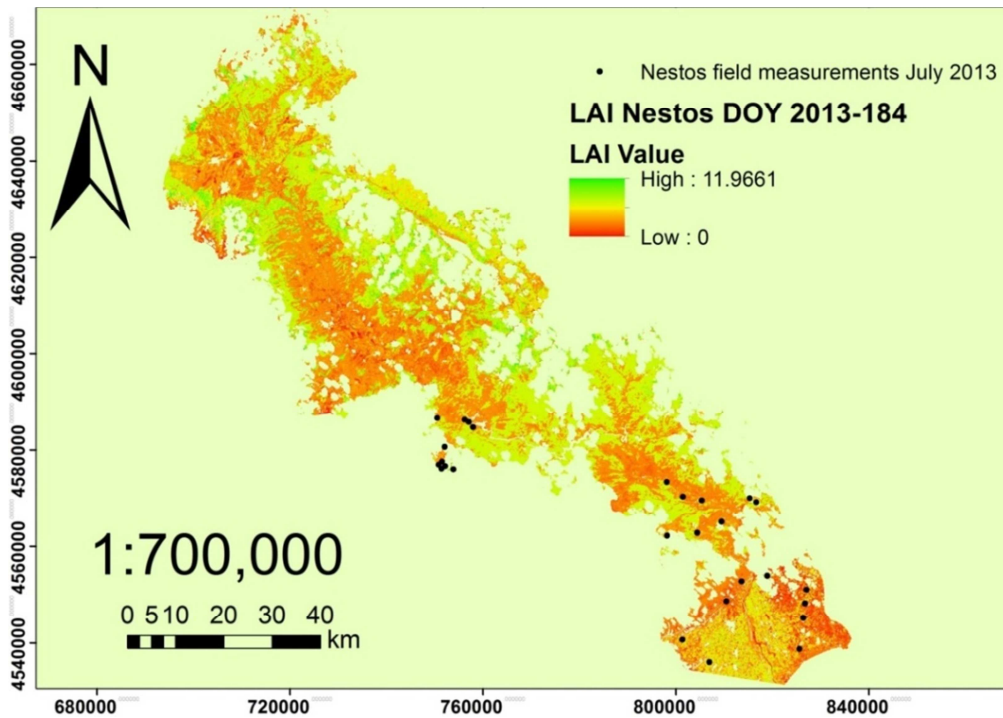
Figure 37: *LAI map at Landsat resolution (30m) for the Nestos study area (DOY 2013-184).*

Rijnland study area

For the study area of Rijnland, only two Landsat EVI images were cloud free and included parts of the study area where field measurements were available. The images of DOY 2013-265 (Landsat 7, Figure 38) and DOY 2013-273 (Landsat 8, Figure 39) were used to calculate the LAI values at Landsat resolution level, both using the same regression equations, since they considered the same 16-day period MODIS equations were available.

The LAI map created from the Landsat 7 image included only 25% of the study area due to the effect of low quality data and the data gaps present. Only 7 points of field measurements coincided with pixels of an estimated LAI value. The highest LAI value found was over 8 and mainly represented forested areas. High LAI values could also be found in areas with shrubs, when the lowest LAI values characterized narrowed areas all over the remaining study area, falling into the non-irrigated crops vegetation type.
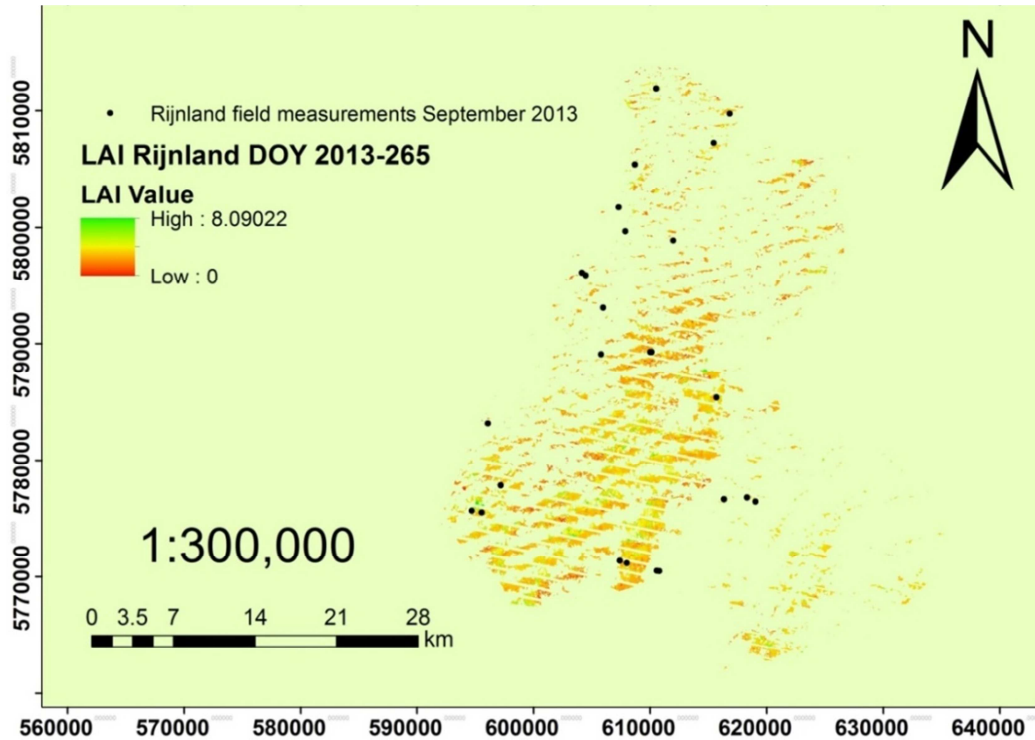
Figure 38: *LAI map at Landsat resolution (30m) for the Rijnland study area (DOY 2013-265).*

The EVI map of DOY 2013-273 was almost cloud-free and resulted into a created LAI map representing approximately 90% of the Rijnland study area. In this case, even though the highest LAI value was a bit lower compared to DOY 2013-265 (7.42), the map is dominated with a rate over 60% of the total area with LAI values higher than 5. Moreover, these high LAI values represent pixels from all available vegetation types in the area, rather than falling only into the broad-leaved and needle-leaved vegetation categories. The early days of September in the region of Netherlands can be characterized by relatively high LAI values for the vegetation, since the local climate (temperature is still high, heavy rainfalls and strong winds occur rarely compared to the winter season) is ideal for the growth of the plants.
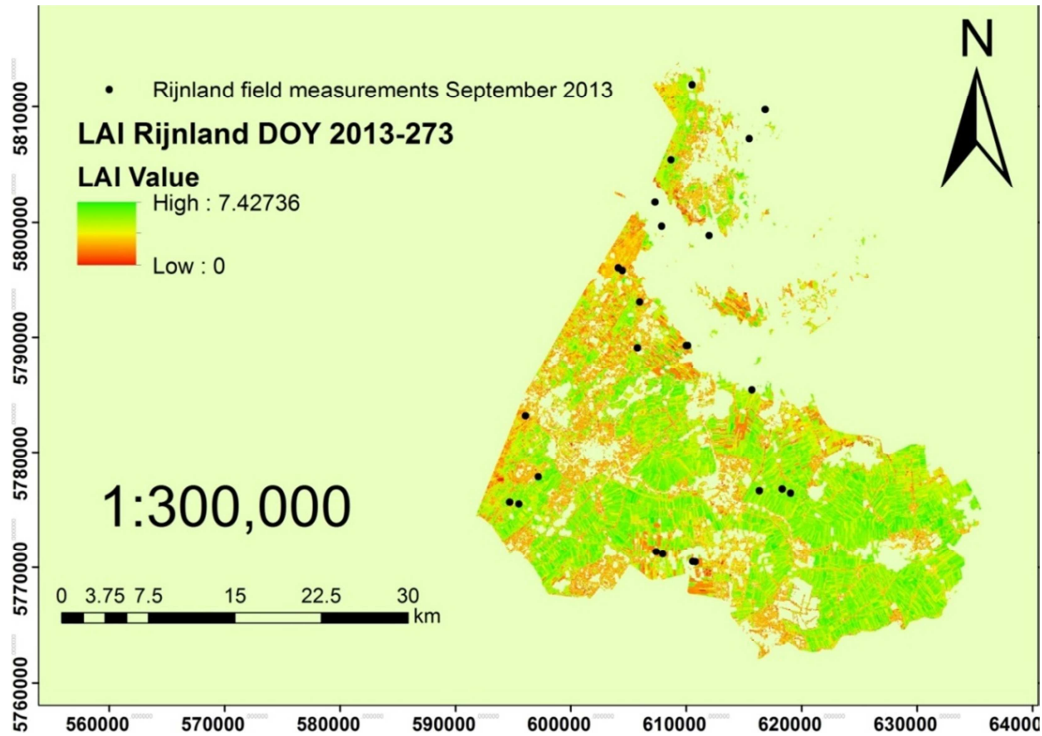
Figure 39: *LAI map at Landsat resolution (30m) for the Rijnland study area (DOY 2013-273).*

Tamega study area

In the Tamega study area, 2 images of EVI Landsat were available for the dataset of field measurements of May and 2 additional for the field measurements of September (one for Landsat 7 and one for Landsat 8 per each period).

For the fieldwork dataset of May, a Landsat 8 EVI image was available for DOY 2013-139, which was dominated by low quality pixels that were discarded, with the remaining 5% of the pixels for which LAI values were calculated not coinciding with a point of a field measurement (Figure 40) and thus, could not be used for validation purposes.

The LAI map of DOY 2013-147 (Figure 41) was created using a Landsat 7 EVI image in the downscaling model. This time, almost 85% of the total study area was assigned a LAI value, with the highest LAI value being close to the LAI map of DOY 2013-139 (7.812). The highest LAI values are characterizing mainly needle-leaved forest areas (central and southern part of the map), followed by relatively high values for the broad-leaved forests (central and northern part of the map) and the irrigated crops (central-East part of the map). On the contrary, lowest LAI values can be found in the majority of the non-irrigated crop areas (South-East and northern part of the map) and for the grasses vegetation type (northern part of the map), while shrublands are characterized by moderate LAI values and are indicated on the map with a light orange or yellow color.

Figure 40: *LAI map at Landsat resolution (30m) for the Tamega study area (DOY 2013-139).*
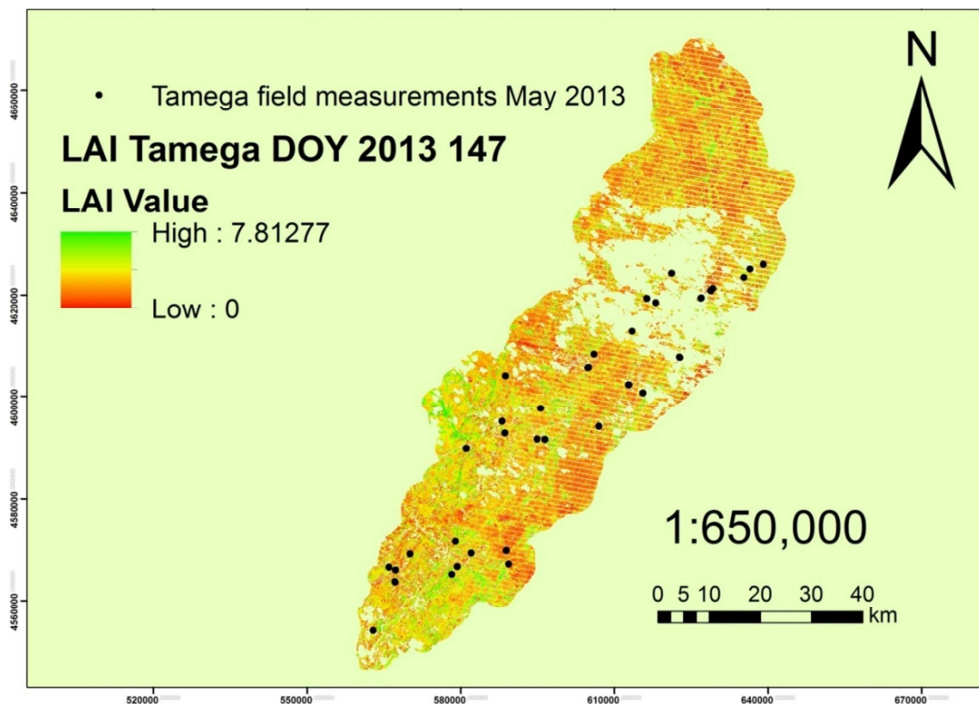
Figure 41: *LAI map at Landsat resolution (30m) for the Tamega study area (DOY 2013-147).*

The final 2 LAI maps created for the Tamega study area were for DOY 2013-243 (Landsat 7, Figure 42) and for DOY 2013-251 (Landsat 8, Figure 43). The highest

LAI values in both cases remained at similar levels as in the LAI maps of May (approximately 7.8) and the distribution of the LAI values followed a similar pattern in both periods. However, the values for the needle-leaved forests were relatively lower in September, compared to the maps of May. Moreover, the shrublands present in the central and southern part of the map had lower LAI values in September. The comparison between the 2 maps of September showed that the map of DOY 2013-243 is characterized by slightly lower LAI values for the total of the study area, compared to the one of DOY 2013-251, even though the pattern of LAI distribution remained the same in both maps. This could be explained by the fact that in both model runs for the creation of the maps, the same MODIS equations per vegetation type were used (DOY 2013-241), since the regression equations of MODIS DOY 2013-257 (closer to DOY 2013-251) were not available due to the low quality MODIS image available for that day. Therefore, the values of the LAI map DOY 2013-243 were evaluated as more accurate and closer to the real conditions present, despite the fact that both cases were assigned LAI values for approximately 95% of the total study area (cloud free images), since the equations used for the calculation of LAI values were closer to that date (2 days) compared to DOY 2013-251 (10 days). For both maps, the majority of available field measurements coincide with areas for which a LAI value was calculated.
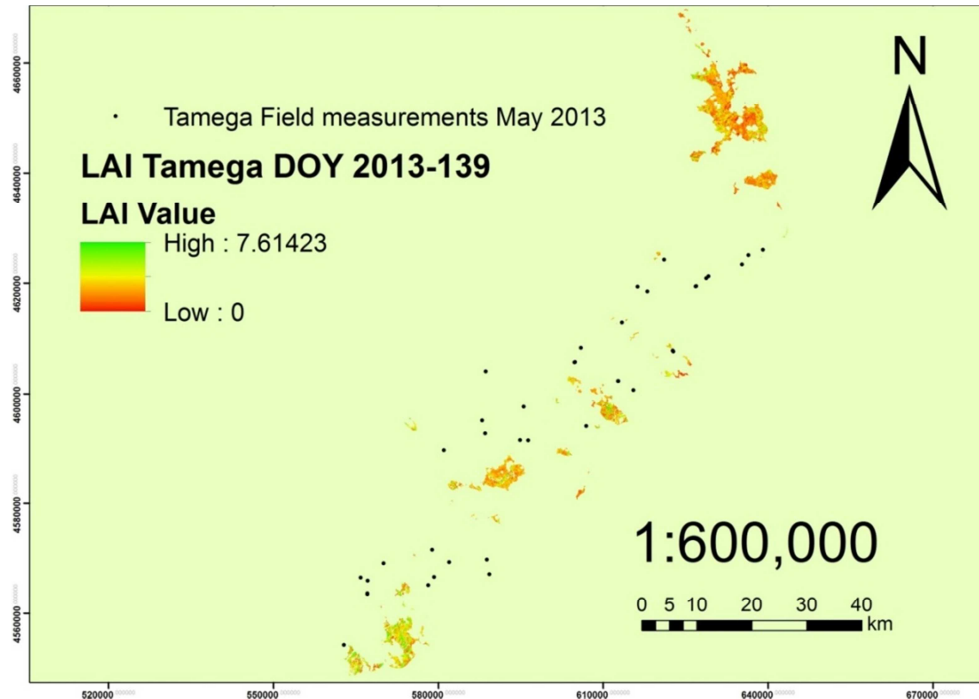

Figure 42: *LAI map at Landsat resolution (30m) for the Tamega study area (DOY 2013-243).*
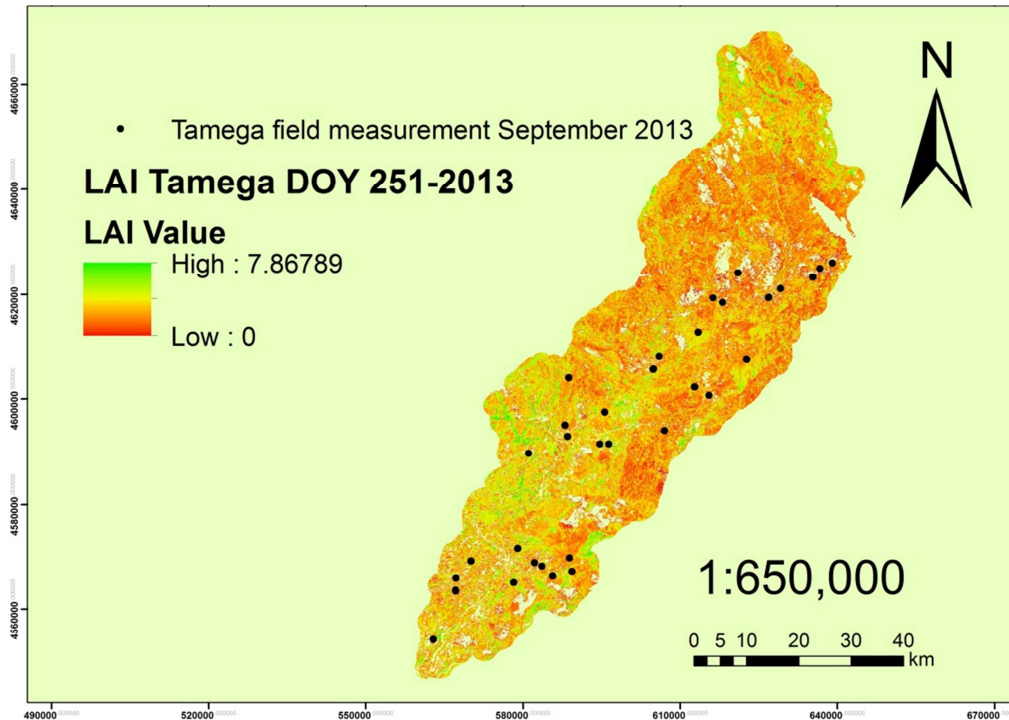
Figure 43: *LAI map at Landsat resolution (30m) for the Tamega study area (DOY 2013-251).*

Umbeluzi study area

Umbeluzi was the only study area for which Landsat EVI images were available for all existing datasets of field measurements. A LAI map was created for the November 2012 period (DOY 2012-309, Figure 44) using Landsat 7 EVI data, covering approximately 25% of the total study area. November is a period when rainfalls are highly possible to occur, which affects the local vegetation and therefore LAI values are the highest found during the whole year (highest LAI value 7.35). The low values in the south-west part of the map, as well as in the north-east, are characterizing non-irrigated crops which probably had no vegetation at that time of the year. Forest areas and shrubs showed the highest LAI values, while grasslands were characterized by relatively moderate LAI values. Several field measurement points from the available dataset of that period coincide with the areas for which LAI values were available.
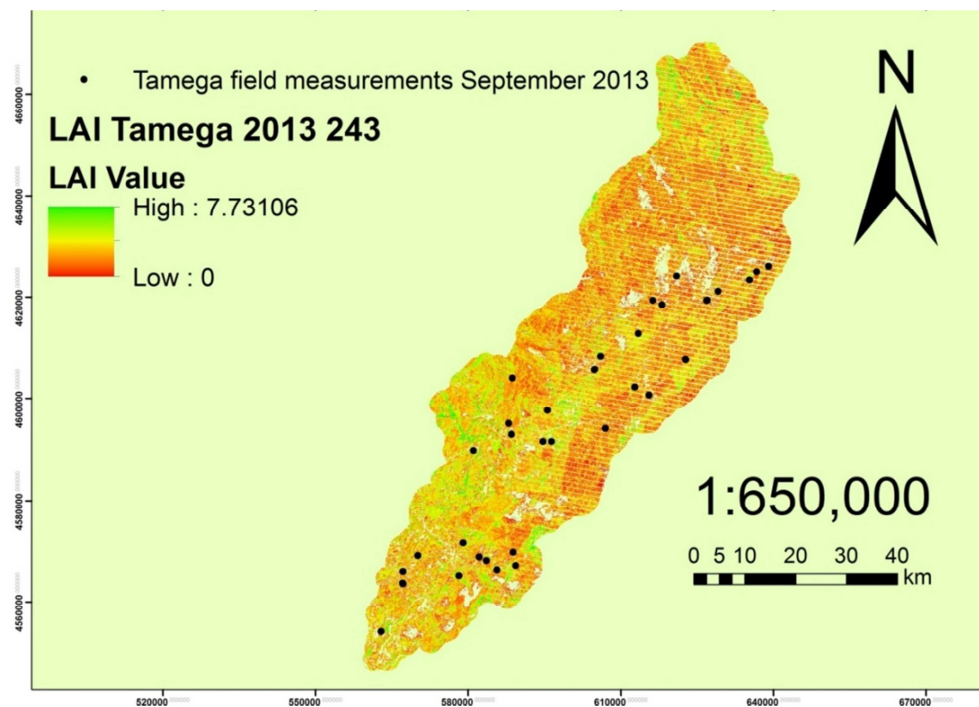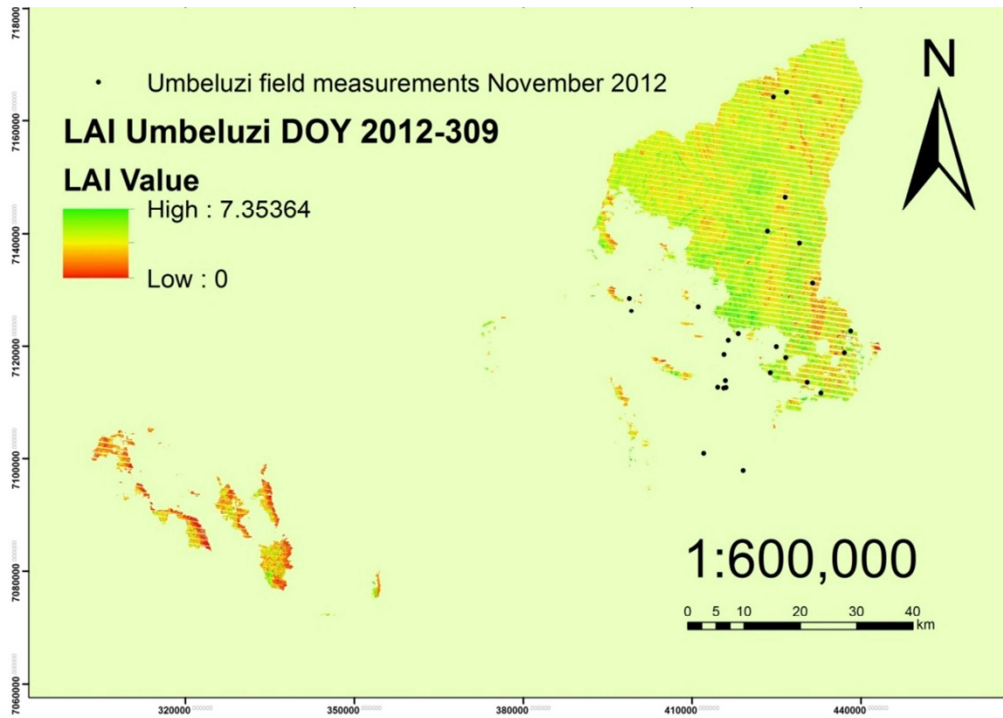
Figure 44: *LAI map at Landsat resolution (30m) for the Umbeluzi study area (DOY 2012-309).*
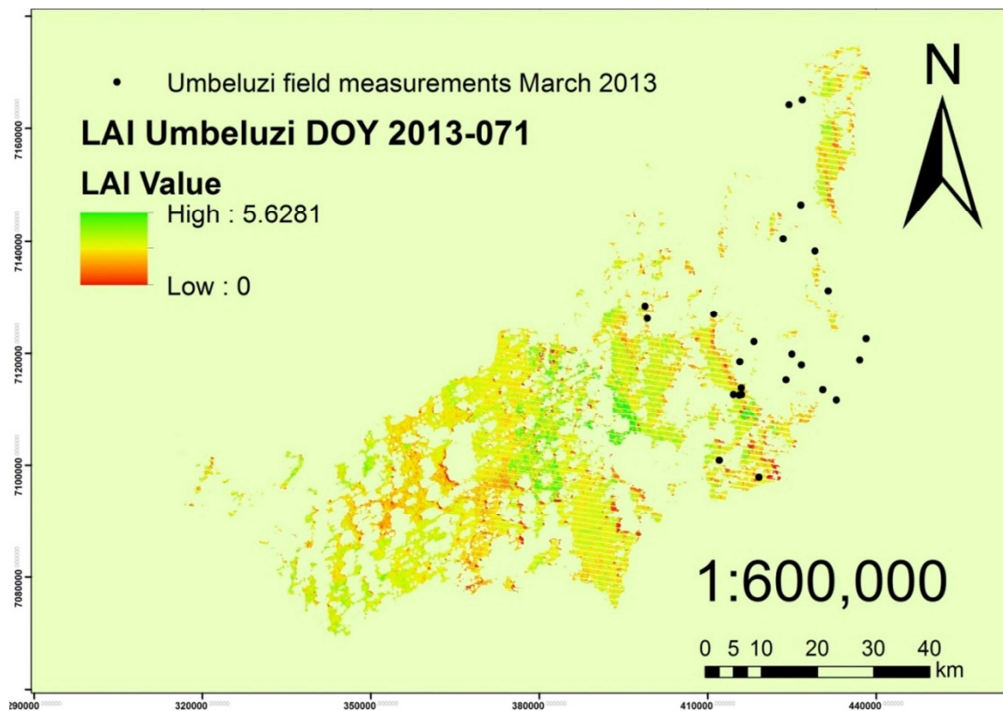
Figure 45: *LAI map at Landsat resolution (30m) for the Umbeluzi study area (DOY 2013-071).*

For the period of July (dry period), the field measurements were acquired between DOY 182 and 184 of 2013 and, therefore, 2 EVI images of Landsat 8 were

available to be used in the downscaling model, one for DOY 2013-175 and an additional one for DOY 191. The MODIS equations used were different for each run of the model (equations of DOY 2013-177 and 2013-193) and in both cases had only 2 days difference from the EVI images.
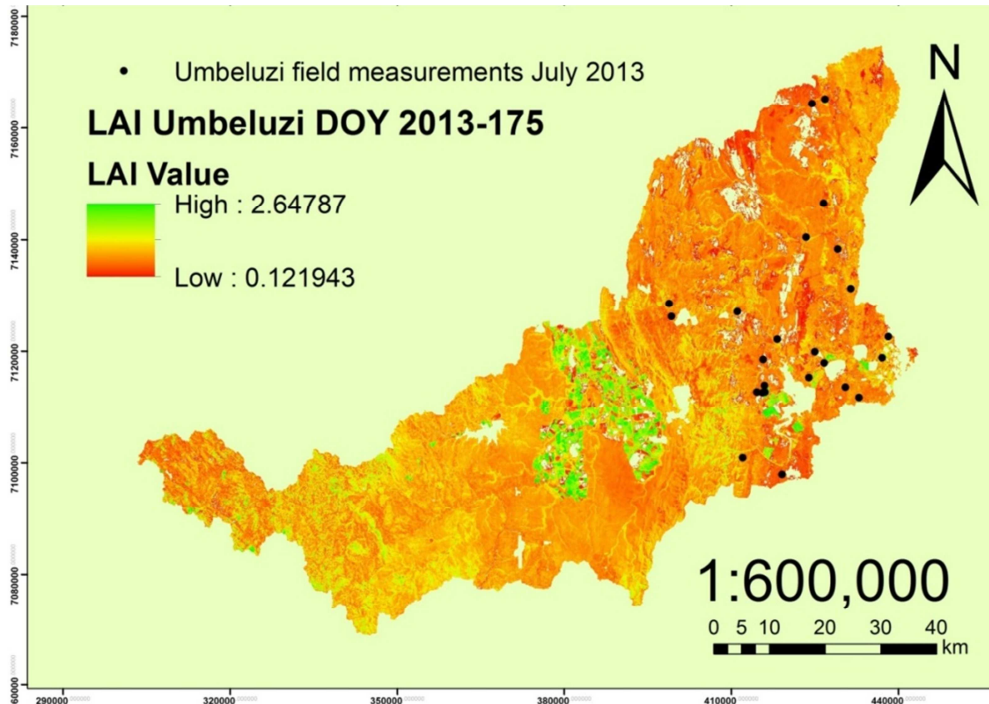
Figure 46: *LAI map at Landsat resolution (30m) for the Umbeluzi study area (DOY 2013-175).*

In case of LAI map DOY 2013-175 (Figure 46), the highest LAI value was calculated as 2.64 for forested areas in the central and central-east part of the map, for which areas the LAI values remain relatively high compared to the rest of the areas during all periods examined. Once again, non-irrigated crops showed the lowest values, but this time for the majority of the related areas. Moderate values of LAI presented with light orange and yellow color on the map represented mainly shrublands, as well as a part of the grasslands present in the study area. Almost 100% of the area was included in the calculation of LAI values, due to the fine weather conditions that provided a cloud-free EVI image. Therefore, almost all field measurements coincided with areas assigned a LAI value.

The LAI map of DOY 2013-191 included approximately 65% of the total study area, showing a much higher LAI value (approximately 4) compared to DOY 2013-175, but for the exact same areas located in the central and central east part of the map (indicated with green color in both maps). The overall pattern of the LAI value distribution was similar when the 2 LAI maps were compared. The values of DOY 2013-191 were relatively lower compared to DOY 2013-175, when the maximum LAI value present was taken into consideration. This was evident in the western part of

the 2 maps when compared, where areas colored with dark orange and orange color in map of DOY 2013-191, were represented with lighter colors (yellow, green) in map of DOY 2013-175. This phenomenon has two possible meanings; first, is that the LAI values of forested areas in the central part of the area increased rapidly (almost doubled) in a period of 16 days, which is not likely to happen. The second and most possible explanation could be that in the regression equations of DOY 2013-191, the regression equation used for the forested areas was not as accurate as the one used for the creation of LAI map DOY 2013-175, returning higher values from those expected. This is also strengthened by the fact that during a dry season, high fluctuations in LAI values, especially in natural vegetation areas, are not likely to occur in such a short period.
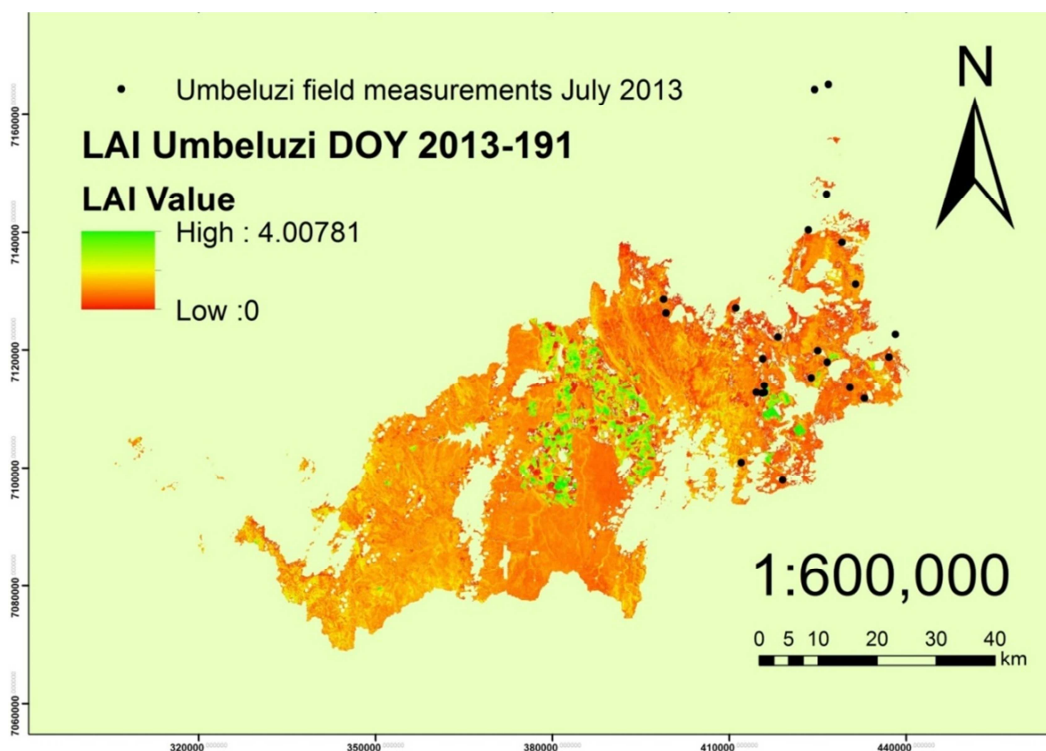
Figure 47: *LAI map at Landsat resolution (30m) for the Umbeluzi study area (DOY 2013-071).*

Details over the level of improvement in the downscaled LAI maps of Landsat resolution level, compared to those of MODIS at 1000m pixel resolution, can be found in Figure 48. The LAI map of DOY 2013-175 of the Umbeluzi study area was the image characterized by the least data gaps from the created LAI-maps series between all study areas (almost 100% of the study area pixels were assigned a LAI value). Compared with the original LAI image of MODIS, there are several features of detailed LAI that are now visible and can be clearly distinguished from the adjacent features. In general, the two images seem to be characterized by a similar pattern in the distribution of the LAI values. The forested area indicated with green color in the

central part of the maps and the crops in the northern part of the maps (indicated with red color) are characterized by the same high (green color) and low (red color) levels of LAI values. On the other hand, differences can be observed mainly in the South-West area of the images, with the low (red color) LAI values of the MODIS image being replaced by moderate (orange and yellow color) LAI values. The "pixelize" effect of the MODIS image due to the large pixel size of MODIS LAI is absent in the Landsat LAI image and transition from one pixel value to another is achieved by a smoother visualization.
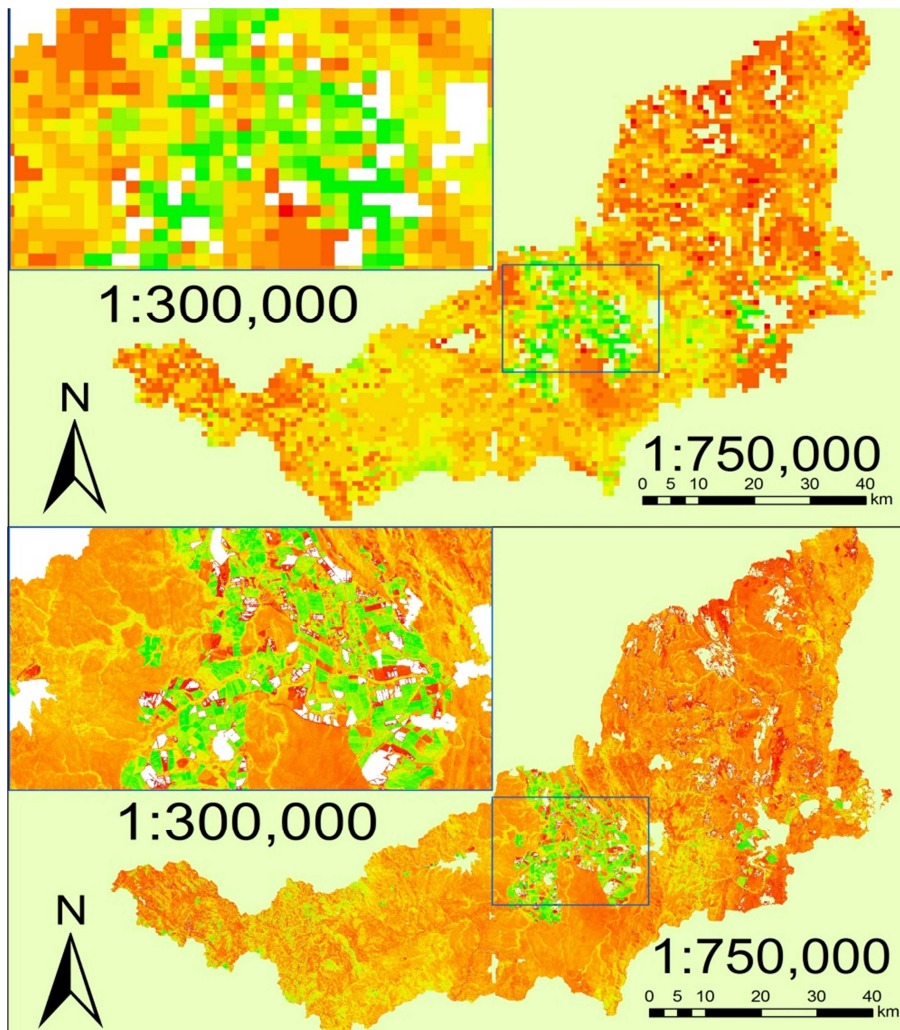
Figure 48: *Comparison of spatial resolution level between a MODIS LAI image (upper part of figure, Umbeluzi, DOY 2013-177) and a downscaled LAI image at Landsat spatial resolution level (lower part of figure, Umbeluzi, DOY 2013 175).*

5.5 Accuracy of Landsat LAI

The results of the correlation analysis are presented in this chapter per study area. In cases of extreme variations between the LAI values measured at the field and estimated for the Landsat LAI maps, scenarios with an approximate 10% exclusion rate of outliers were used. The excluded sample points included sites indicated with a different vegetation type during the fieldwork from the one assigned by the Landsat land-cover maps, or sites in which special conditions were observed during the period the field measurements were acquired, such as heavy rainfalls, presence of mud, etc.

Nestos study area

In the Nestos study area, there were no correlation analysis results for the November dataset of field measurements (Table 12), as the related Landsat EVI images were fully covered by clouds (low quality data).

Table 12: *Correlation analysis between LAI pixel values calculated at Landsat resolution level per DOY and LAI field measurements at the same locations for Nestos study area.*

Site	Date of field survey	Landsat LAI DOY	Number of locations	r	RMSE (m2/m2)	Mean difference (m2/m2)
Nestos	3-5/11/2012	n/a	n/a	n/a	n/a	n/a
	9-10/4/2013	096-2013	4	0.7209*	0.304972	0.10509*
		104-2013	5	0.9275*	0.315112	0.604831*
	2-3/7/2013	184-2013	19	0.2542	1.484	0.0544
		184-2013^	17	0.6396*	1.0126	-0.1921
^ appr. 10% outliers exclusion			*** Statistically significant at 0.05 level**			
n/a = not available			**** Statistically significant at =< 0.001 level**			

During the period of April, the 2 LAI images of DOY 096-2013 and 104-2013 included 4 and 5 pixels respectively that coincided with points of field measurements acquired. The coefficient of correlation values (r) were high for both cases and statistically significant at 0.05 level (0.72 for DOY 096-2013 and 0.93 for DOY 104-2013), even though the amount of sampled pixels was relatively low compared to other cases examined. The RMSE was at similar levels with a value of approximately 0.3, while the mean difference was over 0.6 for the highest r value between the 2 days (DOY 104-2013) and 0.1 for the lowest (DOY 096-2013), indicating a systematic over-estimation pattern of LAI values for the map of DOY 104-2013, at least for the pixel values that could be validated.

For the period of July, the r value dropped to approximately 0.25 for the image of DOY 184-2013, which was the lowest between all datasets examined. The RMSE value was also high (1.484), while the mean difference remained at lower levels, compared to the results of April (0.054). In order to examine the effect of outliers, 2 out of the 19 available LAI field measurements for that date were excluded (LAI

measured overestimation, LAI difference>3): a broad-leaved forest area with a high slope and an area with corn fields which was indicated as grasslands in the Landsat land-cover map. After the exclusion of these samples there was a dramatic increase of the r value (0.6396) which became statistically significant at 0.05 level, as well as a decrease in the RMSE value (1.0126). There was also a change in the mean difference (0.25 in absolute value), which turned the overestimation in values for the field-measured LAI, into underestimated values (negative), which was reasonable since both the excluded samples showed a high over-estimation of LAI values.

Period	April	Period	July
R²	0.575807	R²	0.064609
RMSE	0.457968	RMSE	1.48399
Observations	9	Observations	19
Mean Difference	0.3827	Mean Difference	0.0544
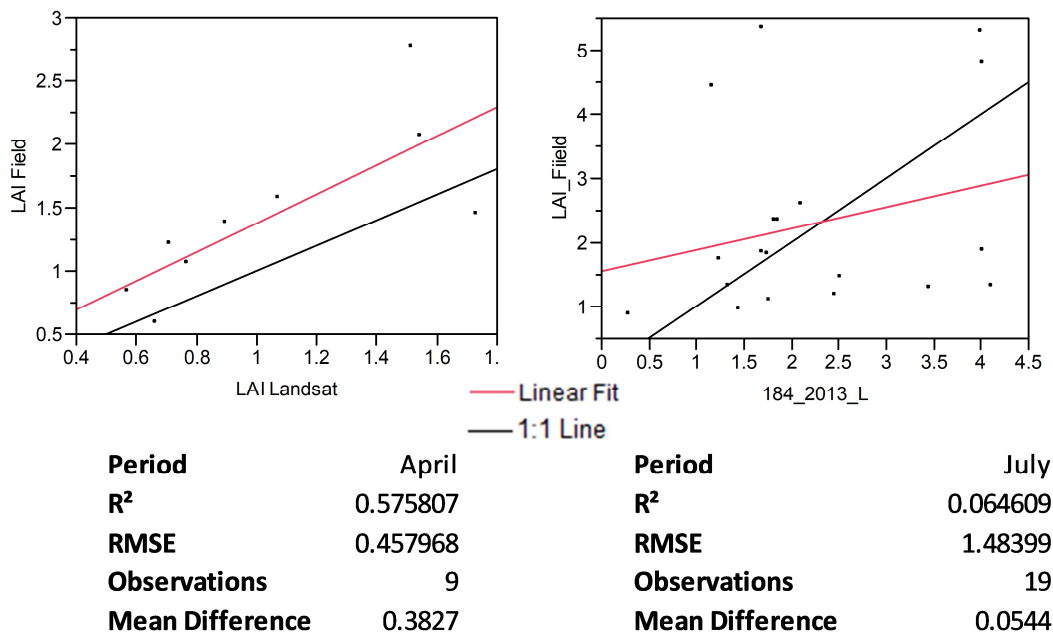
Figure 49: *Scatter plots (y by x) and statistical analysis between LAI field values(y) and Landsat LAI(x) values per period of field data availability for the study area of Nestos.*

The strong correlation found between the limited samples of field measurements and Landsat LAI during April was also evident from the scatter plots of the y by x fit (Figure 49), when on the other hand, for the period of July (DOY 2013-184) the 1:1 line relating field LAI data with Landsat LAI shows larger deviations.

Rijnland study area

The validation between field LAI data and Landsat level LAI data was not possible for the period of June, since none of the field points coincided with a pixel of a calculated LAI value (Table 13).

Table 13: *Correlation analysis between LAI pixel values calculated at Landsat resolution level per DOY and LAI field measurements at the same locations for Rijnland study area.*

Site	Date of field survey	Landsat LAI DOY	Number of locations	r	RMSE (m2/m2)	Mean difference (m2/m2)
Rijnland	26-29/6/2013	n/a	n/a	n/a	n/a	n/a
	23-26/9/2013	265-2013	7	0.8065*	0.6511	1.7764**
		273-2013	15	0.4801*	1.1576	0.8693*
		273-2013^	13	0.7875**	1.0232	0.5945*
^ appr. 10% outliers exclusion			*** Statistically significant at 0.05 level**			
n/a = not available			**** Statistically significant at =< 0.001 level**			

For the period of September, the LAI map of DOY 265-2013 included 7 samples, for which the analysis showed a correlation coefficient of 0.8065, which was statistically significant at 0.05 level. While the RMSE remained at a moderate value level (0.6511) the mean difference was high (1.776) but at the same time statistically significant at 0.001 level. The number of samples increased to 15 for the image of DOY 273-2013 resulting into a moderate r value of 0.4801, statistically significant at 0.05 level. The RMSE increased to 1.15, but the mean difference decreased to approximately half the value of DOY 265-2013 (0.87), but this time it was statistically significant at 0.05 level. Outlier sample analysis showed that 2 of the samples had overestimated values (>2.50) for the field measured data, with the one sample characterized as shrubland in the Landsat land-cover map while it was found to be a coniferous forest on a dune during the fieldwork and the other characterized as broadleaved forest during the fieldwork while it was classified as a needle-leaved forest according to the Landsat land-cover map. For the rest of the samples, the ground truth points showed high level of agreement with the land-cover types as indicated in the Landsat LAI maps. In some cases, grasslands were falsely identified as agricultural areas, however, the associated LAI values of each pair of samples when compared were close. The analysis of the remaining samples for both the LAI maps created for the September period, showed a moderate agreement of field data with Landsat LAI data as demonstrated in the y by x scatter plot of Figure 50.
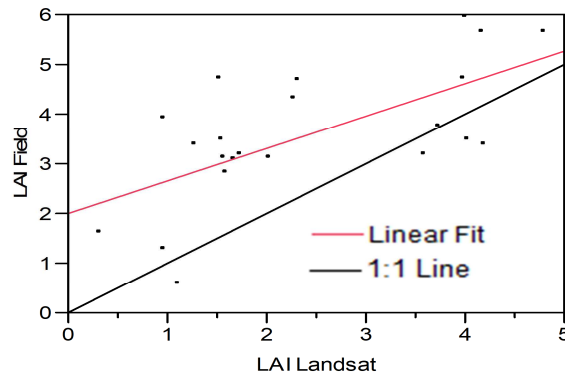
Period	September
R²	0.426626
RMSE	1.0494
Observations	22
Mean Difference	1.1579

Figure 50: *Scatter plot (y by x) and statistical analysis between LAI field values(y) and Landsat LAI(x) values (September 2013) for the study area of Rijnland.*

Tamega study area

For the period of May, only one (DOY 147-2013) of the 2 images included areas with available field measurements in the Tamega study area. 21 samples were used for the correlation analysis (Table 14), giving a moderate to low r value (0.3246) and a RMSE value of 0.88. The mean difference of LAI values (0.455) was statistically significant at 0.05 level and decreased to half when outlier samples were excluded. From the 2 samples excluded as outliers for that period, one was classified as shrubs while at field it was found to be a pine-tree forest, while the other was found to be an irrigated field while it was classified as a non-irrigated field. The difference in LAI values was lower in this case compared to the outlier exclusion process of Rijnland and Nestos and didn't exceed a value of 2. However, it was once again following an overestimation pattern of LAI field measured values. The r value increased after the exclusion of outliers to a moderate level (0.458), becoming also statistically significant at 0.05 level, while the RMSE value remained almost the same.

Table 14: *Correlation analysis between LAI pixel values calculated at Landsat resolution level per DOY and LAI field measurements at the same locations for Tamega study area.*

Site	Date of field survey	Landsat LAI DOY	Number of locations	r	RMSE (m2/m2)	Mean difference (m2/m2)
Tamega	20-23/5/2013	139-2013	n/a	n/a	n/a	n/a
		147-2013	21	0.3246	0.8809	0.4551*
		147-2013^	19	0.4581*	0.8628	0.2746*
	2-5/9/2013	243-2013	32	0.6366**	1.0825	0.1888
		243-2013^	29	0.7188**	0.7582	-0.0523
		251-2013	35	0.6428**	1.0593	0.2486
		251-2013^	32	0.7096**	0.7986	0.037
^ appr. 10% outliers exclusion			*** Statistically significant at 0.05 level**			
n/a = not available			**** Statistically significant at =< 0.001 level**			

The dataset of the September field measurements, showed the best correlation results between all areas located in the Northern Hemisphere (Nestos, Rijnland, Tamega), considering that both created LAI maps of that period (DOY 243-2013 and 251-2013) showed moderate to high r values (0.636 and 0.643) during the correlation analysis, with over 30 samples included in both cases examined and with the related r values being statistically significant at 0.001 level. The mean difference was characterized by similar relatively low values (0,188 and 0,248), while the RMSE was over 1 in both cases. Outlier analysis showed that 2 sample areas were found to be broad-leaved forests in the field instead of needle-leaved forests as indicated in the Landsat land cover map, while an additional one was falling into the shrubs vegetation type according to Landsat and was found to be a broad-leaved forest too. Exclusion of the outliers improved the r value result by approximately 0.07 in both cases, as well as the RMSE value by approximately 0.3 units. The mean difference value decreased to almost 0, showing a slight underestimation of measured LAI values for DOY 243-2013 (-0.05) and a slight overestimation for the values on DOY 251-2013 (0.03).

Period	May	Period	September
R²	0.1053	**R²**	0.4463
RMSE	0.8809	**RMSE**	1.0546
Observations	21	**Observations**	67
Mean Difference	-0.455	**Mean Difference**	0.22

Figure 51: *Scatter plots (y by x) and statistical analysis between LAI field values(x) and Landsat LAI values(y) per period of field data availability for the study area of Tamega.*

Scatter plots of y by x fit between the field measured LAI and the LAI values estimated at Landsat level (Figure 51), showed a low level of agreement for the period of May, while the level of agreement became moderate when the LAI values calculated for the 2 Landsat LAI maps created for the September period, were examined. In both plots no outliers were removed.

Umbeluzi study area

The correlation analysis between the field measurements of November and the LAI values calculated at Landsat level (Table 15), showed a negative r value (-0.178), although the amount of samples was not very limited compared to other cases examined. From the 10 samples used for the correlation analysis, 3 showed a LAI value difference > 2 and fell into the mixed vegetation category according to the field measurements (shrubs, trees and grass), while Landsat land-cover classified these areas as non-irrigated crops in two cases and shrubs in the third. It was evaluated as pointless to include an outlier analysis for this case, since after the exclusion of 1 outlier (10% of the total sample) the r value remains negative, after the exclusion of 2 outliers it doesn't exceed 0.2 and only increases over 0.5 when all 3 outliers are excluded (30% of the initial total of samples).

Table 15: *Correlation analysis between LAI pixel values calculated at Landsat resolution level per DOY and LAI field measurements at the same locations for Umbeluzi study area.*

Site	Date of field survey	Landsat LAI DOY	Number of locations	r	RMSE (m2/m2)	Mean difference (m2/m2)
Umbeluzi	7-9/11/2012	309-2012	10	-0.1783	1.3419	0.6534
	12-14/3/2013	071-2013	2	n/a	n/a	n/a
	1-3/7/2013	175-2013	21	0.935**	0.2013	-0.2754**
		175-2013^	19	0.9622**	0.1613	-0.2383**
		191-2013	16	0.874**	0.3127	-0.1924*
		191-2013^	14	0.9113**	0.2778	-0.129
^ appr. 10% outliers exclusion			*** Statistically significant at 0.05 level**			
n/a = not available			**** Statistically significant at =< 0.001 level**			

Correlation analysis between the field measured LAI values of March and those estimated at the Landsat level was not possible, since only 2 of the field measurement samples coincided with the areas for which a LAI value was calculated.

For the 21 samples taken into consideration for DOY 175-2013, the correlation coefficient value was the highest found reaching 0.935, with the mean difference result showing an underestimation of measured LAI values (-0.2754), while both the coefficient values mentioned were found to be statistically significant at 0.001 level. The RMSE was found to be 0.2013 and slightly improved to 0.1613 (along with the rest of the coefficient values) when 2 outlier samples were excluded, both found to fall into the mixed vegetation category (grass, shrubs, spare trees), while Landsat land-cover classified them as shrubs. The difference between the measured and the calculated LAI of the 2 excluded outliers never exceeded a value of 0.8, which is low compared to the other cases where outliers were excluded. For DOY 191-2013 16 samples were available showing a correlation r value of 0.874 statistically significant at 0.001 level and a mean difference value of -0.192, statistically significant at 0.05 level. The RMSE was low (0.312), although it slightly increased compared to DOY 175-2013. Outlier exclusion (same 2 outliers as excluded on DOY 191-2013) slightly improved the r value result (0.911) as well as the mean difference result (-0.129) and the RMSE value (0.278).

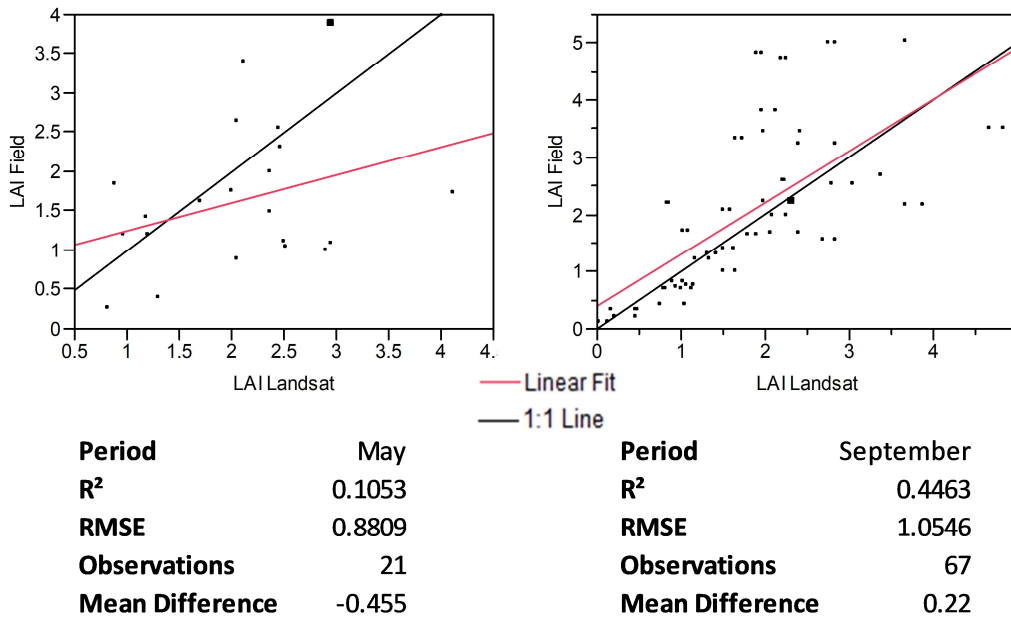|				
Period	November 2012		**Period**	July
R²	0.031808		**R²**	0.744406
RMSE	1.3419		**RMSE**	0.292474
Observations	10		**Observations**	37
Mean Difference	0.6534		**Mean Difference**	-0.2395
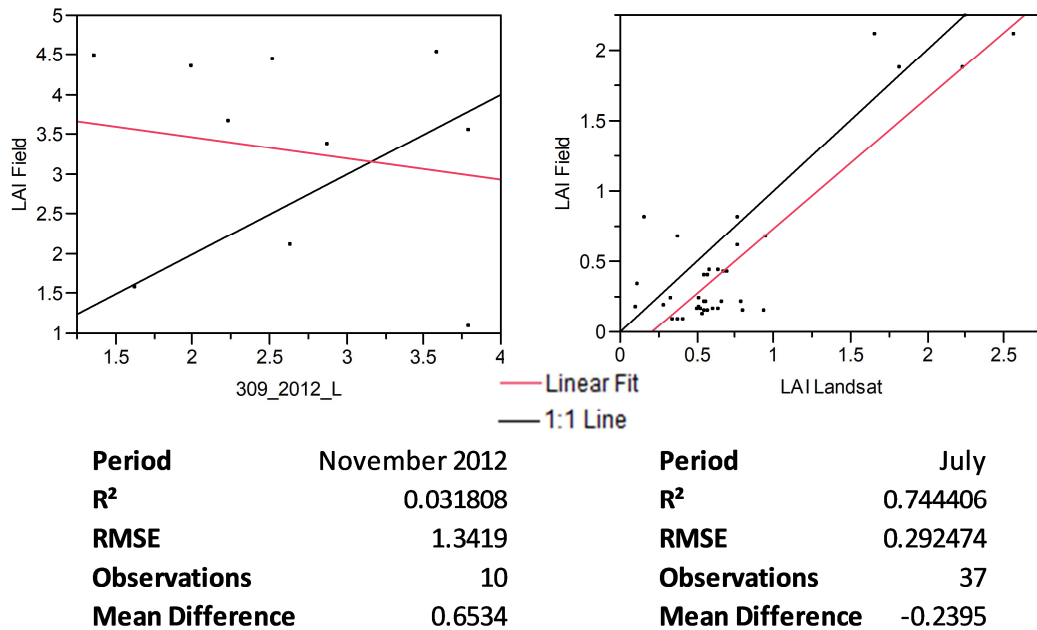
Figure 52: *Scatter plots (y by x) and statistical analysis between LAI field values(y) and Landsat LAI values(x) per period of field data availability for the study area of Umbeluzi.*

The y by x scatter plots (Figure 52) between field measured LAI and the values estimated by the model at Landsat level for the study area of Umbeluzi, showed the lowest level of agreement found between all cases examined for the period of November 2012 (1 Landsat LAI map available). For the datasets of field measurements of July, the scatter plot showed relatively low LAI values in both created LAI maps (DOY 2013-175 and 2013-191), and the highest level of agreement found between all study areas examined, for the LAI values measured and those estimated at Landsat resolution.

6 Discussion

The relationship between EVI and LAI data of MODIS 1000m pixel resolution was found to be better described by a linear fit equation in all study areas, although a polynomial fit had slightly better results of R^2 when regression analysis was applied. This choice was related to the fact that even though relevant research as presented in chapter 2 was supporting both cases, the actual fit of the equation lines when the related samples were plotted showed a stronger linear relationship for the majority of the samples and, moreover, it was believed that the complexity of a polynomial equation would affect the overall results of the LAI downscaling model. A solution to improve the results of the regression analysis as proposed by Wang et al. (2005a) would be to use more VIs in the equations. However, this would dramatically affect the amount of data needed for all study areas. Among the study sites, the strongest relationship between EVI and LAI was found for the study areas located in the Southern Hemisphere, for which the diversity of vegetation species within the same vegetation type is not as high as in the study areas of the Northern Hemisphere.

The QC process showed that during cloudy conditions the MODIS algorithm did not succeed to effectively detect and flag all the low quality data, which is evident from the sudden increase and decrease of LAI values during consecutive 16-day periods of the time-series, something also found to be true by Wang et al. (2005b). According to Ganguly et al. (2012), the LAI algorithm of MODIS cannot retrieve LAI values more precisely than its inputs, while he also warns that classification inaccuracies of land-cover data are a critical source of error in the LAI retrieval process. Such inaccuracies were present in all study areas, at least for some of the points available from the field measurement samples that were proved to be of a different vegetation type according to the Landsat land-cover maps, when compared to the real conditions found in the field. Tian et al. (2000) also added that the effect of misclassification is even larger, when the spatial resolution of the data used is coarse, such as the MODIS data used in the present study.

The majority of the R^2 values of the regression analysis per DOY were found to be lower, compared to the R^2 value estimated when all data of the time-series were considered. The analysis of the slope of the model showed that sensitivity of LAI to EVI reached a maximum level between November and February for the study areas located in the Southern Hemisphere, and between June and August for those located in the Northern Hemisphere. The highest sensitivity of EVI to LAI values was characterizing the broadleaved forests in all study areas, with needleleaved forests following in areas that such a vegetation type was present. The sensitivity of EVI to LAI was lower in periods of high biomass and the sensitivity range was different across sites, probably also related to biomass quantity or type, since the EVI has improved sensitivity at high biomass regions (Wang et al., 2005b).

Vegetation type had the highest influence in the variation of the relationship between EVI and LAI, as R^2 values were found to be high in all cases examined during the regression analysis. However, this is also the parameter that in case of biome misclassification of the data could fatally impact the quality of the associated results.

In areas with needle-leaved vegetation the EVI related to LAI with lower R^2 values for the study area of Nestos, but not for Rijnland and Tamega, something that could be caused by increased diversity of tree species in one area (increased heterogeneity) but not in the other. The highest sensitivity of EVI to LAI values was characterizing the broadleaved forests in all study areas, with needleleaved forests following in areas where such a vegetation type was present. The R^2 value of irrigated-crop areas was the highest found among the examined cases.

After examining how EVI relates to LAI by taking into account different factors each time (chapters 4.2.1, 4.2.2, 4.2.3), it was evident that the most accurate approach to apply regression analysis between EVI and LAI would be to use both the characteristics of vegetation type and seasonality (DOY). Experimenting with the data showed that these were the most crucial characteristics, which could improve the accuracy of the equations used for the LAI maps (Landsat resolution) and consequently produce results (LAI values) of as high confidence and accuracy level as possible. In general, the R^2 values dramatically improved in the regression analysis per DOY and vegetation type, compared to the results per DOY only (chapter 5.2.2), an improvement which for some vegetation types reached a value of 0.5 in absolute values, mainly for the vegetation classes of irrigated and non-irrigated crops, for the study areas where such vegetation types were present. Moreover, since the analysis of the EVI and LAI data showed variations in their relationship between even consecutive 16-day periods (MODIS EVI time-step), it was decided that it would be wise to estimate through the EVI – LAI regression analysis not only one equation per study area and vegetation type, but several more, depending on the amount of available Landsat EVI images per study area and the dates these images were associated with. This could help to evaluate whether the associated results of the regression analysis applied for consecutive dates were remaining consistent, or if they included errors, indicated by a constant increasing and decreasing of LAI values through time. From the vegetation types examined, the grasslands showed moderate to high values of R^2 during the regression analysis in all study areas, except Rijnland, probably due to inundation phenomena present because of the flat landscape of the Netherlands. Grasslands are characterized by vertical and lateral homogeneity as well as a high ground cover, and because of this, the view angle influence is considered to be minor for this vegetation type (Myneni et al., 2002). Mowing effects could also be present, however, it was not possible to verify if such phenomena occurred.

The patterns of LAI distribution in the LAI maps created at the Landsat spatial resolution were evaluated as reasonable, after visually examining the related data. The temporal variability present in the time-series of Landsat LAI maps could be explained by geo-location errors or by variable atmospheric conditions that the MODIS algorithms could not detect correctly (Tan et al., 2006). The application of correlation analysis between the LAI pixel values calculated at Landsat resolution level and the values of field measurements for the same locations and dates, displayed a wide range of values for the correlation coefficients across the different study areas and seasons. In general, the results of the correlation coefficient values

in all study areas were statistically significant (at 0.05 level for the majority of the cases and at 0.001 level for specific dates per study area), even in cases where the amount of sample points available for the analysis was lower than 10. In all study areas for which a Landsat 7 image was used to run the model, the available samples had extra limitations due to the data gaps created by the SLC-off effect of the sensor. The mean difference for most of the cases examined showed an overestimation of the field measured LAI compared to the Landsat calculated LAI, except for the Umbeluzi study area, which was the only study area with at least an available dataset of field measurements for validating Landsat LAI values, located in the Southern Hemisphere.

Overall, the study areas of the Southern Hemisphere showed higher correlation values compared to those of Northern Hemisphere during the comparison of the field measured LAI with this estimated at Landsat level. This is mainly related to the local climate of these study areas and the fact that during dry seasons, the temperature is relatively higher than for the Northern Hemisphere areas, having as a result more areas with low LAI and EVI values (dry vegetation), which are characterized by a stronger linear relationship compared to those of high LAI-EVI values. An example for this situation, which also showed the best results for the correlation analysis process among all datasets examined (in all study areas), was found between the LAI values measured in July and the values estimated in the Umbeluzi LAI maps for DOY 175-2013 and 191-2013. This could be explained by taking into consideration the local climate of Umbeluzi, in which during the dry season the tendency of LAI values is to remain at very low levels for the biggest part of the study area (<1.5), with EVI values ranging according to the vegetation type also to low levels. Because of this, overestimation and underestimation of LAI values remains to minimum levels possible, showing high linearity of the distribution of the sample points on the x-y axes.

Comparison of the correlation coefficients with a similar study, which attempted to downscale MODIS LAI to the spatial resolution of 250m (Silleos et al., 2014), showed that the results for the same study areas and datasets of field measurements were improved, with field measured LAI showing high r values (>0.7) in 5 out of 11 cases, when in their study only one case had such a high r value. Additionally, there was a slight improvement in the moderate level r values (0.4-0.7), as most of the cases examined in this study with moderate values were over the value of 0.5, while in their study all moderate values except for one were found to be lower than 0.5. A similarity found was that for the same study area and period (Umbeluzi, November 2012), a similar negative value was found (-0.10), slightly better than the one found from Silleos et al. (2014) (-0.15). Since the comparison between the 2 studies showed that the related results were improved when the spatial resolution improved, it would be interesting to examine whether the suggested methodology could find application while using Sentinel data (15m pixel) instead of Landsat, or if instead of 1000m LAI datasets of MODIS were used MODIS LAI datasets with a pixel size of 500m.

Several limitations were identified during the progress of this study. The compositing method used for the MODIS products likely has contributed to the observed differences during consecutive DOY. The LAI values of the MODIS 8-day composite product might describe vegetation conditions of a different DOY compared to the values of 16-day composite product of EVI. The MODIS algorithm chooses the day with the highest fPAR value to be the LAI value characterizing the 8-day composite period. On the other hand, the highest EVI value that is also closer to nadir is selected to be the EVI value characterizing the related 16-day composite period (VI Manual, landval.gsfc.nasa.gov/ProductStatus.php?ProductID=MOD13). Consequently, in vegetation types such as irrigated or non-irrigated crops, where agricultural activity is present, a high LAI value of a crop during the last days before harvesting might be paired with a low value of EVI for the same DOY, if the closer to nadir EVI value coincides with the days just after harvesting. The same phenomenon could also occur in dates when rapid green-up of leaves happened in a period of 10-12 days, as warned by Ahl et al. (2006). Additionally, the standard VI and LAI products of MODIS may contain significant directional noise as supported by Huete et al. (2002). Tian et al. (2000) found that a 1-km MODIS LAI pixel could be underestimated by 5% compared to field measurements, if resolution of the data is not taken into consideration.

On the other hand, beside the problems related to the MODIS algorithm, limitations also occurred due to saturation of the EVI signal in high densities of biomass (as demonstrated in Figure 20), which has been documented for other VIs too (Turner et al., 1999). As explained by Myneni et al. (2002), pixels in data of low resolution, such as the LAI product of MODIS, are likely to contain an amount of radiative contribution from the background. While saturation problems could explain the low R^2 values of the regression analysis results between EVI and LAI in the needle-leaved vegetation type, this was not the case for shrublands and grasslands since saturation is not likely to affect sparse biomes (Fensholt et al., 2004). Considering the needle-leaved forest vegetation type, the results could also be affected by the presence of lower layers of leaves in areas of dense canopies, causing insensitivity of reflectance to model vegetation (Fensholt et al., 2004).

Finally, it was found that variability of species within the same vegetation type but of a different phenology, could dramatically affect the results of the regression analysis. Heterogeneity causes greater underestimation levels of LAI and the magnitude of underestimation increases as vegetation heterogeneity increases (Fensholt et al., 2004). Therefore, the ideal situation would be to classify the data that fall into that category into more detailed vegetation classes (depending on the diversity of the dominating species per study area), in order to improve the linearity of the LAI-EVI relationship and succeed to acquire more accurate LAI values when the related regression equations are used in the LAI downscaling model.

7 Conclusions

The results of this study showed that vegetation type has the highest influence on the relationship between LAI and EVI, which was found to be best described by a linear equation, for all study areas examined. The sensitivity of EVI to LAI was lower in periods of high biomass production. In some case studies, a high percentage of variability of LAI was explained by the linear regression equation with EVI, thus the equation could be used to predict LAI with a high degree of confidence. The limited amount of pixels available for some vegetation types in specific DOYs, the diversity of vegetation species within the same vegetation types in the study areas, as well as saturated EVI values and misclassification of the vegetation type of the samples, were the most possible factors affecting the regression analysis results.

The created LAI maps showed visual similarity of high level in the patterns of LAI value distribution, when compared to the equivalent lower resolution LAI maps (MODIS).

Through the stage of the validation analysis between field-measured LAI and the corresponding values estimated by the model, the lowest correlation coefficient values were estimated during the rainy season for all the study areas, when the availability of cloud free pixels in the LAI images was low. For most of the cases examined, the model gave statistically significant results (at 0.05 and 0.001 level) with the r coefficient values ranging from negative (-0.1783, one case) and relatively low (2 cases, 0.25 and 0.32) values, to moderate (3 cases, 0.4-0.7) and high (5 cases, 0.7-0.935). This practically means that 8 out of 11 cases examined showed at least a moderate level correlation and 5 out of 11 showed high correlation level for the associated LAI values.

8 **References**

Ahl, D.E., Gower, S.T., Burrows, S.N., Shabanov, N.V., Myneni, R.B., and Knyazikhin, Y., 2006. Monitoring spring canopy phenology of a deciduous broadleaf forest using MODIS. Remote Sensing of Environment, 104(1): 88-95.

Alexandridis, T., Gitas, I., and Silleos, N., 2008. An estimation of the optimum temporal resolution for monitoring vegetation condition on a nationwide scale using MODIS/Terra data. International Journal of Remote Sensing, 29(12): 3589-3607.

Alexandridis, T.K., Cherif, I., Kalogeropoulos, C., Monachou, S., Eskridge, K., and Silleos, N., 2013a. Rapid error assessment for quantitative estimations from Landsat 7 gap-filled images. Remote sensing letters, 4(9): 920-928.

Alexandridis, T.K., Stavridou, D., Strati, S., Monachou, S., and Silleos, N., 2013b. LAI Measurement With Hemispherical Photographs At Variable Conditions For Assessment Of Remotely Sensed Estimations, ESA Special Publication, pp. 252.

Boskidis, I., Gikas, G.D., Sylaios, G., and Tsihrintzis, V.A., 2011. Water quantity and quality assessment of lower Nestos river, Greece. Journal of Environmental Science and Health, Part A, 46(10): 1050-1067.

Bréda, N.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. Journal of experimental botany, 54(392): 2403-2417.

Brogaard, S., and Ólafsdóttir, R., 1997. Ground-truths or Ground-lies?: Environmental sampling for remote sensing application exemplified by vegetation cover data. Lund electronic reports in physical geography.

Butson, C.R., and Fernandes, R.A., 2004. A consistency analysis of surface reflectance and leaf area index retrieval from overlapping clear-sky Landsat ETM+ imagery. Remote Sensing of Environment, 89(3): 369-380.

Chambel-Leitão, P., Almeida, C., Jauch, E., Trancoso, R., Neves, R., and Leitão, J.C., 2011. Evapotranspiration forecast using SWAT model and weather forecast model.

Chen, J., and Sun, L., 2010. Using MODIS EVI to detect vegetation damage caused by the 2008 ice and snow storms in south China. Journal of Geophysical Research: Biogeosciences, 115(G3).

Chen, J.M., and Cihlar, J., 1996. Retrieving leaf area index of boreal conifer forests using Landsat TM images. Remote Sensing of Environment, 55(2): 153-162.

Clevers, J., 1988a. The application of the weighted near-infrared-red vegetation index for estimating LAI at the vegetative and generative stage of cereals, Proc. 16th ISPRS-Congress, Kyoto, Japan.

Clevers, J., 1988b. The derivation of a simplified reflectance model for the estimation of leaf area index. Remote Sensing of Environment, 25(1): 53-69.

Clevers, J., 1991. Application of the WDVI in estimating LAI at the generative stage of barley. ISPRS journal of photogrammetry and remote sensing, 46(1): 37-47.

Clevers, J.G., 1986. Application of remote sensing to agricultural field trials. Landbouwuniversiteit.

Cohen, W.B., Maiersperger, T.K., Gower, S.T., and Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sensing of Environment, 84(4): 561-571.

Colombo, R., Bellingeri, D., Fasolini, D., and Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sensing of Environment, 86(1): 120-131.

Colwell, J.E., 1974. Vegetation canopy reflectance. Remote Sensing of Environment, 3(3): 175-183.

Curran, P., and Williamson, H., 1986. Sample size for ground and remotely sensed data. Remote Sensing of Environment, 20(1): 31-41.

Eklundh, L., Harrie, L., and Kuusk, A., 2001. Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. Remote Sensing of Environment, 78(3): 239-251.

Fassnacht, K.S., Gower, S.T., MacKenzie, M.D., Nordheim, E.V., and Lillesand, T.M., 1997. Estimating the leaf area index of north central Wisconsin forests using the Landsat Thematic Mapper. Remote Sensing of Environment, 61(2): 229-245.

Fensholt, R., Sandholt, I., and Rasmussen, M.S., 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. Remote Sensing of Environment, 91(3): 490-507.

Foody, G.M., Boyd, D.S., and Cutler, M.E., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sensing of Environment, 85(4): 463-474.

Ganguly, S., Nemani, R.R., Zhang, G., Hashimoto, H., Milesi, C., Michaelis, A., Wang, W., Votava, P., Samanta, A., and Melton, F., 2012. Generating global leaf area index from Landsat: Algorithm formulation and demonstration. Remote Sensing of Environment, 122: 185-202.

Gao, F., Masek, J., Schwaller, M., and Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE transactions on geoscience and remote sensing, 44(8): 2207-2218.

Goel, N.S., 1989. Inversion of canopy reflectance models for estimation of biophysical parameters from reflectance data. Theory and applications of optical remote sensing: 205-250.

Gupta, R., Prasad, T., and Vijayan, D., 2000. Relationship between LAI and NDVI for IRS LISS and Landsat TM bands. Advances in Space Research, 26(7): 1047-1050.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83(1): 195-213.

Huete, A., Justice, C., and Van Leeuwen, W., 1999. MODIS Vegetation Index (MOD13); Algorithm Theoretical Basis Document; 1999. Verion.

Huete, A., Liu, H., Batchily, K.v., and Van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment, 59(3): 440-451.

Jassas, H., Kanoua, W., and Merkel, B., 2015. Actual evapotranspiration in the Al-Khazir Gomal Basin (Northern Iraq) using the surface energy balance algorithm for land (SEBAL) and water balance. Geosciences, 5(2): 141-159.

Johnson, L.F., 2003. Temporal stability of an NDVI-LAI relationship in a Napa Valley vineyard. Australian Journal of Grape and Wine Research, 9(2): 96-101.

Kimes, D., Knyazikhin, Y., Privette, J., Abuelgasim, A., and Gao, F., 2000. Inversion methods for physically-based models. Remote Sensing Reviews, 18(2-4): 381-439.

Knyazikhin, Y., Glassy, J., Privette, J., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J., Votava, P., and Myneni, R., 1999. MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document. Theoretical Basis Document, NASA Goddard Space Flight Center, Greenbelt, MD, 20771.

Knyazikhin, Y., and Marshak, A., 2000. Mathematical aspects of BRDF modeling: adjoint problem and Green's function. Remote Sensing Reviews, 18(2-4): 263-280.

Maiersperger, T., Scaramuzza, P., Leigh, L., Shrestha, S., Gallo, K., Jenkerson, C., and Dwyer, J., 2013. Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data. Remote Sensing of Environment, 136: 1-13.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., and Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990-2000. IEEE Geoscience and Remote Sensing Letters, 3(1): 68-72.

Meroni, M., Colombo, R., and Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. Remote Sensing of Environment, 92(2): 195-206.

Myneni, R., 2012. MODIS LAI/FPAR product user's guide. Boston University: Boston, MA, USA, pp. 1-7.

Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., and Smith, G., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sensing of Environment, 83(1): 214-231.

Pedro, A., Morais, M., Rosado, J., SILVA, H., Serafim, A., Neves, R., Brito, D., Potes, M., Salgado, R., and LILLEBØ, I., 2011. Hydrological Modeling in temporary streams: A case study in Pardiela basin, Southern Portugal, The 12 nd International Specialized Conference on Watershed & River Basin Management. Internacional Water Association (IWA), pp. 13-16.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., and Nickless, D., 2004. Terra and Aqua MODIS products available from NASA GES DAAC. Advances in Space Research, 34(4): 710-714.

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., and Gao, F., 2013. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. 2331-1258, US Geological Survey.

Schumann, A., 1993. Development of conceptual semi-distributed hydrological models and estimation of their parameters with the aid of GIS. Hydrological Sciences Journal, 38(6): 519-528.

Shen, L., Li, Z., and Guo, X., 2014. Remote Sensing of Leaf Area Index (LAI) and a Spatiotemporally Parameterized Model for Mixed Grasslands. International Journal of Applied, 4(1).

Silleos, N., Strati, S., Cherif, I., Topaloglou, C., Alexandridis, T.K., Iordanidis, C., Stavridou, D., Monachou, S., Kalogeropoulos, C., and Bilas, G., 2014. Weekly time series of LAI maps at river basin scale using MODIS satellite data, 1st international GEOMAPPLICA conference, pp. 8-11.

Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., and Macomber, S.A., 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote Sensing of Environment, 75(2): 230-244.

Soudani, K., François, C., Le Maire, G., Le Dantec, V., and Dufrêne, E., 2006. Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. Remote Sensing of Environment, 102(1): 161-175.

Stenberg, P., 1996. Correcting LAI-2000 estimates for the clumping of needles in shoots of conifers. Agricultural and Forest Meteorology, 79(1): 1-8.

Tan, B., Woodcock, C., Hu, J., Zhang, P., Ozdogan, M., Huang, D., Yang, W., Knyazikhin, Y., and Myneni, R., 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. Remote Sensing of Environment, 105(2): 98-114.

Teixeira, A.H.d.C., 2010. Determining regional actual evapotranspiration of irrigated crops and natural vegetation in the São Francisco River Basin (Brazil) using remote sensing and Penman-Monteith Equation. Remote Sensing, 2(5): 1287-1319.

Tian, Y., Zhang, Y., Knyazikhin, Y., Myneni, R.B., Glassy, J.M., Dedieu, G., and Running, S.W., 2000. Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data. IEEE transactions on geoscience and remote sensing, 38(5): 2387-2401.

Topaloglou, C., Monachou, S., Strati, S., Alexandridis, T., Stavridou, D., Silleos, N., Misopolinos, N., Nunes, A., and Araújo, A., 2013. Modeling LAI based on land cover map and NDVI using SPOT and Landsat data in two Mediterranean sites: Preliminary results, First International Conference on Remote Sensing and Geoinformation of Environment. International Society for Optics and Photonics, pp. 879505-879505-9.

Townshend, J.R., and Justice, C.O., 1988. Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. International Journal of Remote Sensing, 9(2): 187-236.

Turner, D.P., Cohen, W.B., Kennedy, R.E., Fassnacht, K.S., and Briggs, J.M., 1999. Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. Remote Sensing of Environment, 70(1): 52-68.

van Vliet, A., de Groot, R.S., Bellens, Y., Braun, P., Bruegger, R., Bruns, E., Clevers, J., Estreguil, C., and Flechsig, M., 2003. The European Phenology Network. International Journal of Biometeorology, 47(4): 202.

van Wijk, M.T., and Williams, M., 2005. Optical instruments for measuring leaf area index in low vegetation: application in arctic ecosystems. Ecological Applications, 15(4): 1462-1470.

Vermote, E.F., El Saleous, N.Z., and Justice, C.O., 2002. Atmospheric correction of MODIS data in the visible to middle infrared: first results. Remote Sensing of Environment, 83(1): 97-111.

Vermote, E.F., Tanré, D., Deuze, J.L., Herman, M., and Morcette, J.-J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE transactions on geoscience and remote sensing, 35(3): 675-686.

Walker, J., De Beurs, K., Wynne, R., and Gao, F., 2012. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote Sensing of Environment, 117: 381-393.

Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., and Liang, S., 2004. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. Remote Sensing of Environment, 92(4): 465-474.

Wang, Q., Adiku, S., Tenhunen, J., and Granier, A., 2005a. On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sensing of Environment, 94(2): 244-255.

Wang, Q., Tenhunen, J., Dinh, N.Q., Reichstein, M., Otieno, D., Granier, A., and Pilegarrd, K., 2005b. Evaluation of seasonal variation of MODIS derived leaf area index at two European deciduous broadleaf forest sites. Remote Sensing of Environment, 96(3): 475-484.

Wang, Y., Tian, Y., Zhang, Y., El-Saleous, N., Knyazikhin, Y., Vermote, E., and Myneni, R.B., 2001. Investigation of product accuracy as a function of input and model uncertainties: Case study with SeaWiFS and MODIS LAI/FPAR algorithm. Remote Sensing of Environment, 78(3): 299-313.

Weiss, M., Baret, F., Smith, G., Jonckheere, I., and Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. Agricultural and Forest Meteorology, 121(1): 37-53.

Welles, J.M., and Norman, J., 1991. Instrument for indirect measurement of canopy architecture. Agronomy journal, 83(5): 818-825.

Zheng, G., and Moskal, L.M., 2009. Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. Sensors, 9(4): 2719-2745.

9 Appendix A

LDOPE is the software provided by LPDAAC in order to be able to process .hdf MODIS images. The interface used by the software is command line (MS-DOS). The user has to create custom scripts, depending on the type of processing needed to be applied on the downloaded MODIS images. In this study, MODIS product MCD15A2 needed conversion of the related layer (band) "fPAR-LAI QC" from the 16-bit format it was initially, to a simple format, where the pixels would be divided into only two classes; flagged with value 0 for data of good quality and with value 1 for data of any other quality. This process was applied in order to use only data where the quality was of high standard and there was an increased level of confidence associated with it. When the equivalent layers of EVI were examined for their accuracy, it was observed that the 0 (good data) values of LAI Quality Assurance (QA) layer were also 0 (good data) for the equivalent EVI pixel reliability layers and for all pixels in all study areas.

The script written for this purpose was the following:

```
 @echo off
cls
FOR %%G IN (MCD15A2*.hdf) DO (call :subroutine15A2 %%G)
GOTO :eof
:subroutine15A2
SET FF=%1
ECHO %FF%
unpack_sds_bits -sds=FparLai_QC -bit=0 %FF% -of=.\QC_outputs\QC_LAI.%FF:~0,16%.hdf
cp_proj_param -ref=%FF% -of=.\QC_outputs\QCLAI.%FF:~0,16%_coords.hdf
.\QC_outputs\QC_LAI.%FF:~0,16%.hdf
del .\QC_outputs\QC_LAI.%FF:~0,16%.hdf
```