Mapping Mediterranean natural vegetation species using heterogeneous endmember analysis of hyperspectral imagery

Jonathan Doelman
September 2012
Universiteit Utrecht
Department of Physical Geography, Faculty of Geosciences
Supervisors: Elisabeth Addink, Steven de Jong

Preface

This research report is the result of a fieldwork campaign carried out from September until October 2011 in southern France, and the subsequent data analysis and analysis work performed from November 2011 to January 2012. The project is part of the Master of Physical Geography at the Faculty of Geosciences of Utrecht University. Elisabeth Addink and Steven de Jong are greatly acknowledged for their dedicated and enthusiastic supervision, and Giovanni Buzzo and Rik van Benthem for their great company during the fieldwork.

Abstract

Vegetation is a crucial component of system Earth as it plays an essential role in the water balance and the carbon dioxide cycle. Knowledge on the distribution and dynamics of vegetation is therefore very important. Imaging spectroscopy is a promising tool to monitor and understand complex vegetation patterns and dynamics. To map vegetation species is one of the most challenging objectives, as reflectance spectra of vegetation are very similar. Conventionally, areas are classified per species which assumes homogeneous compositions. However, natural vegetation compositions are often mixed – i.e. heterogeneous. Mapping species in natural environments with a homogeneous approach is therefore difficult.

This research proposes a new method of species mapping: spectral unmixing on the basis of heterogeneous endmembers as opposed to homogeneous endmembers. Next to that, two upcoming analysis methods are compared to their conventional version: object-based compared to pixel-based image analysis and image analysis based on original compared to continuum removed reflectance spectra.

Six Mediterranean vegetation species were mapped in the Peyne catchment, southern France. These species typically occur in dense forests of mixed composition. Three different approaches were applied to investigate the effects: 1) linear spectral unmixing on the basis of heterogeneous and homogeneous endmembers. 2) Image analysis with an object-based approach and a pixel-based approach. 3) Image analysis with continuum removed reflectance spectra and original reflectance spectra. Lastly, the accuracy was assessed and correlations were checked to determine the differences in reliability.

It is concluded that linear spectral unmixing on the basis of heterogeneous endmembers produces substantially better results than linear spectral unmixing on the basis of homogeneous endmembers. Average root mean square error for all species for the heterogeneous approach is 23 compared to 32 for the homogeneous approaches.

No differences in accuracy were found for the object-based compared to the pixel-based approach, and for the image analysis based on original reflectance spectra compared to continuum removed reflectance spectra. A substantial negative correlation between average root mean square error per plot and the level of heterogeneity was found, indicating bias in favour of heterogeneous pixels.

Key words
Mediterranean vegetation patterns and species
Imaging spectroscopy: HyMap
Heterogeneous endmembers
Object-based image analysis
Continuum removed reflectance spectra

Contents

List of figures

List of tables

List of equations

List of appendices

List of acronyms

ARU – Arbutus unedo
CAS – Castanea sativa
ERA – Erica arborea
LAI – leaf area index
LOO analysis – leave-one-out analysis
LSU – linear spectral unmixing
OBIA – object based image analysis
PHL – Phillyrea latifolia
QIL – Quercus ilex
QPU – Quercus pubescens
RMSE – root mean square error
ROI tool – region of interest tool
SMA – spectral mixture analysis

1. Introduction

The monitoring of vegetation dynamics is increasingly important as climate change and human interference are expected to cause unprecedented change in the local and global environment (Solomon et al., 2007; Lambin et al., 2001). Vegetation is a crucial component of system Earth as it stores carbon dioxide, produces oxygen and acts as a buffer in the hydrological cycle. Moreover, it prevents soil erosion and produces a wide range of natural resources for human use. Therefore, it is important to understand and be informed about changes taking place in vegetation dynamics (Xie et al., 2008).

The Mediterranean ecosystem is especially vulnerable. Mean precipitation is expected to decrease which makes desertification a serious threat (Christensen et al., 2007), there is a long history of human overexploitation and population pressure continues to be very high (Shoshany., 2000). Due to intense human interference and distinct environmental conditions, ecosystems with unique floristic composition and characteristic vegetation dynamics developed. From the few studies that assessed region-wide vegetation trends, it was found that forest cover in the African and Asian part of the Mediterranean is decreasing by 2% per year, whereas forest cover is increasing with 1,5% in the European part. These developments threaten ecosystems in the south and east of the Mediterranean, whereas ecosystems in the north will experience substantial modifications.

In order to prevent deterioration of ecosystems and to initiate conservation and restoration activities, region-wide information on the state of ecosystems and its vegetation dynamics is required (Shoshany, 2000; Xie et al., 2008). Spaceborne, multispectral remote sensing is a suitable tool that is commonly used to provide region-wide information on the general distribution, basic dynamics and characteristics of vegetation. However, to gain more understanding of the occurring developments, and to detect small-scale modifications or deteriorations of the ecosystem, more detailed remote sensing is necessary. The conventional, multispectral remote sensing satellites are not able to detect the smaller, local differences in for example water content, nutrient availability or species distribution. These parameters are crucial however, to enhance understanding of trends and developments that take place.

Imaging spectroscopy (hyperspectral remote sensing) is a technique developed in the 1990s in which the reflected solar radiance of the Earth is measured by acquiring hundreds of images in registered, contiguous spectral bands in the visible and infrared wavelengths. In this way, much more detailed information of reflectance surfaces is obtained compared to conventional, multispectral remote sensing. On the basis of this data, more advanced research can be performed on vegetation dynamics and characteristics, opening up a wide range of possibilities for future studies and

applications (Ustin et al., 2004). Up to now, most imaging spectroscopy is performed from airplanes – i.e. airborne remote sensing. This puts limits on the spatial extent of analyses. The launch of NASA's experimental Hyperion instrument on the EO-1 satellite in the year 2000 announced the start of spaceborne imaging spectroscopy (NASA, 2012). In 2015, another hyperspectral instrument will be launched: the German EnMAP (EnMAP, 2012). Thus, imaging spectroscopy is a promising tool for vegetation research. It has the potential to provide the necessary detail on both small and large scales to monitor and understand complex vegetation dynamics.

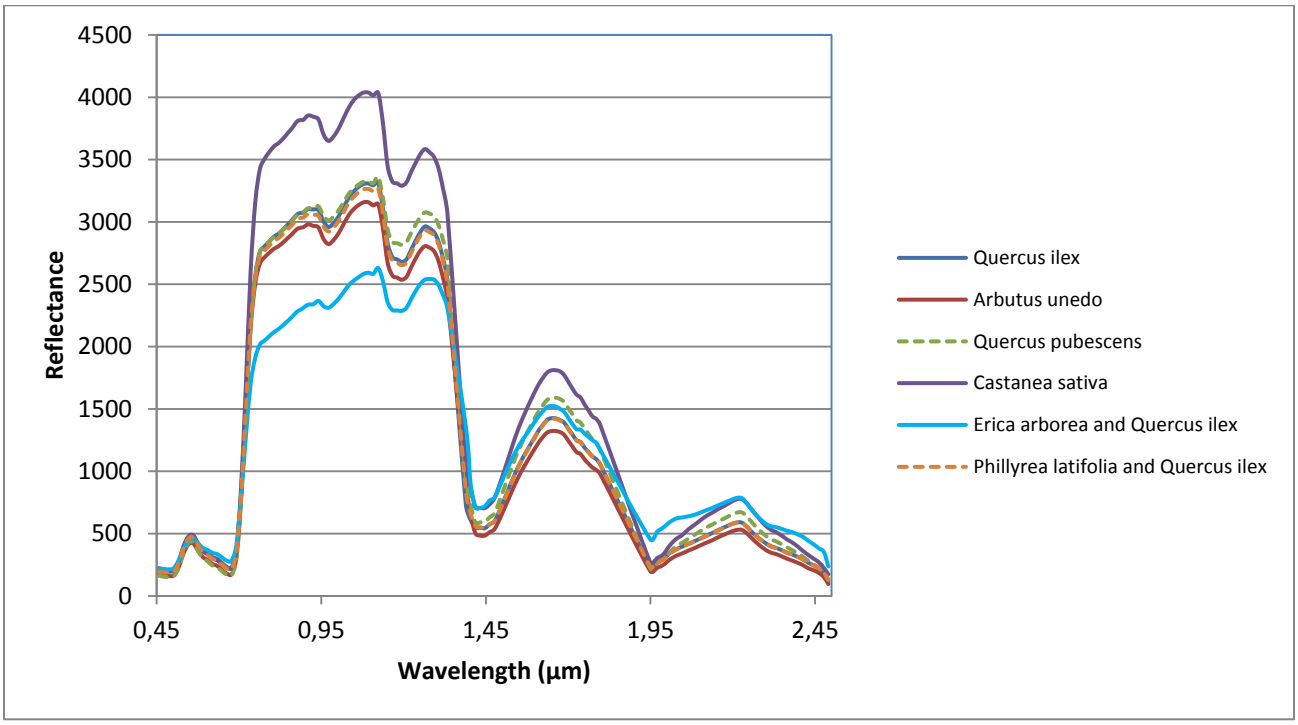

Fig. 1.1: example of reflectance spectra extracted from HyMap imagery of six important Mediterranean natural vegetation species

Vegetation research by means of imaging spectroscopy has addressed many different topics: for example biochemical properties, water content, biomass, stress level, leaf area index and more (Ustin et al., 2004). One of the most challenging topics regards the mapping of vegetation species. The main problem of this subject is that reflection spectra of different plant species are very similar. Figure 1.1 shows the spectra of six important Mediterranean species: their spectra are very similar and therefore difficult to distinguish from each other, illustrating the challenges faced by studies aiming to map vegetation species.

There is a substantial number of studies that investigated the possibilities to statistically distinguish between vegetation species on the basis of imaging

spectroscopy. Hyperspectral reflectance spectra of leaves have been measured in the laboratory, and reflectance spectra of canopies have been measured in the field. Statistical tests showed that there are indeed significant differences between reflectance spectra of species, making it possible to distinguish between them (Gong et al., 1997; Schmidt et al., 2001&2002; Van Aardt et al., 2001; Adam et al., 2009; Manevski et al., 2011). To apply this to hyperspectral imagery in order to map species has proven to be a bigger challenge. Clark et al. (2005) and Lucas et al. (2008) delineated large tree crowns in Australian eucalypt and Costa Rican tropical forests respectively, subsequently classifying them with reasonable success using discriminant analysis algorithms. Three other studies apply spectral unmixing to map vegetation species with hyperspectral imagery. This algorithm assumes that a spectrum is a mix of various components. Of each component, a known spectrum is required – a so-called endmember – on the basis of which unmixing is performed. The algorithm then calculates the degree of match of the endmembers in relation to the original reflectance spectrum. Li et al. (2005) and Youngentob et al. (2011) could discriminate rather accurately between two species in wetlands and eucalypt forests respectively. Sobhan et al. (2007) were able to determine the number of species present per pixel in the Apennine mountains with moderate accuracy.

Vegetation species mapping studies generally aim to create a classification of a certain area per species. This implies that each pixel can only be classified as one species, i.e. it has to be of homogeneous composition. This was the case for the tree crown mapping studies (Clark et al., 2005; Lucas et al., 2008) where only large, homogeneous trees were considered. The same applies to the unmixing projects of Li et al. (2005) and Youngentob et al. (2011) where large areas of vegetation cover were classified as one out of two species. However, natural vegetation often does not have a homogeneous composition. Vegetation species usually occur mixed on a sub-pixel level, resulting in heterogeneous canopy compositions. Sobhan (2007) acknowledged this, and investigated which different species were present per pixel, instead of classifying pixels as one species or another. They also performed their analysis by means of spectral unmixing, and based their unmixing on homogeneous endmembers – i.e. reflectance spectra of single species. Li et al. (2005) and Youngentob et al. (2011) did this as well. But when the natural vegetation considered is of heterogeneous composition, spectral unmixing might work better if the endmembers used are also of heterogeneous species composition. To find out if spectral unmixing produces better species mapping on the basis of heterogeneous endmembers compared to homogeneous endmembers – the conventional approach – was the main objective of this research.

The study area is located in the catchment of the river Peyne, and is characterized by dense forest cover. The dominant species are *Quercus ilex, Arbutus*

unedo, Quercus pubescens, Castanea sativa, Erica arborea and *Phillyrea latifolia*. These species most often occur mixed and single specimen canopies are relatively small. This results in heterogeneous canopy compositions on a sub-pixel level. Mapping species by classification into single species classes is therefore not suitable. The purpose is to map the abundance of each species throughout the study area. It is expected that spectral unmixing performs better on the basis of heterogeneous endmembers, because the vegetation composition in the area is very heterogeneous. Next to that, some species – *Erica arborea* and *Phillyrea latifolia* – only occur in combination with other species on a sub-pixel level. Therefore, these species cannot be detected if spectral unmixing on the basis of homogeneous endmembers is applied. To compare the results, spectral unmixing was performed on the basis of both homogeneous and heterogeneous endmembers (Fig 1.2).

Besides the effect of spectral unmixing on the basis of heterogeneous compared to homogeneous endmembers, the effects of two other methodologies on species mapping with imaging spectroscopy were explored: first, image analysis is traditionally performed on the basis of pixels. In recent years, object-based image analysis has become a popular alternative to this approach as a method to improve mapping results (Blaschke, 2010). The object-based approach applies image segmentation by grouping together pixels with high spectral similarity. This creates objects – e.g. patches of similar vegetation composition – and is reported to improve results for studies on vegetation parameters (Addink et al., 2007). In this research, it was tested if the object-based approach yields better results than the pixel-based approach by performing species mapping with both methodologies (Fig. 1.2). Second, imaging spectroscopy generally uses original reflectance spectra as measured by airborne or spaceborne spectrometers. By applying the continuum removal procedure, spectra can be normalized which reduces differences in brightness and enhances reflectance features (Clark and Roush et al 1984). It is known to improve results in mineral detection (Kruse et al., 1993). Also in vegetation studies enhanced results are reported, though with less confidence (Schmidt & Skidmore, 2003, Youngentob et al., 2011). Both the conventional, and the continuum removed approach were applied in this research to assess the positive or negative effects of both methodologies on species mapping (Fig. 1.2). Finally, the accuracies of the vegetation mapping results were assessed to check the reliability of the applied methodologies.

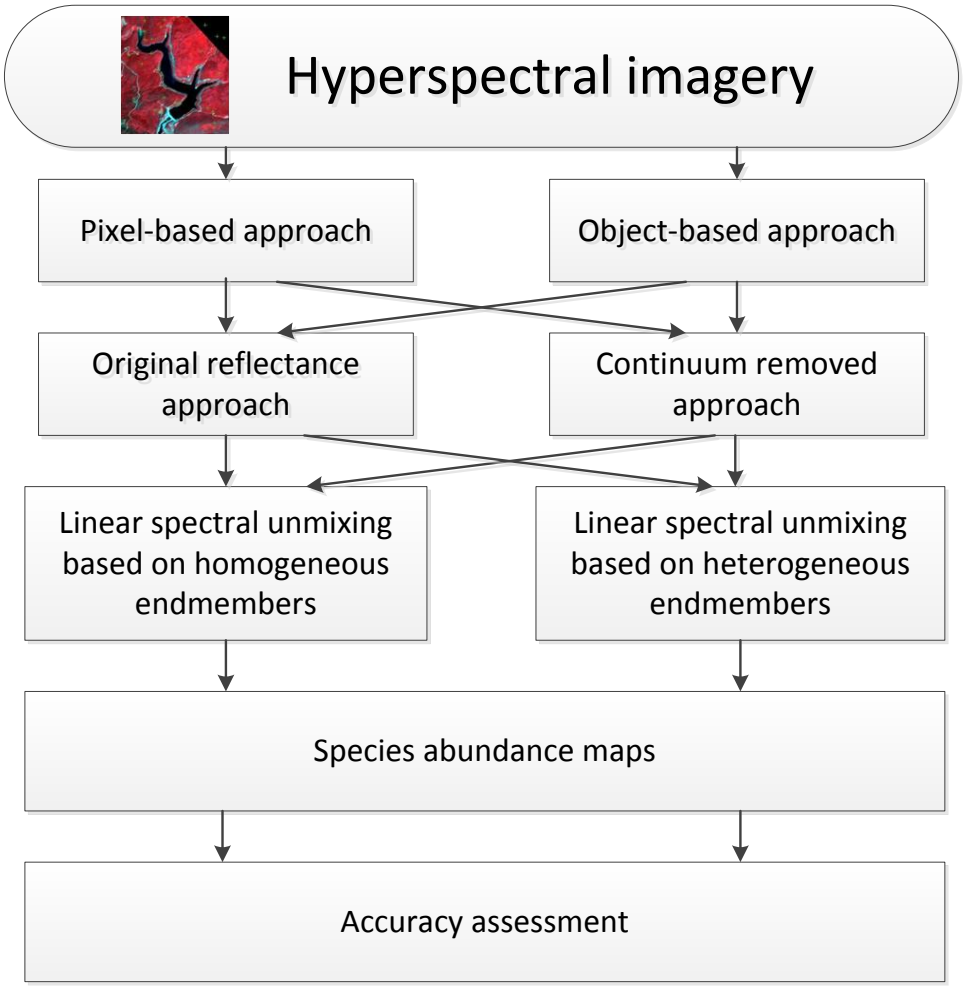

Fig. 1.2: diagram of the species mapping process

To summarize, in this research the following five research questions were investigated:

1. Is it possible to map the spatial abundance of six natural vegetation species in the forested area of the Peyne river catchment in southern France?
2. Which spectral unmixing methods yields the best results: a new proposed method on the basis of heterogeneous endmembers or a conventional method on the basis of homogeneous endmembers?
3. Which image analysis approach yields the best results: the method of object-based image analysis or the traditional method of pixel-based image analysis?
4. Which image analysis approach yields the best results: analysis based on continuum removed reflectance spectra or analysis based on original reflectance spectra?
5. Which species mapping methodology yields the highest accuracy, and are there any correlations or biases?

2. Background

In this chapter, the academic background of the main topics of this research is regarded. First, Mediterranean vegetation properties are discussed. Subsequently, imaging spectroscopy and some related methodologies will be elaborated upon. Finally, the application of imaging spectroscopy to vegetation is regarded.

2.1 Mediterranean vegetation

The Mediterranean climate is characterized by warm, dry summers en temperate, wet winters. Water availability is limited during the summer which reduces plant growth and requires vegetation to adapt. Therefore, evergreen sclerophyll species are dominant in most Mediterranean ecosystems (Joffre & Rambal, 2001; Ritter, 2006). These species have an advantage over deciduous species due to their strong stomatal control which limits loss of moisture. They retain a larger percentage of their nutrients by not shedding leaves every year. Therefore, they can efficiently sustain themselves in nutrient-poor environments. The fact that they are evergreen means that they can be productive at any time of the year when moisture availability is sufficient and temperatures are high enough. This gives evergreen species an advantage over deciduous species.

In the Mediterranean ecosystems of southern France roughly four vegetation types can be distinguished: mixed deciduous oak forest, maquis, garrigue and landes. Mixed deciduous oak forest is considered to be the climax vegetation type in large parts of the Mediterranean, but due to human induced soil degradation its current extent is limited (Miller & Hajek, 1981; Nijland, 2011). Maquis, garrigue and landes are more degraded types of vegetation, which are described by Tomaselli (1981) as matorral: he defines this as ‘a formation of woody plants, whose aerial parts are not differentiated into trunks and leaves, because they are much ramified from the base’. Three types are defined: tall, middle and short matorral respectively. In practice, it turns out that no clear distinction between the various groups is possible: types are often intermixed and transitions are gradual.

2.1.1 *Mixed deciduous oak forest*

The mixed deciduous oak forest is dominated by *Quercus pubescens* and *Quercus ilex*. Typically it grows ten to twenty meters high and has substantial undergrowth, most often consisting of *Buxus sempervirens* (Fig. 2.1). Other deciduous tree species are occasionally found, most importantly *Castanea sativa* and *Acer monopessulanum*. The mixed deciduous oak forest is mainly present in areas with relatively more precipitation, less agricultural land use and limited soil degradation.

Fig. 2.1: mixed deciduous oak forest vegetation type

2.1.2 *Maquis (tall matorral)*

Maquis or tall matorral is the tallest and densest version of matorral. Typically it grows between two and five meters high, with little undergrowth as limited light penetrates the canopy. Ground and standing litter is often abundant due to unfavourable climatic conditions for organic decomposition. *Quercus ilex* and *Arbutus unedo* are the dominant species of the maquis vegetation type (Fig. 2.2). *Erica arborea* often occurs when canopy cover is slightly discontinuous. Occasionally, *Phillyrea latifolia* is encountered.

Fig. 2.2: Maquis vegetation type (tall matorral)

2.1.3 *Garrigue (middle matorral)*

Garrigue or middle matorral is typically discontinuous of nature, has a height between a half and two meters, and is dominated by many different species. Most abundant are *Quercus coccifera*, *Erica arborea* and more scattered *Quercus ilex*, *Pistacia lentiscus* and *Spartium junceum*. The undergrowth consists of many herbaceous species such as *Cistus albidus*, *Thymus vulgarus*, *Ulex parviflorus* and grasses (Fig. 2.3).

2.1.4 *Landes (short matorral)*

Landes or short matorral typically has very sparse vegetation and many areas with rocks or bare soil. Height is generally less than half a meter and it is dominated by *Quercus coccifera, Cistus albidus*, *Thymus vulgarus*, *Ulex parviflorus* and *Lavendula latifolia* (Fig. 2.4).

Fig. 2.3: Garrigue vegetation type (middle matorral)

Fig. 2.4: Landes vegetation type (short matorral) with pine forest in the distance

2.2 Imaging spectroscopy

In imaging spectroscopy the reflected solar radiance of the Earth is measured by acquiring hundreds of images in registered, contiguous spectral bands in the visible and infrared wavelengths. Thus, it is possible to obtain a complete reflectance spectrum for each picture element (van der Meer & de Jong, 2002). In this way, detailed spectral data is available for every pixel of an observed area. From this, a wide range of information on surface and object properties can be derived.

In this research, airborne hyperspectral imagery was used. This type of imagery provides continuous spectra of objects at the Earth surface with high spectral and spatial resolution. It was expected that information on vegetation can more successfully be derived with this type of imagery as opposed to more conventional multispectral imagery, because the larger number of bands contain more information. A disadvantage is the spatial limitation, as all measurements have to be made from airplanes. Satellite imagery has much larger geographic coverage, but on the other hand has lower spectral and spatial resolution. However, spaceborne imaging spectroscopy is experiencing significant developments: in recent years NASA successfully launched and operated the Hyperion instrument (NASA, 2012), and in 2015 Germany will launch EnMAP (EnMAP, 2012). Both have spatial resolutions of about 30 meters, and measure reflectance in more than 200 spectral bands. The possibilities of hyperspectral remote sensing are expanding rapidly.

The imagery used has been acquired by the HyMap hyperspectral scanner from the Australian company HyVista, which was manufactured by Integrated Spectronics Pty Ltd. It provides imagery with 126 bands in the wavelength region from 0.45 to 2.5 µm, and bandwidths between 15-20 nm. The instantaneous field of view is 2.5 mrad along track and 2.0 mrad across track, the field of view is 61.3 degrees (512 pixels) and the ground instantaneous field of view varies from three to six meters. The sensor consists of four modules (Table 2.1) (HyVista, 2012):

Module	Spectral Range	Bandwidth across module	Average spectral sampling interval
VIS	0.45 – 0.89 µm	15 – 16 nm	15 nm
NIR	0.89 – 1.35 µm	15 – 16 nm	15 nm
SWIR 1	1.40 – 1.80 µm	15 – 16 nm	13 nm
SWIR 2	1.95 – 2.48 µm	18 – 20 nm	17 nm

Table 2.1: HyMap sensor modules technical details (HyVista, 2012)

In imaging spectroscopy many different techniques are developed to analyse imagery data. In this study, several methods were applied. The theory and some academic background of these methods is discussed in the following paragraphs.

2.2.1 *Object-based image analysis*

In object-based image analysis image segmentation is applied as one of the first steps in the analysis process. Neighbouring pixels with high spectral similarity are grouped into meaningful objects that form the basis of the subsequent analysis. This approach has become an increasingly popular tool over the last decade in image analysis for remote sensing (Blaschke, 2010). With advances in remote sensing techniques – specifically continuously increasing spatial resolutions – also small objects on the ground can be distinguished which enhances the possibilities of analyses based on objects.

This approach has various advantages. First, it reduces the effect of geometric inaccuracies, which is crucial in linking image data and ground observations. Second, the subjects of study are objects with actual meaning instead of artificial, squared pixels – in the case of this research vegetation patches. Third, the object-based approach yields additional variables than can be used in the analysis and can significantly improve results (Lillesand et al., 2008). Specifically, variables that are based on the relationships among objects such as connectivity and proximity to similar objects. In this way object-based image analysis is more similar to human visual interpretation of images because it takes contextual arguments into account. Addink et al., (2007) have shown that also for vegetation parameters, an object-based approach yields more accurate results than a pixel-based approach. They modelled biomass and leaf area index on the basis of HyMap imagery. It turned out that the object-based approach yields more accurate results than the pixel-based approach, and that the object-based approach has a certain optimal object size which produces the best results.

2.2.2 *Continuum removed image analysis*

The continuum removal procedure is regularly used in remote sensing studies. Especially in mineral detection it is acknowledged to improve results (Kruse et al., 1993). Also in vegetation studies it has led to better outcomes, though these are sometimes criticized (Schmidt & Skidmore, 2003, Youngentob et al., 2011). The procedure is a normalisation technique that fits a curve over the original spectrum – a so-called “convex hull”. The original data is subtracted from this curve and subsequently standardized to produce a normalized spectrum with a range from zero to one (Fig. 2.6). In this way, the technique normalizes brightness while emphasizing reflectance features (Clark & Roush, 1984).

In vegetation studies continuum removal is occasionally known to enhance species detection accuracy. Schmidt & Skidmore (2003) found mixed results in their research on spectral discrimination between coastal wetland vegetation types: in the visible wavelengths spectral discrimination was enhanced, whereas it was weakened in

the near-infrared and the shortwave infrared. Youngentob et al., (2011) on the other hand, found an accuracy for *Eucalyptus* subgenera of 75% for original reflectance spectra as opposed to 83% for continuum removed spectra.

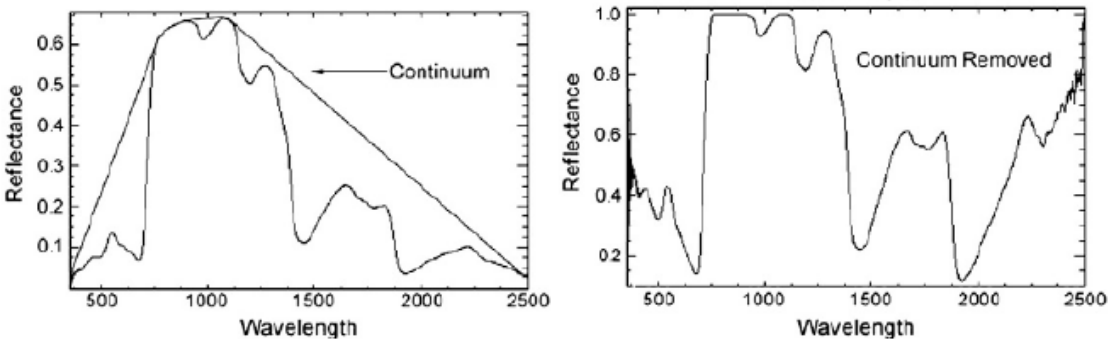

Fig. 2.6: continuum removal for a vegetation spectrum (Youngentob et al., 2011)

2.2.3 *Linear spectral unmixing*

The most common approach in remote sensing to extract information from images is by assigning pixels to a specific class on the basis of a classification procedure. The reflectance of a certain pixel usually is a mixture of various classes present within the pixel on the ground. For this reason, assigning only one class to a pixel leaves out interesting information (van der Meer & de Jong, 2002).

Singer & McCord (1979) showed that mixing of photons on a macroscopic scale can be considered a linear process. This means that the resulting reflectance spectrum is a linear summation of the classes present at a surface, multiplied by their relative fractions. ENVI applies the following linear mixing model to produce linear spectral unmixing results:

$$R_i = \sum_{j=1}^{N} f_j * Re_{ij} + \varepsilon_i$$

Eq. 2.1: linear spectral unmixing model (van der Meer & de Jong, 2002)

Where R_i is the reflectance of the modelled spectrum in image band i, f_j is the fraction of each endmember j, Re_{ij} is the reflectance of the endmember spectrum j in band i, N is the number of endmembers and ε_i is the residual error.

2.3 Imaging spectroscopy and vegetation studies

2.3.1 *Spectral behaviour of vegetation*

The high-resolution spectra of vegetation display some interesting features that directly relate to the physical characteristics of vegetation. A vegetation spectrum can be subdivided in three regions (Fig. 2.5): the high absorption area in the visible light

range (400-700 nm), the low absorption area in the near infrared region (700-1300 nm) and the mid-infrared region with characteristic water absorption bands (1300-2600 nm). The exact characteristics differ per plant species and with its status, e.g. age, health or phenological state (McCoy, 2005).

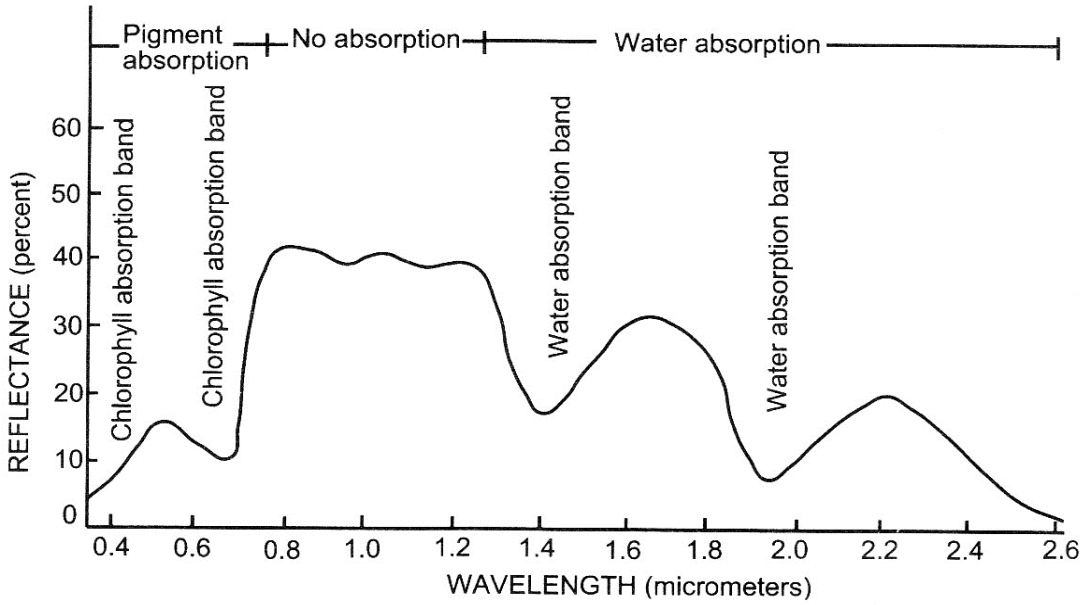

Fig. 2.5: an example of a typical vegetation spectrum (McCoy, 2005)

The high absorption area in the visible light range is most importantly caused by chlorophyll present in vegetation (400-700 nm). These are the pigments that absorb radiative energy which is used to produce organic material. As this is the only part of the spectrum that is visible it determines the visual appearance of vegetation. Typically, most light absorbed by chlorophyll is around 420, 490 and 660 nm which is red and blue light (van der Meer & de Jong, 2002). Consequently, there is a reflectance peak around 500 nm indicating limited preference to absorb radiation around this wavelength. Light around this wavelength is green and thus causes the distinctive green color of vegetation. Chlorophyll most importantly determines the reflectance characteristics in the visible area of the spectrum, but also other pigments such as carotenoids and xantophylls are responsible. Especially when plants are under stress or during leaf senescence these become more distinct as chlorophylls degrade quickly, consequently reducing light absorption (Merzlyak et al., 1999).

In the low absorption area in the near infrared region most radiation is reflected or transmitted (700-1300 nm). As opposed to the visible light range the interaction in this area is determined predominantly by leaf structure rather than pigments (Roy, 1989). This results in the relatively high percentages of reflectance and transmittance, which in turn causes the distinctive feature between approximately 690 to 720 nm. This

is the red edge, which is typical for vegetation and which is often used as an indicator of vegetation status (Horler et al., 1983). The most prominent of these indicators is the normalized difference vegetation index (NDVI) which is a ratio of the red light on one side of the red edge and of infrared on the other side of the red edge.

The mid-infrared area is most importantly typified by major water absorption bands (1300-2600 nm). In the spectrum they are located around 1400 and 1900 nm, and also 1200 and 2500 nm are important water absorption locations (Fig. 2.5). In healthy vegetation, these are generally masked by the water absorption features. In case of leaf senescence they can become distinguishable, which results in observable features caused by biochemicals such as lignin, cellulose, starch, proteins and nitrogen (McCoy, 2005).

2.3.2 Applications of imaging spectroscopy in vegetation studies

Imaging spectroscopy has only recently become a popular tool in remote sensing vegetation studies (Xie et al., 2008). Compared to multispectral imagery, hyperspectral imagery has much more potential to study vegetation because it does not have dozens, but several hundreds of bands. Consequently, hyperspectral imagery contains much more information and is therefore very suitable to investigate vegetation characteristics, or discriminate between vegetation types or species.

Many studies have been executed to explore the potential of imaging spectroscopy to discriminate between vegetation species. Because vegetation spectra are very similar (Fig. 1.1), this is a difficult task. But when enough spectral samples are taken, statistical discriminant analysis generally find significant differences between plant species (Gong et al., 1997; Schmidt et al., 2001&2002; Van Aardt et al., 2001; Adam et al., 2009; Manevski et al., 2011). This proves the possibility to discriminate between species, and indicates the possibility to also discriminate between species in a geographical perspective.

The mapping of vegetation species on the basis of hyperspectral imagery has been moderately successful though. An approach that yielded reasonable results classifies tree crowns: individual tree crowns are delineated and subsequently classified on the basis of their reflectance spectra by means of discriminant analysis. This yielded good results in the tropical forests of Costa Rica, with an overall accuracy of 86% (Clark et al., 2005). Also, good results were achieved in Australian forest communities with accuracies of 87% and 76% for training and testing datasets respectively (Lucas et al., 2008). For this approach however, tree crowns need to be large and distinguishable from other vegetation. In natural environments, this is often not the case. Several other studies apply spectral unmixing algorithms on the basis of endmembers to classify vegetation. Li et al. (2005) discriminated two important species in Californian wetlands

for which they report high accuracy, and so did Youngentob et al. (2011) for Eucalypt forests in Australia (83%). Sobhan used the same method to map species present per pixel in the Apennine mountains of Italy, and reported an accuracy of 69%. These studies show that there are good possibilities to map species with hyperspectral imagery. However, the methodologies vary widely and the reliabilities of the studies are often difficult to compare.

3. Study area

The study area of this research is located in southern France in the Languedoc-Rousillon region, Hérault department. The field campaign focused on the catchment of the river Peyne, a small tributary of the Hérault river (Fig. 3.1) located on the southern slopes of the Massif Central. An important landmark in the centre of the study area is the Barrage des Olivettes, an artificial lake used for irrigation and flood protection. The study area has an approximate size of 16 km^2. Two major cities in the region are Montpellier (45 km to the east) and Béziers (25 km to the south).

This region is a long-term study area of the Utrecht University Department of Physical Geography. Therefore, a wide range of data from both field and earth observation campaigns is available. Next to that, a large variety of geological substrates, vegetation types and land-uses is present in a relatively small area, and stable and dry summer weather improves the possibilities for successful remote sensing. For these reasons, the Peyne study area is suitable for this research project.

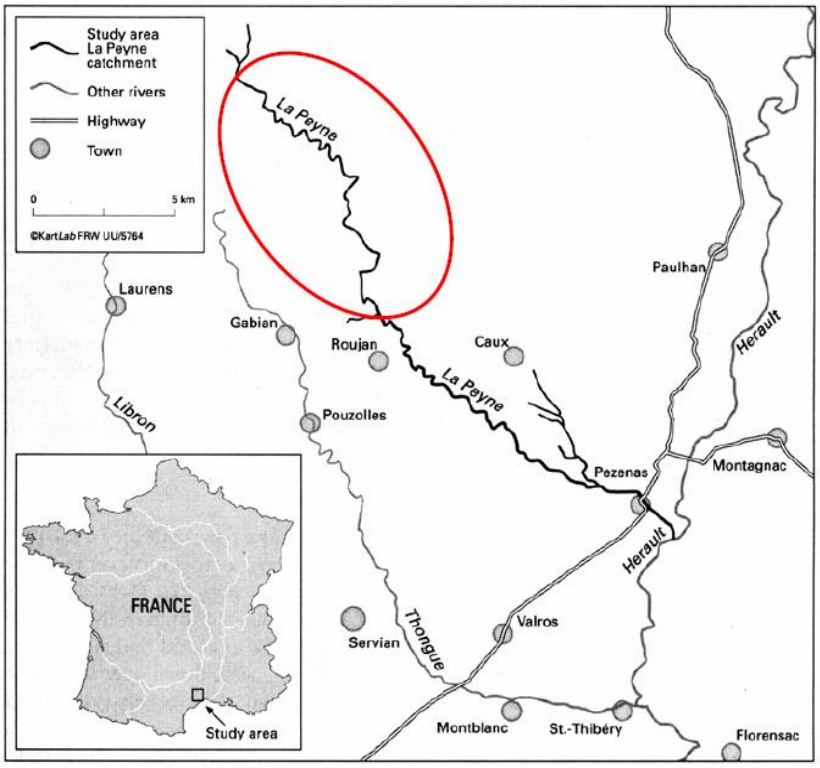

Fig. 3.1: location of the Peyne catchment area

3.1 Climate

The climate in the study area is characteristic for a Mediterranean region. Typically, Mediterranean regions are located between 30° and 45° latitude and to the east of oceans or seas. Their climate is characterized by warm, dry summers and temperate, wet winters (Fig. 3.2), and they form a transition zone between humid, temperate climates and warm, arid climates. Therefore, they are classified as sub-humid to semi-arid (Joffre & Rambal, 2001; Ritter, 2006).

On a local level the climate also shows a significant trend: the elevation of the study area increases when moving from the coastal plain of the Mediterranean Sea up the Massif Central in the north. In line with the increasing elevation precipitation rates also increase (Fig. 3.3)(Bonfils, 1993). This in turn affects the characteristics of the vegetation in the area with relatively more drought-resistant species in the south and more deciduous species in the north.

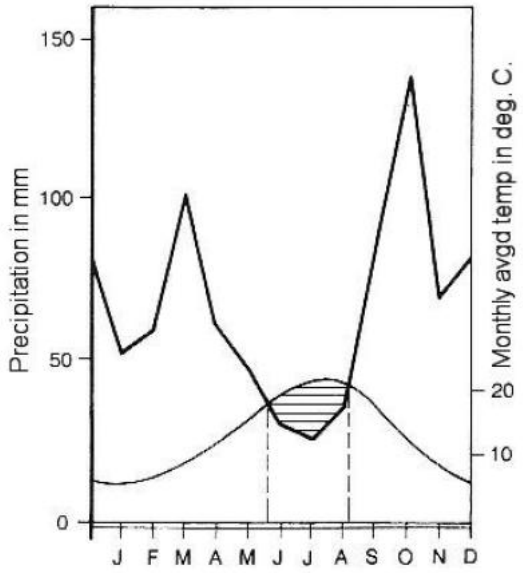

Fig. 3.2: Annual distribution of precipitation and temperature in St. Chenin, Hérault dept. (I.N.R.A., 1993)

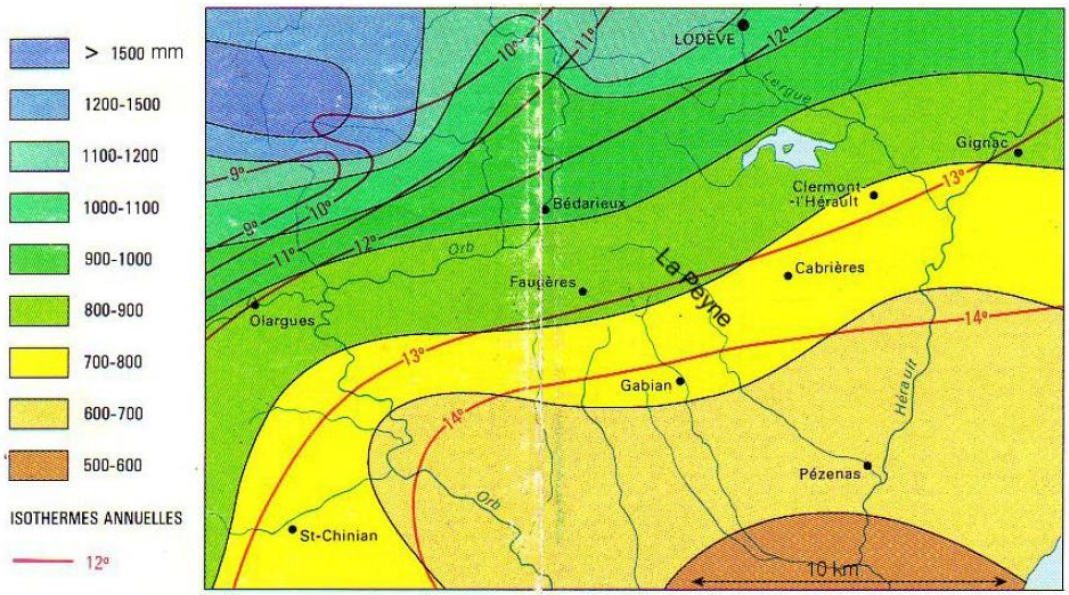

Fig. 3.3: Yearly average precipitation map including isotherms (Bonfils, 1993)

3.2 Geology and soils

The study area is located on the southern slopes of the Massif Central where the transition to the coastal plains of the Mediterranean Sea takes place. One of the characteristics of this transition zone is the complex geology of the area with a high variability of geological substrates. The tectonic forces that formed the Massif Central deformed the original geological formations and caused various substrates from different ages to surface at similar elevations. The most dominant types are sedimentary sandstones, mudstones and limestones, all of which are moderately metamorphized (Nijland, 2011; Sluiter, 2005). In addition to these geologically old formations, volcanic tuffs and basalt outflows are present at many locations. These originated in different periods of volcanic activity during the Quaternary, remnants of which are found throughout the Massif Central. The most recent lithological formations are found on the valley bottoms and consist of alluvial substrates. However, these only comprise a small part of the area's surface.

Soil depth in the Peyne catchment area is generally shallow due to slow soil formation and human-induced soil degradation. Soils only develop slowly in this area because heat and humidity have peaks in different seasons, while both are necessary to decompose organic and mineral material in order to form soils. In addition, the steep slopes induce high rates of erosion which is enhanced by forest removal for agricultural purposes and intense rainfall events in spring and autumn. For these reasons, only shallow soils can be found which classify as lithosols, regosols or luvisols (Bonfils, 1993).

3.3 Land use

The Peyne catchment area has a long history of human presence and agricultural exploitation. Especially during the 18th and early 19th century rural population density was high, agriculture was intense and forest cover was relatively limited. With the onset of industrialization large portions of the rural population migrated to urban centres. Consequently, agricultural land was abandoned and often reverted to natural vegetation and forest regrowth (Fig. 3.4). Currently, some of the most suitable land is used for agriculture – mainly viticulture – and residential purposes. The majority of the Peyne catchment area consists of former agricultural land that is occupied by natural vegetation and forest regrowth.

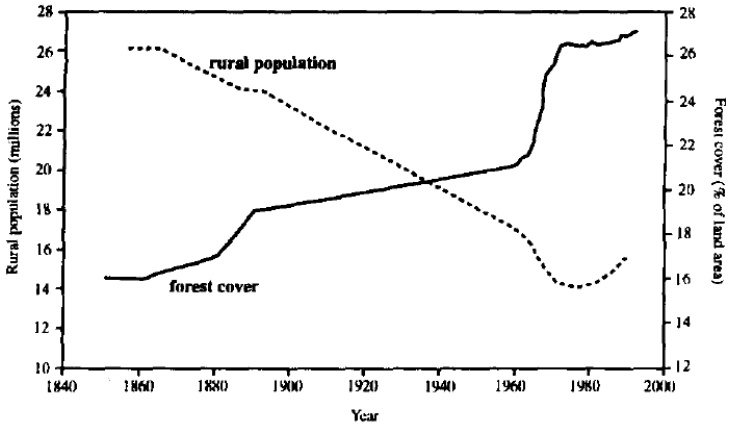

Fig. 3.4: forest area and population trends in France (Mather et al., 1999)

3.4 Natural vegetation

The natural vegetation in the study area is mainly comprised of the maquis (tall matorral) vegetation type. In the north, mixed deciduous oak forest is present as this area is characterized by higher precipitation rates and less soil erosion. At the southern boundary of the area garrigue (middle matorral) and landes (low matorral) occurs.

In the Peyne catchment large tracts of forest were cultivated as coppices. This practice was used for firewood and charcoal production, and had a large effect on the nature of the forest as trees often grow with many separate stems from a single root system. The cultivation of coppices still continues, but on a smaller scale. Many forests are affected by the practice though, causing forest degradation.

This study focused on maquis and mixed deciduous oak forest vegetation types. The canopy of these vegetation types is mostly comprised of six species, which are the main subject of research: *Quercus ilex, Arbutus unedo, Quercus pubescens, Castanea sativa, Erica arborea* and *Phillyrea latifolia* (Fig. 3.5). If garrigue and landes were included, many more different species would be present which would make species

mapping much more challenging. Therefore, the study area and the strata used in the field campaign are mainly located in areas with maquis and mixed deciduous oak forest vegetation types (Fig. 3.6).

Fig. 3.5: the six vegetation species that are the main subject of this research: A) *Quercus ilex,* B) *Arbutus unedo,* C) *Quercus pubescens,* D) *Castanea sativa,* E) *Erica arborea* and F) *Phillyrea latifolia*

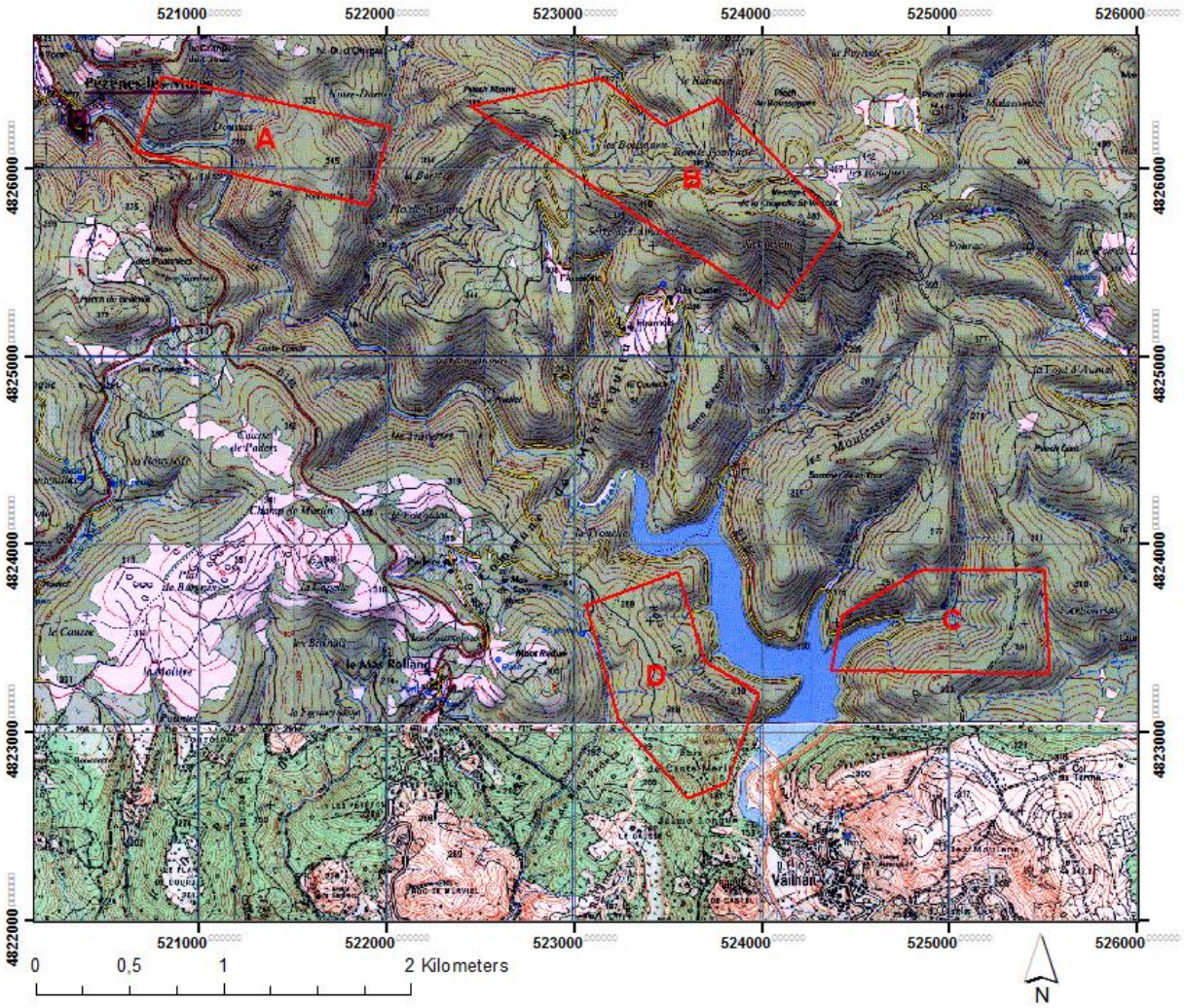

Fig. 3.6: study area with the strata used as a basis for the field work

4. Methods

The methodology of this study consists of three main components: the field data collection, the image analysis and the accuracy assessment. The field data collection was carried out in September and October 2011 during a field campaign in the Peyne catchment, southern France. The following sections will describe the methods of field plot selection and description in detail. Subsequently, the image analysis methodology will be discussed in the following order: the used image material, the pixel- compared to the object-based analysis, the original reflectance spectra compared to the continuum removed reflectance spectra and the linear spectral unmixing. Finally, the accuracy assessment is regarded: the application of the leave-one-out approach, the root mean square error, the confusion matrices and the correlation tests.

4.1 Field data collection

The goal of the field work was to obtain a representative database with the characteristics of the natural vegetation in the Peyne catchment area, which is sufficiently large for image interpretation and accuracy assessment. For this purpose, 169 field plots were selected of which a detailed quantitative description of site and vegetation characteristics was made. In this section the working methodology in the field is discussed, which consisted of plot selection and plot description.

4.1.1 *Plot selection*

The selection of the field plots had to satisfy a range of requirements: 1) they had to be representative of the range of natural vegetation in the selected study area in terms of vegetation composition and characteristics, 2) a random component had to be included in the selection in order to prevent statistical bias, and 3) they had to be reasonably accessible in order to make data collection logistically feasible. Four strata were selected that were expected to satisfy the aforementioned requirements (Fig. 3.5). The selection was based on image analysis performed in advance of the field work, data available from earlier studies in the area, and road accessibility.

Within the selected strata, a nested random sampling method was adopted to determine the plot locations (Fig. 4.1) (as suggested by Nijland, 2009). An initial nodal point - determining the first plot location - was identified in situ based on accessibility and species composition of interest. From this point, seven follow-up points were sequentially selected. These points were each located at a distance of 50 m from the previous point, and located in a randomly chosen direction. The random direction was determined on the basis of a list of random numbers in the range of 0 to 360. The

follow-up points were found using compass and a GPS device, which ensured the plot selection to be statistically unbiased.

External factors could influence the data acquisition through for example weather conditions or phenology (e.g. leaf shedding). To prevent the concentration of these influencing factors in the data of single measurement locations, the strata were visited in a rotating order. In this way, the influence of possible external factors that showed a trend over time was minimized.

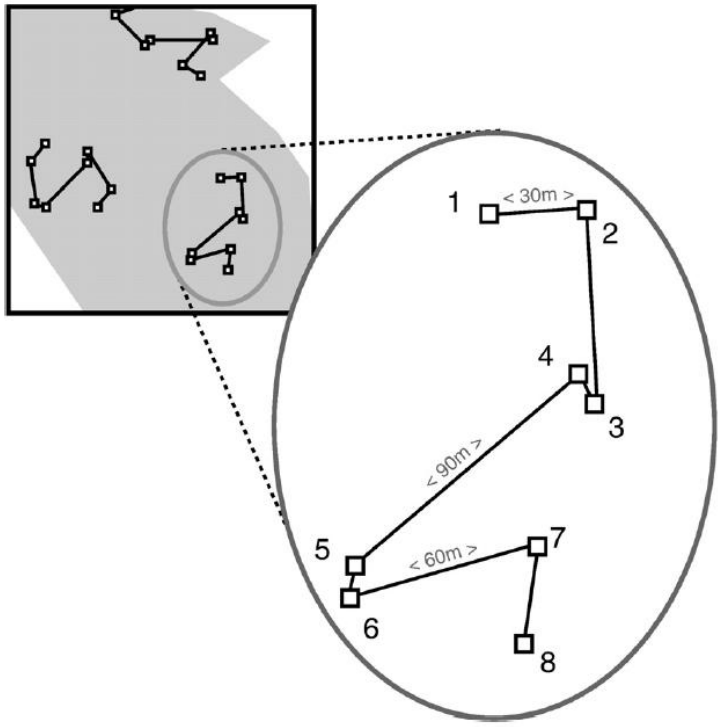

Fig. 4.1: example of stratified nested random sampling scheme (Nijland et al, 2009)

4.1.2 *Plot description*

At each selected location a square plot of 25 m^2 (5x5 m) was measured and laid out using a tape-measure. The choice for 25 m^2 plots is a compromise between the need to describe at least one pixel (the HyMap imagery has a spatial resolution of 5x5 m) and the logistical limitations of describing larger plots within a reasonable time span.

The following plot characteristics are registered:

- Date and time
- Stratum number, cluster number and site number
- Geometric location: XY coordinates measured with a Garmin eTrex [WGS84, UTM31N]
- The vegetation cover density by visual estimate: low, middle or high

- The land cover class according to Tomaselli (1981): high matorral (dense, discontinuous or scattered) or middle matorral (dense, discontinuous or scattered)
- The plot average vegetation height
- The percentages of plot cover type by visual estimate that are covered by a closed canopy or not, and if their respective surfaces are vegetated or not: canopy, out-canopy vegetated or out canopy non-vegetated
- The percentages per species that make up the canopy composition
- The type and species of vegetation that are dominant in out-canopy vegetated area
- The characteristics of the surface cover of the out-canopy non-vegetated area

In addition, an overview photo of the plot was taken as a reference and notes were taken if the plot had any characteristics that might influence the spectral analysis or the measurements.

Species	**Formula**	**Source**
Quercus ilex *Quercus pubescens* *Castanea sativa* other species	ln AB = 4.900 + 2.277 * ln D50	Ogaya et al., 2003 Quercus Ilex
Arbutus Unedo *Erica arborea* *Buxus sempervirens* *Phillyrea latifolia* other species	ln AB = 4.251 + 2.463 * ln D50	Ogaya et al., 2003 Arbutus unedo
AB: aboveground biomass individual [g] *D50: stem diameter at 50cm [cm]*		

Table 4.1: Equations used for the field estimates of aboveground biomass (Ogaya et al., 2003)

Secondly, the aboveground biomass of each plot was estimated using allometric relations between stem diameter and biomass as described by Ogaya et al. (2003). They developed relations for the two species most dominant in the Peyne catchment area: *Quercus ilex* and *Arbutus unedo*. The biomass of other species was also estimated using either of these two relationships based on morphological similarity to *Quercus ilex* and *Arbutus unedo* (Table 4.1). For this purpose, in each plot the diameter at breast height of all specimens with circumference larger than two centimetres was determined by measuring the circumference at breast height (±50 cm).

Finally, the vegetation cover was measured using hemispherical photographs (Fig. 4.2). In each plot five photographs were taken: four at each corner and one in the middle. These photographs were processed using the Can-Eye software package

developed by INRA, Avignon France (Weiss & Baret, 2010; Jonckheere et al., 2004; Weiss et al., 2004). The program uses five photos per plot to increase the accuracy. First, the user has to mask out any areas with solar flare or non-canopy objects. Then, a binary classification is made indicating canopy cover and non-canopy cover based on a 324 colour matrix. This is an important step and therefore user supervised. Finally, the program calculates the vegetation cover, leaf area index and other vegetation characteristic parameters by means of a look-up-table comparison based on the gap-fraction distribution of the canopy. In addition, a correction is applied for non-random leaf distribution: the clumping factor. Thus, an objective measure of the vegetation cover and other vegetation characteristics is obtained.

Fig. 4.2: example of hemispherical photograph used for vegetation cover estimates

4.2 Image analysis

The image analysis process consists of several steps. This section describes these aspects in the same sequence as they have been applied. Most operations were performed with the remote sensing software ENVI 4.7 and the office suite Microsoft Excel 2003. In addition, the GIS-software packages eCognition 8.7 and ArcGis 10 were used.

4.2.1 *Imagery*

The airborne hyperspectral HyMap imagery used was acquired over the Peyne catchment area on the 23rd of July 2008. It consist of 126 spectral bands (Table 2.1) and

has a spatial resolution of 5x5 m, and it has been pre-processed to correct for noise and geometric errors according to the methods described by Schläpfer & Richter (2002) and Richter & Schläpfer (2002). For research purposes the study area was limited to the central portion of the HyMap image which is mostly covered by dense, natural vegetation (Fig. 3.5) A NDVI mask was applied to exclude non-vegetated areas from the analysis. In ENVI, bands 16 (0.6773 µm) and 24 (0.7920 µm) were used to calculate the index. The cut-off value to produce the mask was designated at 0.5.

4.2.2 *Pixel- and object-based analysis*

It was evaluated whether a pixel-based or an object-based approach works best for species mapping. In order to draw conclusions on this question, image analysis was performed using both approaches. Eventually, the respective results were compared (Fig. 4.3).

In the pixel-based approach, the corrected, masked HyMap image was used. However, when a geographical location was taken in the field there was generally a certain error involved in the range of three to six meters. As the pixel-size is five by five meters, it is common that a sample site does not exactly match a pixel. For this reason, an averaging procedure was applied. Each pixel was recalculated by means of the ENVI convolution tool, to be the average of nine pixels: the pixel itself and its eight neighbouring pixels. In this method it is assumed that the area in the immediate surrounding is representative of the pixel considered. In this way, the effects of geometric error are limited because essentially spectral data from fifteen by fifteen meter plots are considered.

In the object-based approach, eCognition 8.7 was used to perform image segmentation. The multiresolution segmentation algorithm was applied in order to exploit all spectral information embedded in the HyMap image. The scale parameter, which determines the maximum heterogeneity of the objects, was set to 25. This scale value resulted in objects in the natural vegetation area of approximately 100 to 200 pixels in size (2500 m^2 to 5000 m^2). Thus, average object diameter is approximately 50 to 70 meters, which is in the range of optimal pixel size for leaf area index predictions according to Nijland et al. (2009). The ratio between spectral and shape determines the importance of the reflectance spectra in the segmentation process as opposed to shape. It was set at 0.9 : 0.1, in order to maximize the influence of reflectance spectra while still including some effect of the shape. This reduced the number of objects with unusual shapes, that would cause difficulties in the analysis process and were more likely caused by landscape characteristics than by vegetation species composition. This effect was enhanced by setting the compactness criterion to 1.0.

For every object, the average reflectance spectrum was calculated. In this way, all pixels that were located in the same object had the same reflectance spectrum. Subsequently, the objects and their values were exported in vector format – an ESRI shapefile – to the ArcGis software environment. In ArcGis, every layer was identified as a separate attribute (i.e. 126 attributes). Using the composite bands tool all attributes were re-assembled into a multidimensional image with a cell-size of five by five meters. This image was then exported to ENVI and analyzed with the same methodology as the image used for the pixel-based analysis.

4.2.3 *Original reflectance spectra and continuum removed spectra*

It was also investigated whether image analysis based on the original reflectance spectra performs better than image analysis based on continuum removed spectra. The beneficial effects of continuum removal in vegetation studies are ambiguous. To evaluate the effects in the case of this study, the image analysis was performed on the basis of both the original reflectance spectra and the continuum removed spectra (Fig. 4.3).

To study the original reflectance spectra, the pixel-based and the object-based HyMap imagery of the Peyne study area was used. In this case the images did not have to be further adapted. For the continuum removed approach the same HyMap imagery was processed with the continuum removal tool in the ENVI software environment. Thus, two additional images were produced: a pixel-based, continuum removed HyMap image and an object-based, continuum removed HyMap image.

4.2.4 *Linear spectral unmixing*

The first step of the linear spectral unmixing procedure was the identification of suitable endmembers. Two sets of endmembers were selected to form a spectral library based on solely homogeneous vegetation patches, and to form a spectral library based on homogeneous as well as heterogeneous vegetation patches.

A wide range of Mediterranean vegetation species is present in the Peyne catchment area. The image analysis however, focused on the natural forest vegetation present in the selected study area. On the basis of the field data it was decided that six species would be studied: *Quercus ilex*, *Arbutus unedo, Quercus pubescens*, *Castanea sativa*, *Erica arborea*, and *Phillyrea latifolia.* These are the species that have substantial presence in the top canopy (Appendix A) and therefore determine the characteristics of the reflectance spectra. Some other species were found, but they have limited influence on the reflectance spectra and would be difficult to detect. For this reason, they were left out of the analysis.

Homogeneous endmember training plot classes	Number of training plots	Heterogeneous endmember training plot classes	Number of training plots
100% *Quercus ilex*	15	100% *Quercus ilex*	15
100% *Arbutus unedo*	8	75% *Quercus ilex* 25% *Arbutus unedo*	11
100% *Quercus pubescens*	12	50% *Quercus ilex* 50% *Arbutus unedo*	7
100% *Castanea sativa*	5	75% *Arbutus unedo* 25% *Quercus ilex*	6
		50% *Quercus ilex* 50% *Quercus pubescens*	3
		100% *Quercus pubescens*	12
		75% *Erica arborea* 25% *Quercus ilex*	3
		75% *Erica arborea* 25% *Arbutus unedo*	2
		75% *Quercus ilex* 25% *Phillyrea latifolia*	6
		100% *Castanea sativa*	5

Table 4.2: endmember classes for linear spectral unmixing and their respective number of training plots

For the homogeneous spectral library, four endmembers were selected that had sufficient training plots with 100% of a single species to build a library: *Quercus ilex*, *Arbutus unedo*, *Quercus pubescens* and *Castanea sativa* (Table 4.2; Appendix B). Three of the *Castanea sativa* training plots were situated outside the study area, but were nonetheless selected to make the *Castanea sativa* training-set more reliable. The percentage *Arbutus unedo* in the top canopy of the 100% *Arbutus unedo* training plots was in the range of 70% - 90%. These were assumed to represent 100% *Arbutus unedo* training plots to be able to perform an analysis based on homogeneous plots including these four prominent species.

For the heterogeneous spectral library, various combinations of the six species were selected. The selection of these endmembers was based on the heterogeneous combinations of species that occur most regularly. Ten endmembers were selected based on three homogeneous and seven heterogeneous groups of training plots (Table

4.2; Appendix B). Again, three *Castanea sativa* training plots located outside the study area were selected to make the *Castanea sativa* training-set more reliable.

In ENVI, the spectra of all training plots were located using their geographic coordinates and exported from the respective HyMap imagery as ASCII files using the region of interest tool. This was done for all image analysis approaches: both original reflectance and continuum removed spectra, from the object based as well as the pixel-based images. Subsequently, the spectra were imported into Excel where the averages per class were calculated. These averages were the endmembers that were used to build spectral libraries.

In total, eight spectral libraries were built that were subsequently applied to the respective HyMap images using linear spectral unmixing (Fig. 4.3). This resulted in eight abundance images with four bands for the homogeneous spectral libraries and ten bands for the heterogeneous spectral libraries. For every pixel, these bands displayed the fractional values per endmember.

Finally, the spectral unmixing output maps were used to produce species abundance maps. It was assumed that the positive fractional values indicated the relative abundance per endmember. Therefore, in each abundance map all positive fractional values were summed. This was performed in ENVI by masking all the negative values, and then summing all positive values. Subsequently, in each band the pixels with positive values were divided by the total positive value of the pixels in the same locations. Thus, the relative fraction of each endmember with respect to the total positive value was calculated. In this way, linear spectral unmixing on the basis of homogeneous endmembers produced species abundance maps for four species. Linear spectral unmixing on the basis of heterogeneous endmember classes produced abundance maps for mixed classes, of which the fractions per species were summed in order to produce six species abundance maps (Fig. 4.4).

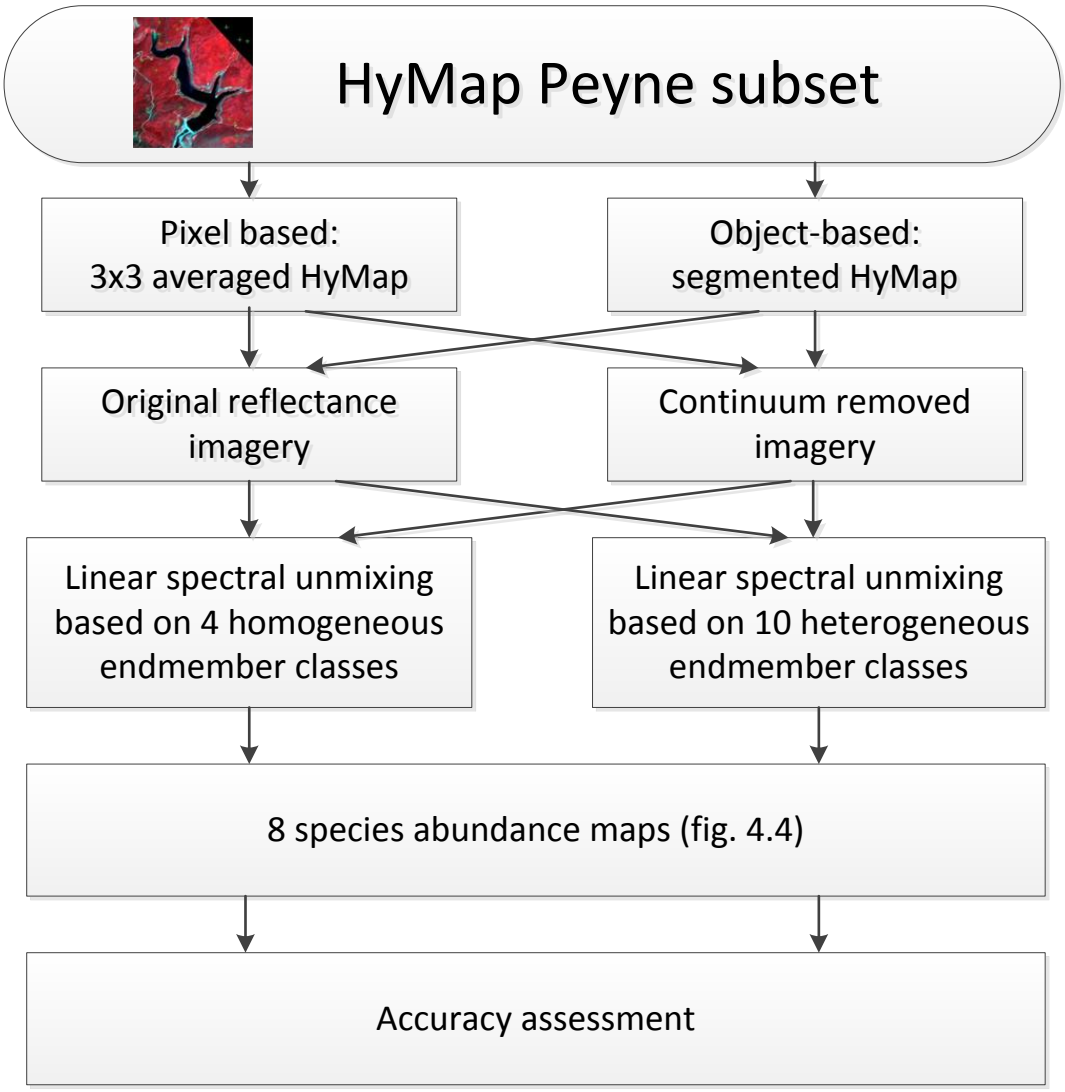

Fig. 4.3: diagram of the image analysis process

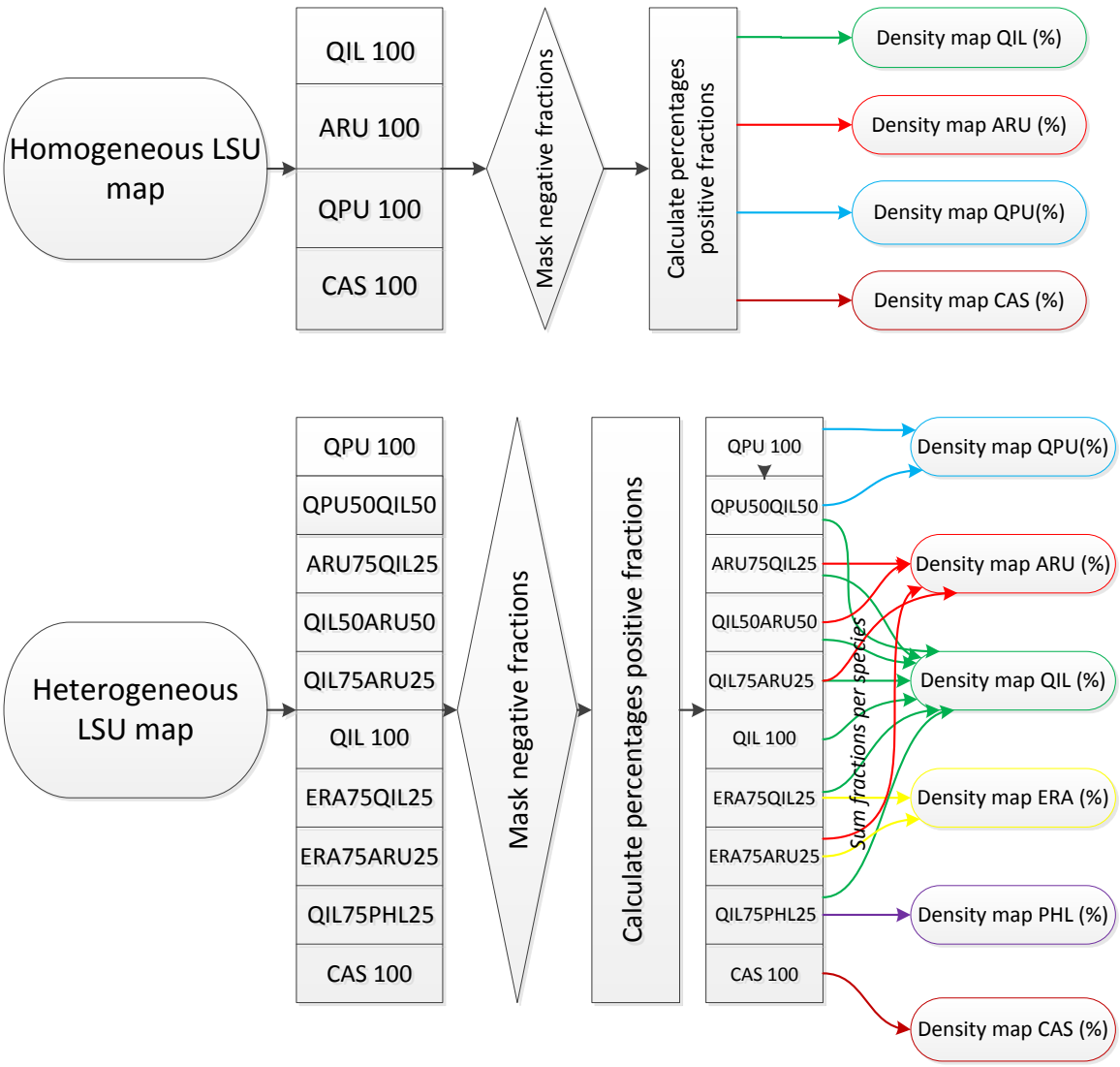

Fig. 4.4: diagram of the species density map creation procedure

4.3 Accuracy assessment

The accuracy of the predicted species abundance maps was assessed by comparing the species abundance per plot estimated during the field work, with the predicted species abundance per plot. This was done in combination with a leave-one-out analysis approach in order to maximize the assessment results. On the basis of this comparison, root mean square error (RMSE) of the species abundance percentages per plot, confusion matrices of the dominant species per plot and correlation parameters between the RMSEs and various physical vegetation parameters were calculated to determine the accuracy.

4.3.1 *Leave-one-out analysis approach*

In this research only a limited amount of field data was available. Specifically for the less common species few plots were collected. The conventional approach is to split the field data into a training group and a validation group. In the case of this study however, it was preferable to use the collected plots for training as well as validation to maximize efficiency on the basis of the limited field data. However, if the same plots were used to train a model, and subsequently to validate the results, the accuracy results would be biased. To avoid this, a leave-one-out analysis approach was applied.

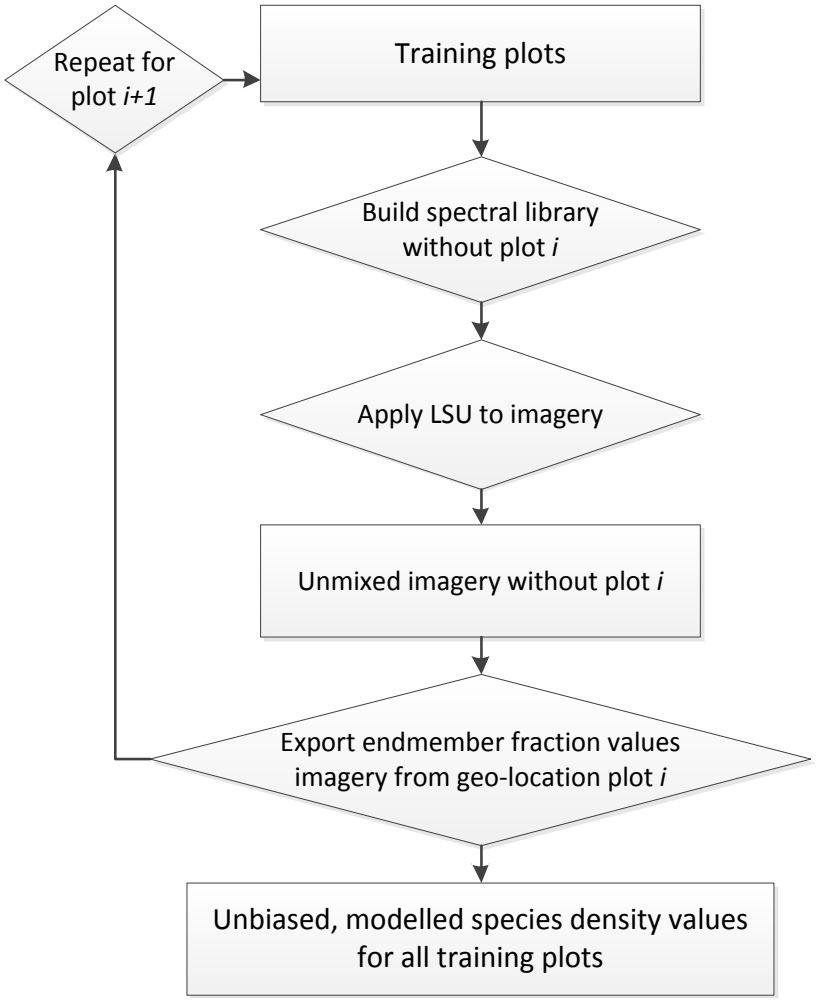

Fig. 4.5: diagram of the leave-one-out approach analysis procedure, used in the accuracy assessment

Linear spectral unmixing was applied on the basis of four homogeneous and ten heterogeneous endmembers (Fig. 4.3). These were trained with 40 and 70 plots respectively (Table 4.2). To use these spectral data for both training and validation, an extra step was included in the linear spectral unmixing stage: for every plot that was used both for training and validation, a separate spectral library was built. In this

spectral library, the plot concerned was not used for training – i.e. it was left out. Subsequently, linear spectral unmixing was performed on the basis of this library: the unmixed fractional values that were predicted for the geographic location of the plot considered, were used to assess its accuracy. This procedure was repeated for every training plot, and applied to the various study cases (Fig. 4.5). Thus, all selected plots were utilized to train the unmixing libraries, and all plots in the natural forest area could be used to assess the accuracy.

4.3.2 *Root mean square error of the species abundance per plot*

The most important results of the image analysis were the species abundance maps. The accuracy of the species abundance maps was assessed by comparing the predicted percentages per plot, for each image analysis approach, with the field data per plot. This was performed by means of the root mean square error (RMSE), a measure frequently used to determine the difference between predicted values and values actually observed on the ground. Thus, it provides a straightforward indication of prediction accuracy. The formula applied in this study is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Eq. 4.1: equation for the root mean square error (Lu, 2003)

Where n is the number of observations, i is the plot considered, $\hat{y}_i$ is the predicted value of the plot and y_i the observed value of the plot.

4.3.3 *Dominant species confusion matrices*

Confusion matrices are often used in remote sensing studies to check whether pixels of which the ground truth is known are correctly classified, and what type of errors occur in the model (e.g. Dennison et al., 2003). It is a specific table layout that shows whether a model makes correct predictions by comparing its classifications to the ground truth. A confusion matrix has two sub-accuracy measures and one overall measure: the user's accuracy checks whether a pixel that is predicted to be in a certain class, actually has this value on the ground. The producer's accuracy does the opposite: it checks whether a pixel that is known to have a certain value on the ground, is correctly classified by a model. These two measures provide insight in the type of over- or underestimation that a model produces, and other trends that might occur. The overall accuracy combines the former two by calculating the percentage correctly classified pixels, thus providing a general accuracy measure (Congalton, 1991).

This study focused on species fractions per pixel instead of specific classes per pixel. This made it impossible to apply confusion matrices. Nonetheless, it would be interesting to apply this analysis because it investigates errors of inclusion and

exclusion, and whether these over- or underestimate the predicted or the groundtruth data. Hence, for all plots it was calculated which species was dominant, i.e. which species had the largest fraction. By means of confusion matrices, it was checked whether the image analysis approaches found the same dominant species in these geographic locations. Specifically for the groundtruth data, it occasionally occurred that two species had the same fraction. These plots were omitted from the confusion matrix analysis, as neither of the species present was actually dominant.

4.3.4 *Correlations between RMSE and vegetation parameters*

The aim of this study was to map different vegetation species on the basis of differences in spectral reflectance. The accuracy of the image analysis was assessed by the RMSE and confusion matrices analyses. However, these did not check if the spectral differences between vegetation species were responsible for high or low accuracies, or if other vegetation characteristics had an effect. Three vegetation characteristic parameters were available for every plot: vegetation cover, aboveground biomass and average canopy height. Also, for every plot it was known how many different species are present, which is a measure of heterogeneity. The correlation between the RMSE and these four parameters was calculated to evaluate if they affect the accuracy. The first three were checked as the variables are independent of the type of species. The heterogeneity was checked as it is an important aspect of the image analysis procedure.

A correlation coefficient produces an indication of the dependency of one variable on another. It provides insight in relations between the data considered, and in this way indicates for example causal relations. To calculate the coefficient for the mentioned cases, the Pearson's correlation was utilized:

$$Correlation(x, y) = \frac{\sum_{x=i}^{N}(x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{x=i}^{N}(x - \bar{x})^2 \sum_{x=i}^{N}(y - \bar{y})^2}}$$

Eq. 4.2: correlation equation (Isaaks, 1989)

In this formula the correlation between N measurements x and y is calculated, with $\bar{x}$ and $\bar{y}$ their respective means.

5. Results

The results section provides an overview of the field data and the results of the image analysis. In addition, the results of the accuracy assessment are presented. On the basis of this, conclusions are drawn regarding the research objectives.

5.1 Field data

In total, 169 plots were sampled (Fig. 5.1). The basic properties of each plot were described and geometric location was noted. Additionally, the canopy composition in percentage per species and the vegetation height were estimated, and aboveground biomass and vegetation cover were measured (Appendix A and C).

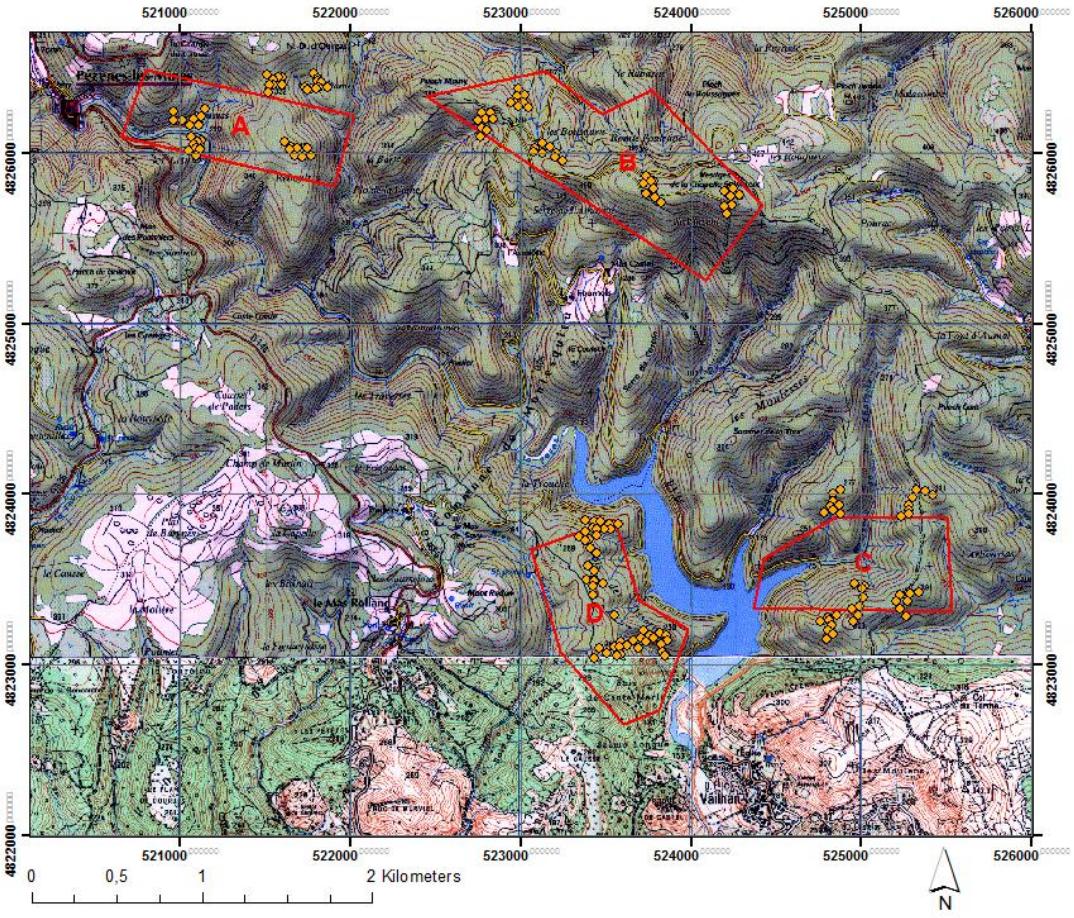

Fig. 5.1: topographic map of study area with strata and plots

The northern plots are mostly dominated by *Quercus ilex* (strata A and B) and *Quercus pubescens* (stratum A) (Table 5.1). This is in accordance with the dominance of mixed deciduous oak forest in the north. The southern plots are also characterized by *Quercus ilex*, but in combination with *Arbutus unedo* (strata C and D). In addition, *Erica arborea* has substantial presence in stratum D. The presence of *Phillyrea latifolia* is limited in all four strata. *Castanea sativa* is only found in a few occasions in stratum B.

stratum	**averaged percentage canopy composition (%)**					
	QIL	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**
all plots (n=169)	56	21	14	1	6	2
A (n=42)	48	4	46	0	0	1
B (n=40)	75	7	9	5	0	4
C (n=40)	55	40	1	0	3	1
D (n=47)	47	32	0	0	18	2

Table 5.1: averaged percentages of canopy composition per species

The northern plots (strata A and B) have a larger average height than the southern plots (strata C and D) (Table 5.2). The vegetation cover is nearly constant throughout the study area. The aboveground biomass shows a trend with relatively higher values in the north (strata A and B) and lower values in the south (strata C and D). This is related to the dominance of maquis in the southern part of the study area, as opposed to the deciduous oak forest in the north. The southern plots also have a lower average height compared to the northern plots. On the other hand, the heterogeneity shows that the northern plots (strata A and B) have on average fewer different species (1.7 – 1.8) than the southern plots (strata C and D) (2.3 – 2.5). In the northern plots, the canopy is often only composed of *Quercus ilex* or *Quercus pubescens*, and sometimes *Castanea sativa*. In the southern plots, *Quercus ilex* and *Arbutus unedo* usually occur mixed, and are often interspersed with *Erica arborea* and sometimes *Phillyrea latifolia*.

stratum	**av. height (m)**	**vegetation cover (%)**	**biomass (ton/ha)**	**heterogeneity (av. nr of species per plot)**
all plots	6.9	75	133	2.1
A	9.5	75	163	1.7
B	7.3	75	166	1.8
C	5.7	76	117	2.3
D	5.3	74	93	2.5

Table 5.2: averaged physical vegetation characteristics: average height, vegetation cover, aboveground biomass and heterogeneity

In accordance with the abundance of *Quercus ilex* as percentage of the canopy (Table 5.1), also the largest number of plots is dominated by *Quercus ilex* (Table 5.3). In the northern plots in stratum A, *Quercus pubescens* dominates a substantial number of plots. *Castanea sativa* is only present in stratum B, but when it is present it always dominates the canopy. In the southern plots, *Arbutus unedo* dominates a substantial number of plots in stratum C and D. *Erica arborea* only dominates a number of plots in stratum D. *Phillyrea latifolia* never dominates the canopy in this study area.

dominant species	QIL	ARU	QPU	CAS	ERA	PHL
all strata	99	25	22	2	10	0
A	21	0	19	0	0	0
B	32	2	3	2	0	0
C	23	12	0	0	1	0
D	23	11	0	0	9	0

Table 5.3: the number of plots in which the respective species are dominant

5.2 Image analysis

The image analysis resulted in eight sets of species abundance maps. The percentages canopy cover based on the plot locations are shown in Table 5.4, and data on species dominance in the plots in Table 5.5. An example of a set of abundance maps is shown in Figure 5.2. An interesting observation is the difference between the homogeneous and heterogeneous approaches. The homogeneous approaches have comparable abundances for *Quercus ilex*, *Arbutus unedo* and *Quercus pubescens* whereas the heterogeneous approaches are dominated by *Quercus ilex* (Table 5.4). The latter situation is in accordance with the field data (Table 5.1), where *Quercus ilex* occupies more than 50% of the canopy. Next to that, it is interesting that only the homogeneous and heterogeneous approaches result in substantial differences. The results from the pixel-based compared to the object-based approaches, and the approaches based on original reflectance spectra compared to continuum removed spectra do not differ much.

The same applies to the data on dominant species per plot (Table 5.5). The homogeneous approaches yield a comparable number of plots dominated by *Quercus ilex, Arbutus unedo* and *Quercus pubescens*. The heterogeneous approaches on the other hand yield the largest number of plots dominated by *Quercus ilex*. The latter situation is expected based on observations in the field (Table 5.3). Again, the differences between the pixel- and object-based approaches, and the analyses based on original and continuum removed spectra are limited.

The set of species abundance maps (Fig. 5.2) provides insight in the differences in spatial distribution of the six species produced by the object-based, continuum

removed, heterogeneous approach. This set of maps is displayed because this image analysis approach had the best results according to the accuracy assessment (section 5.3). *Quercus ilex* has high densities (± 50%) throughout the study area. Also *Arbutus unedo* has relatively high densities, most importantly in the central and southern parts. Both observations are in accordance with field observations. *Quercus pubescens* has lower densities, but is abundant in the north, which matches field observations. *Erica arborea*, *Phillyrea latifolia* and *Castanea sativa* have very low densities throughout the area. *Castanea sativa* does have some small, higher density spots in the north, which is plausible as this species mostly occurs there in small, homogeneous patches. *Erica arborea* has very low densities which is as expected because this species hardly ever dominates the canopy. However, it does have a relative high density in the south-easternmost corner of the study area. This also confirms expectations as this is a Landes area: this type of vegetation is more similar to *Erica arborea* dominated areas than the densely forested areas. *Phillyrea latifolia* has low presence throughout the study area (±7%). It is as expected that this species has low density, but on the basis of field observations a more patchy pattern is expected.

Image analysis approach	**stratum**	**average canopy cover (%)**					
		QIL	*ARU*	*QPU*	*CAS*	*ERA*	*PHL*
pixel-based, original reflectance, homogeneous	**all plots**	**30**	**26**	**36**	**8**		
	A	30	27	35	8		
	B	55	20	10	15		
	C	29	47	15	9		
	D	26	44	24	6		
pixel-based, original reflectance, heterogeneous	**all plots**	**45**	**25**	**18**	**1**	**8**	**3**
	A	46	18	24	1	8	3
	B	47	25	15	2	6	5
	C	44	29	15	2	8	3
	D	42	29	18	0	10	1
		QIL	*ARU*	*QPU*	*CAS*	*ERA*	*PHL*
pixel-based, continuum removed, homogeneous	**all plots**	**28**	**33**	**27**	**12**		
	A	26	23	38	13		
	B	44	20	20	16		
	C	20	46	21	13		
	D	24	41	27	9		
pixel-based, continuum removed, heterogeneous	**all plots**	**42**	**27**	**19**	**1**	**9**	**3**
	A	41	23	24	1	9	3
	B	48	23	18	1	7	4
	C	41	31	17	1	8	3
	D	40	30	16	0	11	2

		QIL	*ARU*	*QPU*	*CAS*	*ERA*	*PHL*
object-based, original reflectance, homogeneous	**all plots**	**32**	**36**	**30**	**2**		
	A	29	21	49	1		
	B	65	20	9	6		
	C	20	57	23	0		
	D	18	44	37	1		
object-based, original reflectance, heterogeneous	**all plots**	**48**	**30**	**7**	**1**	**10**	**4**
	A	50	21	10	1	11	6
	B	47	36	7	0	6	3
	C	50	35	6	1	6	2
	D	44	27	5	1	17	5
		QIL	*ARU*	*QPU*	*CAS*	*ERA*	*PHL*
object-based, continuum removed, homogeneous	**all plots**	**12**	**44**	**40**	**4**		
	A	10	31	54	6		
	B	24	36	33	7		
	C	7	59	32	2		
	D	7	48	43	2		
object-based, continuum removed, heterogeneous	**all plots**	**41**	**30**	**19**	**1**	**3**	**5**
	A	36	26	26	3	3	6
	B	44	30	18	2	2	4
	C	41	32	21	0	1	4
	D	44	32	14	1	4	6

Table 5.4: averaged percentages of canopy composition per species, per image analysis approach

Image analysis approach	**QIL**	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**
pixel based, original reflectance, homogeneous	62	68	30	9		
pixel based, original reflectance, heterogeneous	142	20	6	0	1	0
pixel based, continuum removed, homogeneous	50	61	41	15		
pixel based, continuum removed, heterogeneous	124	23	16	0	4	0
object based, original reflectance, homogeneous	58	59	50	2		
object based, original reflectance, heterogeneous	141	24	3	0	1	0
object based, continuum removed, homogeneous	5	83	77	3		
object based, continuum removed, heterogeneous	116	31	20	1	0	0

Table 5.5: the number of plots in which the respective species are dominant according to the various image analysis approaches*

* The application of continuum removal resulted in erroneous results at some plot locations. Therefore the pixel-based continuum removed image analysis misses data for two plots in the field dominated by ARU, and the object-based continuum removed image analysis misses data for one plot in the field dominated by QPU.

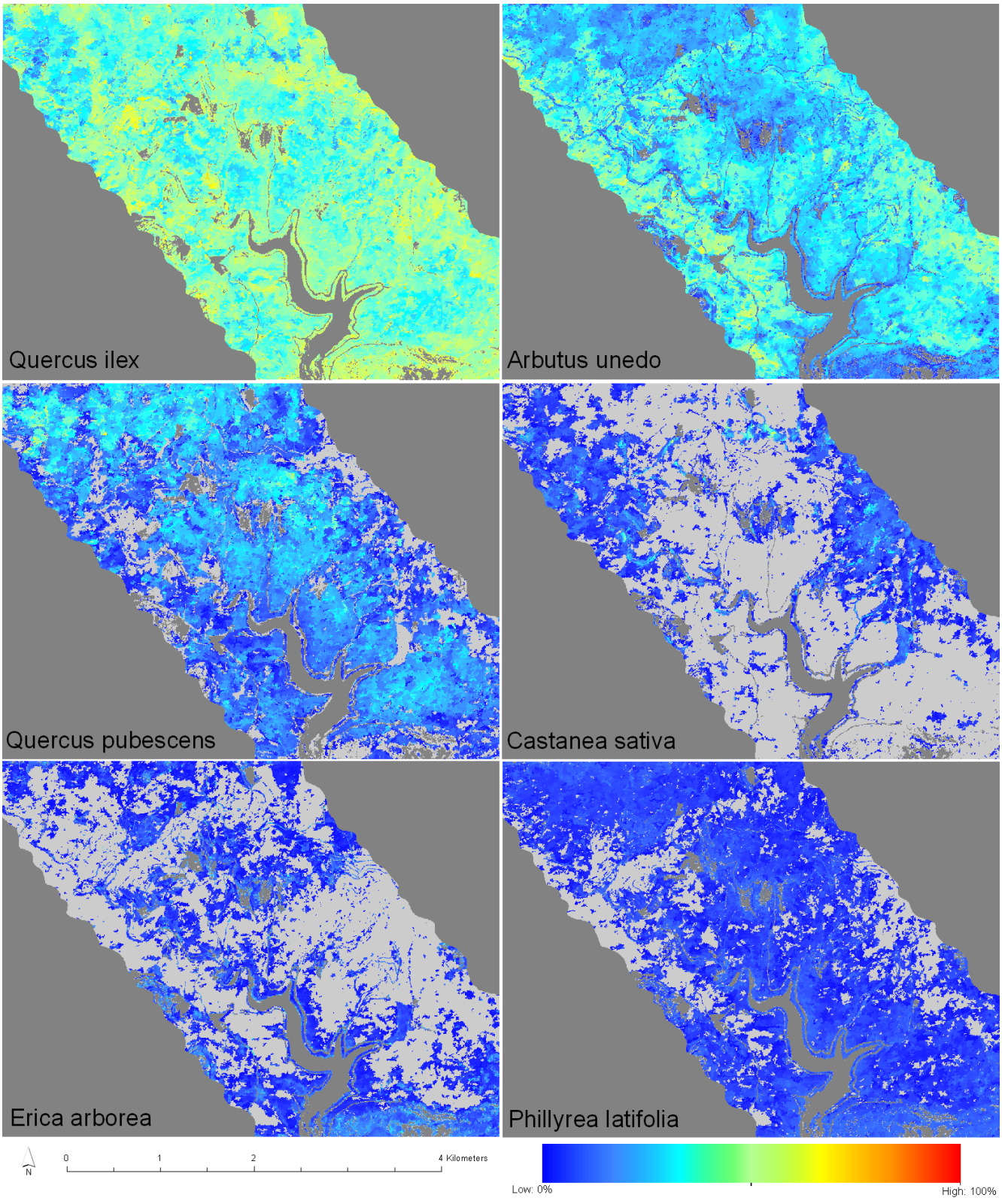

Fig. 5.2: abundance maps of six species based on the object-based, continuum removed, heterogeneous image analysis approach. In the light grey areas, the considered species is not present (i.e. 0%). The dark grey areas are masked out because there is too little vegetation present (NDVI < 0.5) or because they are located outside the hyperspectral image.

5.3 Image analysis accuracy assessment

5.3.1 *Root mean square error of the species abundance per plot*

The predicted species abundances per plot for each image analysis approach were compared to the field data per plot by means of the root mean square error (RMSE). The RMSE values that resulted from this comparison are rather large (Table 5.6). The values are on average in the same range as the observed, groundtruth values (Table 5.1). This indicates substantial errors.

Nonetheless, some interesting trends can be observed. 1) the pixel-based and object-based approaches do not yield different results: the RMSEs of the cases that are either pixel- or object-based are very similar. 2) the analyses based on original reflectance spectra compared to those based on continuum removed spectra also show little differences based on the RMSEs: average errors per species are similar when comparing the original reflectance approaches with the continuum removed approaches. 3) on the other hand, the differences between the RMSEs based on homogeneous spectra compared to those based on heterogeneous spectra do show substantial differences: of the four species predicted by all approaches, average RMSE for the homogeneous approaches is 32% as opposed to 23% for the heterogeneous approaches (Fig. 5.3). Clearly, the RMSEs of the homogeneous image analysis approaches are much larger than those of the heterogeneous image analysis approaches, which proves that the heterogeneous approach produces better mapping results than the homogeneous approach.

	Average RMSE per species (%)					
Image analysis approach	***QIL***	***ARU***	***QPU***	***CAS***	***ERA***	***PHL***
pixel based, original reflectance, homogeneous	40	37	31	18		
pixel based, original reflectance, heterogeneous	31	22	28	10	14	7
pixel based, continuum removed, homogeneous	42	33	34	23		
pixel based, continuum removed, heterogeneous	31	22	27	11	14	6
object based, original reflectance, homogeneous	41	34	32	9		
object based, original reflectance, heterogeneous	31	26	28	10	16	8
object based, continuum removed, homogeneous	52	44	35	12		
object based, continuum removed, heterogeneous	32	24	26	9	14	8

Table 5.6: average root mean square errors in percentages between the predicted species abundance per plot and the field observations per plot – averaged per species and per image analysis approach

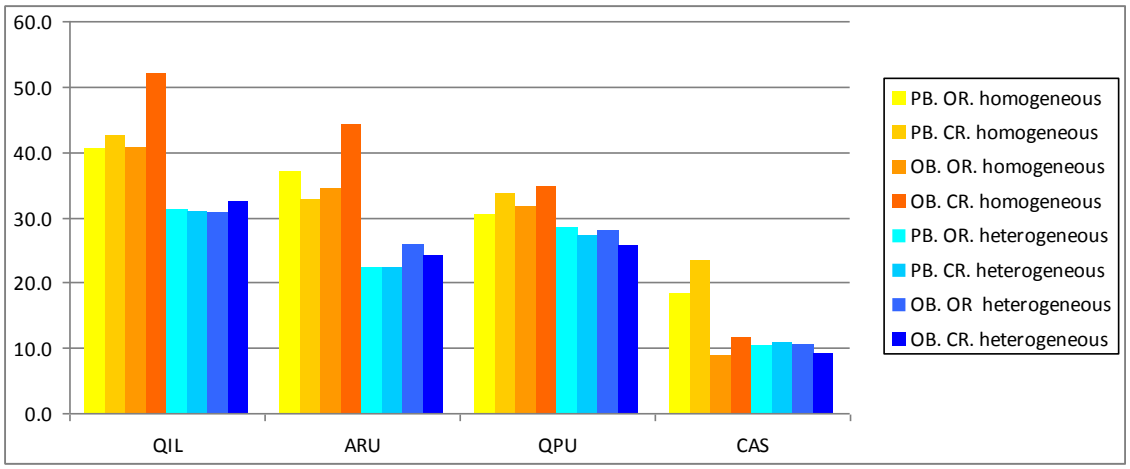

Fig. 5.3: comparison of heterogeneous and homogeneous RMSEs between the predicted species abundance per plot and the field observations per plot, for the four species predicted by all image analysis approaches. (PB=pixel-based, OB=object-based, OR=original spectra, CR=continuum removed spectra)

5.3.2 *Dominant species confusion matrices*

Table 5.7a to 5.7h display confusion matrices of the species that dominate the canopy in the geographic locations of the plots, comparing the results from the eight image analysis approaches and the field data. The most interesting question of this analysis is whether the eight image analysis approaches applied in this research were able to correctly predict dominant species.

The most important trend found is that the heterogeneous image analysis approaches perform better in predicting dominant species according to the overall accuracy measures than the homogeneous image analysis approaches: the average homogeneous overall accuracy is 0.43 whereas the average heterogeneous overall accuracy is 0.62. Comparing all image analysis approaches, the object-based continuum-removed heterogeneous approach has the highest overall accuracy with 0.65. Moreover, its user's and producer's accuracies are also generally the largest. However, it did not correctly detect any of the dominant *Erica arborea*. The pixel-based, continuum removed, heterogeneous image analysis approach is the only one that did correctly find *Erica arborea*. This approach on the other hand did not detect any of the *Castanea sativa* plots, which the former as well as three out of the four homogeneous approaches found.

Besides the better results of the heterogeneous approaches compared to the homogeneous approaches, it is interesting to investigate some trends that can be observed from the confusion matrices. Only the matrix results from the heterogeneous approaches are considered because these produced the best results (Table 5.7b; 5.7d; 5.7f; 5.7h). Typically, the most dominant species is *Quercus ilex:* this species is dominant in the majority of the plots according to the image analyses (Table 5.5), even though on

the ground the dominant species is regularly *Arbutus unedo, Quercus pubescens* or *Erica arborea* (Table 5.3). This shows that the dominance of *Quercus ilex* is overestimated in the analysis. Furthermore, the only situation in which *Quercus ilex* was regularly erroneously predicted, is when it is mistaken for *Arbutus unedo*. This is probably caused by the fact that *Quercus ilex* and *Arbutus unedo* often occur in mixed compositions in the field, making it difficult to distinguish between them.

	Image analysis prediction				
Field data	**QIL**	**ARU**	**QPU**	**CAS**	*producers's accuracy*
QIL	44	38	13	4	0.44
ARU	7	13	2	3	0.52
QPU	3	9	9	1	0.41
CAS	0	1	0	1	0.50
user's accuracy	0.81	0.21	0.38	0.11	0.45

Table 5.7a: confusion matrix of the dominant species per plot of the pixel based original reflectance, homogeneous, image analysis approach*

	Image analysis prediction						
Field data	**QIL**	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**	*producer's accuracy*
QIL	88	11	0	0	0	0	0.89
ARU	18	6	0	0	1	0	0.24
QPU	17	0	5	0	0	0	0.23
CAS	2	0	0	0	0	0	0.00
ERA	10	0	0	0	0	0	0.00
PHL	0	0	0	0	0	0	-
user's accuracy	0.65	0.35	1.00	-	0.00	-	0.63

Table 5.7b: confusion matrix of the dominant species per plot of the pixel based, original reflectance, heterogeneous image analysis approach*

	Image analysis prediction				
Field data	**QIL**	**ARU**	**QPU**	**CAS**	*producer's accuracy*
QIL	38	39	17	5	0.38
ARU	2	11	5	5	0.48
QPU	2	6	9	5	0.41
CAS	0	1	1	0	0.00
user's accuracy	0.90	0.19	0.28	0.00	0.40

Table 5.7c: confusion matrix of the dominant species per plot of the pixel based, continuum removed, homogeneous image analysis approach*†

	Image analysis prediction						
Field data	**QIL**	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**	*producer's accuracy*
QIL	80	12	5	0	2	0	0.81
ARU	16	6	1	0	0	0	0.26
QPU	13	0	9	0	0	0	0.41
CAS	1	1	0	0	0	0	0.00
ERA	7	1	0	0	2	0	0.20
PHL	0	0	0	0	0	0	-
user's accuracy	0.68	0.30	0.60	-	0.50	-	0.62

Table 5.7d: confusion matrix of the dominant species per plot of the pixel based, continuum removed, heterogeneous image analysis approach*†

	Image analysis prediction				
Field data	**QIL**	**ARU**	**QPU**	**CAS**	*producer's accuracy*
QIL	47	33	18	1	0.47
ARU	1	20	4	0	0.80
QPU	3	1	18	0	0.82
CAS	1	0	0	1	0.50
user's accuracy	0.90	0.37	0.45	0.50	0.58

Table 5.7e: confusion matrix of the dominant species per plot of the object based, original reflectance, homogeneous image analysis approach*

	Image analysis prediction						
Field data	**QIL**	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**	*producer's accuracy*
QIL	83	16	0	0	0	0	0.84
ARU	22	3	0	0	0	0	0.12
QPU	17	2	3	0	0	0	0.14
CAS	0	2	0	0	0	0	0.00
ERA	10	0	0	0	0	0	0.00
PHL	0	0	0	0	0	0	-
user's accuracy	0.63	0.13	1.00	-	-	-	0.56

Table 5.7f: confusion matrix of the dominant species per plot of the object based, original reflectance, heterogeneous image analysis approach*

	Image analysis prediction				
Field data	**QIL**	**ARU**	**QPU**	**CAS**	*producer's accuracy*
QIL	5	50	43	1	0.05
ARU	0	20	5	0	0.80
QPU	0	7	13	1	0.62
CAS	0	1	0	1	0.50
user's accuracy	1.00	0.26	0.21	0.33	0.27

Table 5.7g: confusion matrix of the dominant species per plot of the object based, continuum removed, homogeneous image analysis approach*†

	Image analysis prediction						
Field data	**QIL**	**ARU**	**QPU**	**CAS**	**ERA**	**PHL**	*producer's accuracy*
QIL	78	17	4	0	0	0	0.79
ARU	17	8	0	0	0	0	0.32
QPU	9	1	15	0	0	0	0.71
CAS	0	0	0	1	0	0	0.50
ERA	5	1	0	0	0	0	0.00
PHL	1	0	0	0	0	0	-
user's accuracy	0.71	0.30	0.79	1.00	-	-	0.65

Table 5.7h: confusion matrix of the dominant species per plot of the object based, continuum removed, heterogeneous image analysis approach*†

* The total number of field data (i.e. plots) is less than 169 because various plots were not dominated by a single species and therefore left out of the analysis. † The application of continuum removal resulted in erroneous results at some plot locations. Therefore the pixel-based continuum removed image analysis misses data for two plots in the field dominated by ARU, and object-based continuum removed image analysis misses data for one plot in the field dominated by QPU.

5.3.3 *Correlations between RMSE and vegetation parameters*

To check the effects of four vegetation parameters on the classification accuracy, correlation coefficients were calculated between RMSE and the parameter values (Table 5.8). The parameters considered were the average height, the aboveground biomass, the vegetation cover and the heterogeneity. These coefficients show how substantial the relation between these vegetation characteristics and the error of the image analysis is. This indicates their effect on the analysis or the potential bias of the results.

The average height has negligible correlation coefficient values for the homogeneous image analysis approaches (-0.19 to 0.12), but small positive correlations for the heterogeneous image analysis approaches (0.38 to 0.63) (Table 5.8). The same applies to the aboveground biomass which has negligible correlation with the RMSEs of the homogeneous image analysis approaches (0.04 to 0.11), but does have a small positive correlation with the RMSEs of the heterogeneous image analysis approaches (0.35 – 0.37). These correlation coefficients are too small however to indicate substantial effects of the vegetation characteristics on the classification accuracy. The vegetation cover has negligible correlation coefficients for all image analysis approaches (-0.14 to 0.05). This means that there is no relation between the RMSE and the vegetation cover as measured in the field. The heterogeneity shows no correlation for the homogeneous image analysis approaches (-0.19 to 0.06). For the heterogeneous image analysis approaches there is a substantial negative correlation (-0.59 to -0.50). This means that when more species are present in a plot, the respective RMSE is lower in case of the heterogeneous approaches.

CORRELATIONS	*Average height*	*Aboveground biomass*	*Vegetation cover*	*Heterogeneity (nr of species)*
pixel-based, original reflectance, homogeneous	0.12	0.09	-0.06	0.06
pixel-based, original reflectance, heterogeneous	0.51	0.36	-0.14	-0.51
pixel-based, continuum removed, homogeneous	0.05	0.11	0.00	-0.01
pixel-based, continuum removed, heterogeneous	0.43	0.37	-0.12	-0.51
object-based, original reflectance, homogeneous	0.01	0.11	0.03	0.10
object-based, original reflectance, heterogeneous	0.63	0.35	-0.12	-0.50
object-based, continuum removed, homogeneous	-0.19	0.04	0.05	-0.19
object-based, continuum removed, heterogeneous	0.38	0.36	-0.09	-0.59

Table 5.8: correlation coefficients between the RMSEs of the various image analysis approaches, and four vegetation characteristics: average height, aboveground biomass, vegetation cover and heterogeneity

The negative correlations between heterogeneity and RMSE for the heterogeneous approaches indicate bias of the image analysis methodology in favour of areas with heterogeneous vegetation composition: the RMSE is lower in areas with more species which means that the methodology works better for heterogeneous areas. To illustrate this, the groundtruth values of the four most abundant species are plotted

against the predicted values of the object-based, continuum removed, heterogeneous image analysis approach (Fig. 5.4). This graph shows a consequence of the bias in favour of heterogeneous vegetation composition: species with low abundance (0%-20%) are overestimated, whereas species with high abundance (80%-100%) are underestimated. This results from the bias in favour of heterogeneous vegetation composition, because heterogeneous vegetation composition usually has average species abundances (e.g. 40%-60%) whereas homogeneous vegetation composition usually has very low or very high species abundance (e.g. 0%-10% or 90%-100%). The image analysis most often produced heterogeneous vegetation compositions with average species abundances. Homogeneous vegetation compositions with very low or very high species abundances were produced much less. Thus, the more extreme values of groundtruth species abundance are generally over- or underestimated.

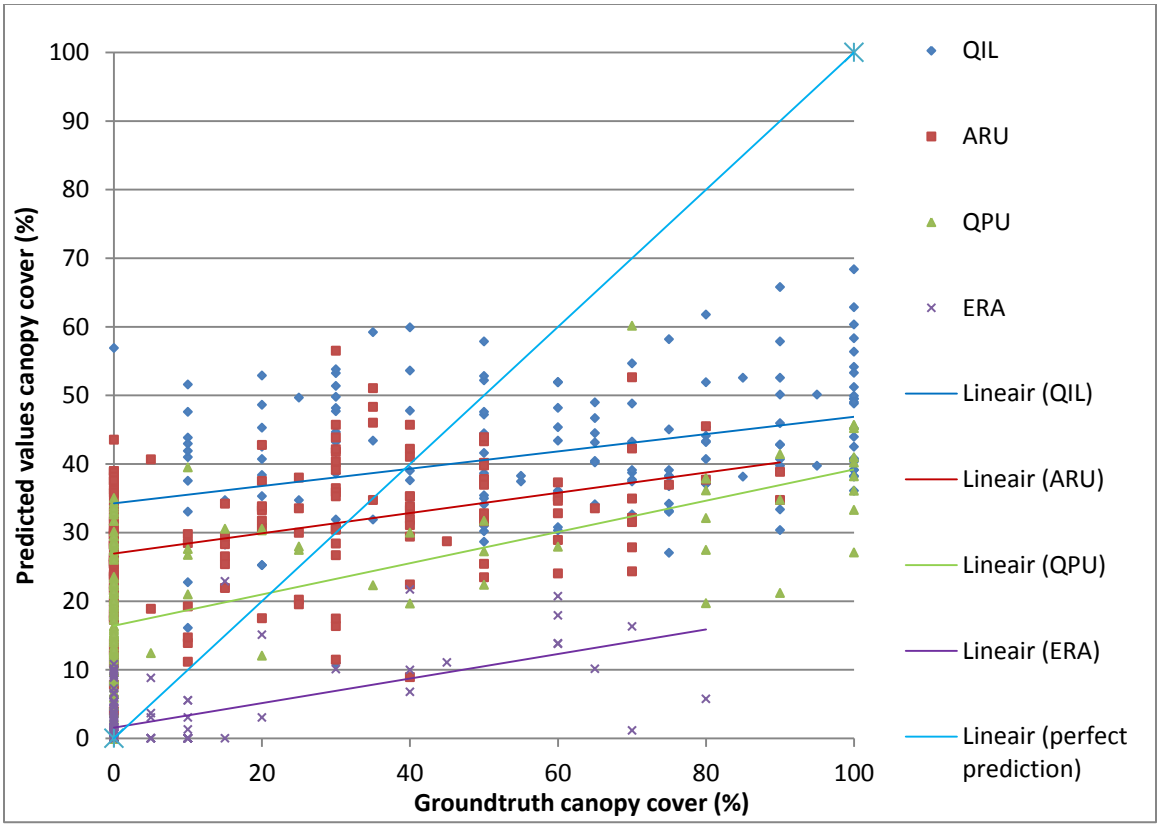

Fig. 5.4: relationship between groundtruth canopy cover values and predicted canopy cover values of four species for the object-based, continuum removed, heterogeneous image analysis approach. The perfect prediction is added for comparison.

Next to the over- and underestimation of the image analysis methodology, figure 5.4 illustrates the limited prediction range: the maximum canopy cover fraction that is predicted is 68% for *Quercus ilex*, 60% for *Quercus pubescens,* 56% for *Arbutus unedo,* and 22% for *Erica arborea.* However, in the field *Quercus ilex* and *Quercus pubescens*

regularly make up 100% of the canopy. *Arbutus unedo* can make up 90% and *Erica arborea* up to 70%. This means that the range of predictable values of the image analysis methodology is substantially smaller than the range of values encountered in the field.

6. Discussion

In this research, Mediterranean vegetation species were mapped by using airborne, hyperspectral HyMap images and homogeneous and heterogeneous endmember analysis in order to determine the best species mapping approach. In addition, it was investigated whether pixel- or object-based image analysis improved the results. Lastly, the advantage or disadvantage of using continuum removed reflectance spectra over original reflectance spectra was evaluated. In this chapter, the entire research process is reflected upon. The various steps are discussed in detail, in the following order: field data collection, image analysis and accuracy assessment. In addition to the discussion, recommendations for follow-up research are provided.

6.1 Field data collection

The field data collection process was determined by a set of requirements: 1) the field plots should represent the selected study area in terms of vegetation characteristics and composition, 2) a random selection factor was required to prevent statistical bias and 3) the plots had to be reasonably accessible. The first two requirements were difficult to combine. The random selection factor as described in section 4.1.1 was nearly always applied when logistically possible. However, this means that dominant species were sampled most often – i.e. *Quercus ilex* and *Arbutus unedo*. Other species, most importantly *Castanea sativa*, *Erica arborea* and *Phillyrea latifolia* were relatively underrepresented in the dataset. Especially as the analysis approach of this research also used the field plots to create spectral libraries for linear spectral unmixing, the limited availability of study plots with the latter three species made their endmembers less reliable. This problem did not necessarily oppose the requirement of representativeness as these are indeed less common. However, training and validation became more problematic and therefore it was more difficult to fulfil the research goals.

Besides the issue of over- and undersampling, in the field some more practical concerns arose. In this study's approach the percentages of species that made up the canopy composition were the most important parameter. However, relatively little time was spent on determining the composition as this was performed by visual estimation. Throughout the field campaign, the same methodology was used to perform the estimations. For this reason, the dataset is consistent and in that sense reliable. Nonetheless, it might have been possible to make more reliable measurements by developing a more automated and quantitative measuring system. Next to that, a portable GPS device was used to note the geometric locations of the field plots. These measurements were often taken below the canopy, which deteriorated reception and consequently reliability. The uncertainty was generally five meters for the

measurements. Especially in combination with additional geometric errors in the image analysis and accuracy assessment this might have increased uncertainty in the analysis, most notably in the pixel-based approaches.

When follow-up research is conducted, or studies with similar field campaigns, it will be important to develop a more automated and quantitative canopy composition measuring systems. For example, a grid with cells of one square meter could be laid out on the surface, and for each cell the overhead vegetation cover could be registered. This would lead to more accurate measurements. On the other hand, this method would cost substantial extra time, so the advantages and disadvantages of adopting such a method have to be carefully considered. Second, the importance of geometric localization should not be underestimated. If similar research is conducted, the use of a more advanced and more accurate GPS device is recommended, for example a DGPS system, in order to minimize geometric error.

6.2 Image analysis

On the basis of field observations, eight image analysis approaches were applied to the hyperspectral imagery to produce eight sets of species abundance maps. In three steps, the following image analysis methods were alternately applied (Fig 4.3): 1) pixel- or object-based image analysis, 2) image analysis with original or continuum removed reflectance spectra, and 3) linear spectral unmixing with homogeneous or heterogeneous endmembers.

In the literature, object-based image analysis is used next to the more traditional pixel-based image analysis and is recognized to improve results in some cases (Blaschke, 2010; Addink, 2007). Also continuum removal is regularly applied and especially in mineral detection is known to improve results (Kruse, 1993). Both methods are popular in remote sensing studies, but whether their application leads to improved results in vegetation studies is debated. In this research neither the pixel- nor the object-based approaches, nor the image analysis with original or continuum removed spectra produced substantially better or worse results. Nonetheless, by continuing the application of the aforementioned methods in future remote sensing studies and comparing the results, effects of the methodologies will become more clear.

The main goal of image analysis was to map species composition in the Peyne catchment area. For this purpose, spectral unmixing on the basis of endmembers was applied to predict the abundance of the species considered. A wide range of spectral unmixing algorithms is available. The most important distinction regards linear and non-linear spectral unmixing (Kesheva, 2003). Linear spectral unmixing assumes a linear relation between the relative surface area of a constituent and its addition to the reflectance spectrum. Non-linear spectral unmixing assumes a more complex situation

where the constituents of a reflecting surface are distributed randomly and the various surfaces do not contribute equally to the overall reflectance. In this study, only linear spectral unmixing is applied as this directly relates to the canopy composition measurements performed in the field. Many algorithms that could enhance results are available, some of which have been applied in other vegetation studies: e.g. Match filtering method by Sobhan (2007) and multiple endmember spectral mixture analysis by Youngentob et al. (2011). This would have required much additional modelling and computational efforts shifting the focus of the research. Therefore, it was decided to use standard linear spectral unmixing, and to concentrate on the effects of other image analysis aspects. Nonetheless, the application of more advanced unmixing algorithms might improve results. For this reason, it is an interesting topic for follow-up studies.

Linear spectral unmixing is based on endmembers. Therefore, the selection of endmember classes is an important step in image analysis. This is also reflected by studies that focus solely on the endmember selection algorithms that will produce the best results: e.g. Tompkins et al. (1997) state that 'it is crucial to carefully choose the endmembers to be able to perform spectral mixing analysis with physically meaningful results'. In this research, no advanced algorithms were applied to select the best endmembers. First, this required too extensive computational efforts which would shift the focus of the study. More importantly, one of the main research objectives considers the difference between homogeneous and heterogeneous endmembers. If the endmembers were selected by means of algorithms, they would not be specifically homogeneous or heterogeneous. Therefore, the endmember selection was performed manually by selecting the field plots most suitable to create endmembers.

The spectral libraries with endmembers were based on a set of training plots manually retrieved from the hyperspectral imagery. The choices made were based on the available training plot compositions, and the most common canopy compositions as known from field observations. In some cases, no perfect solution was possible: for the homogeneous endmembers, sufficient training plots for *Quercus ilex* and *Quercus pubescens* were available. For *Arbutus unedo*, only plots with 70%-90% abundance were present. In order to still be able to detect *Arbutus unedo*, these were considered homogeneous training plots. For *Castanea sativa*, only five training plots were available which reduced linear spectral unmixing reliability. In the case of heterogeneous approaches, four endmembers had five or less training plots which reduced linear spectral unmixing reliability: 100% *Castanea sativa*, 50% *Quercus pubescens* - 50% *Quercus ilex*, 75% *Erica arborea* - 25% *Quercus ilex*, and 75% *Erica arborea* - 25% *Arbutus unedo*. The limitations on endmember training plot selection were partly the result of the stratified random selection scheme which did not allow the selection of plots with less common compositions. In future studies, some additional training plots could be

selected outside the stratified random selection scheme to increase the number of training plots. These plots cannot be used for validation purposes.

Next to these limitations in training plot availability, in the heterogeneous approaches some species were only present in endmembers with a limited percentage (i.e. *Arbutus unedo:* 75%, *Erica arborea*: 75%, *Phillyrea latifolia*: 25%). This limits the analysis as these species could never be detected with abundances larger than the percentages they constitute in the respective endmembers. Moreover, in the homogeneous approach *Erica arborea* and *Phillyrea latifolia* were not present at all which excludes them entirely from the image analysis. In this way, the selection of endmembers influences the image analysis. In the case of the heterogeneous approaches, the composition of the training plots is a good representation of the species canopy compositions encountered in the field, so the limited percentages do not affect the results much. On the other hand, in the case of the homogeneous approaches two species are excluded even though they are regularly present in the canopy. The possibility to detect species that have lower abundance, is an inherent advantage of the heterogeneous approaches compared to the homogeneous approaches.

6.3 Accuracy assessment

The accuracy assessment of this study constitutes three components which will be discussed in the following paragraphs. First, the root mean square errors which compared the predicted species abundance per plot to the groundtruth as known from field observations, to check reliability of the image analysis. Second, the confusion matrices which compared how well the dominant species per plot were detected by the image analysis. Third, correlation coefficients between the average RMSE per plot and four vegetation characteristics, which investigated relations between parameters and possible bias.

From the RMSE accuracy assessment it was found that image analysis on the basis of heterogeneous endmembers produces substantially better results than image analysis on the basis of homogeneous endmembers: of the four species predicted by all approaches (*Quercus ilex, Arbutus unedo, Quercus pubescens* and *Castanea sativa*), average RMSE of the homogeneous approaches is 32% compared to 23% for the heterogeneous approaches. The RMSE is relative to the species abundance values, which means that they differ per species. For example, *Quercus ilex* has relatively high abundance (average per approach: 12% – 48%) and therefore high RMSEs (31% - 52%). *Phillyrea latifolia* however has much lower abundance (average per approach: 3% - 5%) and consequently low RMSEs (6% – 8%). When considering the RMSEs (Table 5.6), it has to be kept in mind that their meaning differs per species in relation to their relative abundance in the canopy (Table 5.4). The total averaged RMSEs per approach shows the

clear advantage of heterogeneous approaches over homogeneous approaches though, which is the most important outcome of the accuracy assessment.

Assessing the accuracy of abundance values is uncommon in vegetation remote sensing studies as most researches use classification approaches (e.g. Clark et al., 2005; Lucas et al., 2008). Therefore, it is difficult to compare the results of this study to other researches. Li et al. (2005) applied spectral mixture analysis to map the fractions of three vegetation species in a coastal salt marsh. They do not report specific error values, but state that geographical trends as observed in the field are similar to those obtained from the image analysis, and that two species are often intermixed because they are spectrally similar. These are similar results to this research, but the number of species studied is three instead of six. Therefore, the research described in this report has achieved similar results with a more challenging approach. Sobhan (2007) performed spectral unmixing of hyperspectral imagery in a Mediterranean environment on the basis of twenty homogeneous endmembers. He investigated if the species present on the ground were correctly detected. The average number of species in the validation plots is three, for which an RMSE of slightly less than one species per validation plot is found (0.73). The accuracies of the respective abundances were not assessed however, which makes a comparison impossible. To conclude, vegetation research on the basis of abundance is uncommon and limited material for comparison is available. In future studies the methods to assess accuracies of species abundance values should be improved.

To obtain an additional measure of prediction accuracy, confusion matrices were applied. As this methodology only works for classification approaches, the matrix analysis was applied to the dominant species of the plots. The main disadvantage of this method is that it does not provide insight in the accuracy of species that typically have low abundance in the canopy – i.e. *Castanea sativa, Erica arborea* and *Phillyrea latifolia*. Indeed these species are hardly ever dominant and therefore their respective user's and producer's accuracy have little meaning for image analysis reliability. For the more dominant species – i.e. *Quercus ilex, Arbutus unedo* and *Quercus pubescens* – the method does provide useful insights. The most important result is that the heterogeneous approaches perform substantially better than the homogeneous approaches: on average respectively 0.62 compared to 0.43. Next to that, it is shown that *Quercus ilex* is regularly erroneously classified to be the dominant species – i.e. it is overestimated. *Quercus ilex* is indeed the most common species in the study area, and the image analysis methodology is biased in favour of this species. In future studies on species mapping, specific attention should be paid to this characteristic in order to prevent the more common species from erroneously dominating image analysis results.

Studies on vegetation species mapping often classify pixels into distinct groups, which can be compared to the estimated dominant species of the validation plots. Based on confusion matrices, Clark et al. (2005) obtain an average accuracy of 0.86 and Lucas et al. (2008) of 0.76. This is substantially better than the 0.62 obtained in the research described in this report. Nonetheless, the two aforementioned studies specifically use homogeneous canopy spectra which makes classification easier than for heterogeneous canopy spectra. Youngentob et al. (2011) studied heterogeneous Eucalypt forests and obtained an accuracy of the dominant species of 0.83. This is better than the accuracies obtained in the research described in this report, but Youngentob et al. (2011) only considered two species groups which is easier to distinguish between than six species.

The final component of the accuracy assessment regarded the investigation of correlations and possible bias of the image analysis methodology. The only coefficients that show substantial correlation resulted from the average RMSE per plot in combination with the level of heterogeneity. The heterogeneous approaches show a distinct negative correlation (-0.51 to -0.59). As is discussed in section 5.3.3, this indicates a serious bias of the image analysis methodology in favour of pixels with heterogeneous vegetation composition. This also results in the overestimation of species with low abundance groundtruth values, the underestimation of species with high abundance groundtruth values, and it limits the predictive range of the image analysis (Fig. 5.4). These characteristics of the image analysis results show a distinct pitfall of the image analysis methodology applied in this study. When future studies plan to apply heterogeneous endmembers in their spectral unmixing analysis, serious attention has to be paid to this issue.

7. Conclusions and recommendations

The main objective of this research was to map six Mediterranean natural vegetation species on the basis of hyperspectral imagery, in the natural forest area of the Peyne catchment, southern France. For this purpose, a new approach was suggested: linear spectral unmixing on the basis of heterogeneous endmembers, as opposed to the conventional approach of homogeneous endmembers. Both approaches were performed, in order to evaluate the difference in results. Next to this, two additional methodologies were tested: image analysis with an object-based approach compared to the traditional pixel-based approach, and image analysis on the basis of continuum removed reflectance data compared to image analysis on the basis of original reflectance data. Finally, an accuracy assessment was performed to check and compare the reliability of the results, and evaluate possible correlations and biases. The final conclusions are as follows:

- Mediterranean vegetation species mapping by means of imaging spectroscopy is possible. However, the uncertainties remain large: RMSEs range from 9% to 52% for the homogeneous approaches (average canopy fractions: 2% to 32%), and from 6% to 31% for the heterogeneous approaches (average canopy fractions: 1% to 48%).

- Linear spectral unmixing on the basis of heterogeneous endmembers produces better mapping results than on the basis of homogeneous endmembers: average RMSE of the four species modelled by all approaches is 32% with homogeneous endmembers and 23% with heterogeneous endmembers. Moreover, the heterogeneous image analysis approaches are capable of mapping six species whereas the homogeneous approaches can map only four species.

- Species mapping using pixel-based image analysis or object-based image analysis does not result in substantially different model performance.

- Species mapping using original reflectance spectra or continuum removed reflectance spectra does not result in substantially different model performance.

- The accuracy assessment showed the differences and similarities between the various analysis approaches. Also, a substantial negative correlation between the RMSE per plot and the level of heterogeneity was found, indicating bias of the image analysis procedure in favour of vegetation with heterogeneous composition.

Future research on natural vegetation species mapping should pay attention to various issues that were encountered during this research. During the field work, practical improvements could be made by improving the measurement methodology of the canopy composition. Visual estimates are reasonably consistent, but a more automated and quantitative method would increase reliability. The accuracy of geometric localization is also important. As canopy cover reduces the reliability of GPS devices, a more advanced instrument is recommended. With regards to the image analysis process, it is recommended to investigate the best spectral unmixing approach. Besides linear spectral unmixing – applied in this research – various other algorithms are available that might improve analysis results (Kesheva, 2003). The same applies to the selection of endmembers. In this research endmembers were selected manually. Many studies have focused on the selection of the perfect endmembers with advanced algorithms, explaining most of the variance (Tompkins, 1997). If the choice of endmembers is based on such procedures, the accuracy of species mapping might be substantially improved. From the accuracy assessment it followed that few studies performed species mapping by detecting abundance, therefore making comparisons of accuracy with other studies difficult. Li et al. (2005) obtained similar results with less species, though clear error tests are not provided. Classification of homogeneous canopies achieve substantially better results: average overall accuracies of 0.86 and 0.76 (Clark et al., 2005; Lucas et al., 2008) as opposed to 0.62 for the research described in this report. Also in heterogeneous environments good results have been obtained (0.83) (Youngentob et al., 2011), but in this case only two species were considered which is easier than six species. In follow-up studies, substantial attention should be paid to accuracy assessment to produce statistical tests that can easily be compared with other studies. Lastly, the correlation tests showed the bias of the model approach in favour of vegetation with heterogeneous compositions. This is a serious disadvantage of this image analysis procedure which has to be kept in mind.

To conclude, spectral unmixing of hyperspectral imagery on the basis of heterogeneous endmembers is an innovative approach in vegetation species mapping. It provides a new step in the research towards detailed, accurate monitoring of vegetation dynamics. With the launch of hyperspectral satellites such as USA's Hyperion (NASA, 2012) and Germany's EnMAP (EnMAP, 2012), the areal extent to which this type of research can be expanded is enormous. Therefore, the future possibilities for vegetation studies by means of imaging spectroscopy are very promising.

References

- Adam, E. and O. Mutanga (2009). Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. ISPRS Journal of Photogrammetry and Remote Sensing 64, 612-620.
- Addink, E.A., S.M. de Jong and E.J. Pebesma (2007). The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. Photogrammatric Engineering and Remote Sensing 73: 905-912.
- Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65: 2-16.
- Bonfils, P. (1993). Carte Pédologicue de la France 1:100 000. Feuille Lodève. INRA, Olivet.
- Christensen, J.H. and B. Hewitson. (2007). Regional Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY, USA: Cambridge University Press
- Clark, M.L., D.A. Roberts and D.B. Clark (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sensing of Environment 96, 375-398.
- Clark, R.N. and T.L. Roush (1984). Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. Journal of Geophysical Research 89: 6329-6340.
- Congalton, R.G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment 37: 35-46.
- Dennison, P.E. and D.A. Roberts (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sensing of Environment 87: 123-135.
- EnMAP. (2012). Environmental Mapping and Analysis Program: http://www.enmap.org/
- Gong, P., R. Pu and B. Yu (1997). Conifer Species Recognition: An Exploratory Analysis of In Situ Hyperspectral Data. Remote Sensing of Environment 62.2, 189-200.
- Horler, D.N.H., M. Dockray and J. Barber. (1983). The red edge of plant leaf reflectance. International Journal of Remote Sensing 4.2: 273-288.
- HyVista. (2012). HyVista Corporation website: http://www.hyvista.com
- Isaaks, E.H. and R.M. Srivastava. (1989) Applied Geostatistics. New York, Oxford: Oxford University Press.
- Joffre, R., and S. Rambal (2001). Mediterranean Ecosystems, in: Encyclopdia of Life Sciences. John Wiley & Sons Ltd, Chichester, UK

- Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M. and F. Baret (2004). Reviews of methods for in situ leaf area index determination. Part I. Theories, sensors, and hemispherical photography. Agriculture and Forest Meteorology 121: 19-35.
- Kesheva, N. (2003). A Survey of Spectral Unmixing Algorithms. Lincoln Laboratory Journal 14.1: 55-78.
- Kruse, F.A. and A.B. Lefkoff (1993). Knowledge-Based Geologic Mapping with Imaging Spectrometers. Remote Sensing Reviews 8: 3-28.
- Lambin, E.F. et al. (2001). The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11: 261-269.
- Lillesand, T.M., R.W. Kiefer and J.W. Chipman (2008). Remote Sensing and Image Interpretation. John Wiley & Sons, USA.
- Lu, D., E. Moran and M. Batistella. (2003) Linear mixture model applied to Amazonian vegetation classification. Remote Sensing of Environment 87: 456-469.
- Lucas, R., P. Bunting, M. Paterson and L. Chisholm (2008). Classification of Australian forest communities using aerial photography, CASI and HyMap data. Remote Sensing of Environment 112, 2088-2103.
- Manevski, K., I. Manakos, G.P. Petropoulos, C. Kalaitzidis (2011). Discrimination of common Mediterranean plant species using field spectroradiametry. International Journal of Applied Earth Observation and Geoinformation 13: 922-933.
- Mather, A.S., J. Fairbairn, C.L. Needle (1999). The course and drivers of the forest transition: The case of France. Journal of Rural Studies 15: 65-90.
- Mc Coy, M. R. (2005). Field methods in remote sensing. New York, USA: The Guilford Press.
- Merzlyak, M.N., A.A. Gitelson, O.B. Chivkunova and V.Y. Rakitin (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum 106: 135-141.
- Miller, P.C. and Hajek, E., (1981). Resource availability and environmental characteristics of Mediterranean-type ecosystems. In: Miller, P.C. (ed). Resource use by chaparral and matorral. Springer-Verlag, New York: 17-41.
- NASA. (2012). Earth Observing-1 General Mission: http://eo1.gsfc.nasa.gov/new/general/firsts/hyperion.html
- Nijland, W., E.A. Addink, S.M. De Jong, F.D. Van der Meer (2009). Optimizing spatial image support for quantitative mapping of natural vegetation. Remote Sensing of Environment 113, 771-780.
- Nijland, W. (2011). Mediterranean evergreen vegetation dynamics. PhD Thesis, Utrecht University, Utrecht.
- Ogaya, R., J. Penuelas, J. Martínez-Vilalta and M. Mangirón (2003). Effect of drought on diameter increment of Quercus Ilex, Phillyrea latifolia, and Arbutus

unedo in a holm oak forest of NE Spain. Forest Ecology and Management 180: 175-184.
- Pu, R. (2009). Broadleaf species recognition with in situ hyperspectral data. International Journal of Remote Sensing 30.11, 2759-2779.
- Richter, R. and D. Schläpfer (2002). Geo-atmospheric processing of airborne imaging spectrometry data Part 2: Atmospheric/topographic correction. International Journal of Remote Sensing 23.48, 2631-2649.
- Ritter, M.E. (2006). The Physical Environment: an Introduction to Physical Geography: http://www.uwsp.edu/geo/faculty/ritter/geog101/textbook/title_page.html
- Roy, P.S. (1989). Spectral reflectance characteristics of vegetation and their use in estimating productive potential. Proceedings Indian Academy of Science 99.1: 59-81.
- Schläpfer, D. and R. Richter (2002). Geo-atmospheric processing of airborne imaging spectrometry data Part 1: parametric orthorectification. International Journal of Remote Sensing 23.13, 2609-2630.
- Schmidt, K.S. and A.K. Skidmore (2001). Exploring spectral discrimination of grass species in African rangelands. International Journal of Remote Sensing 22.17, 3421-3434.
- Schmidt, K.S. and A.K. Skidmore (2003). Spectral discrimination of vegetation types in a coastal wetland. Remote Sensing of Environment 85, 92-108.
- Shoshany, M (2000). Satellite remote sensing of natural Mediterranean vegetation: a review within an ecological context. Progress in Physical Geography 24.2, 153-178.
- Singer, R.B. and T.B. McCord (1979). Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance. Proceedings of the 10th lunar and planetary science conference, USA: 1835-1848.
- Sluiter, R. (2005). Mediterranean land cover change. PhD thesis, Utrecht University, Utrecht.
- Sobhan, I. (2007). Species discrimination from a hyperspectral perspective. PhD thesis, ITC Enschede.
- Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller. (2007); Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY,USA: Cambridge University Press.
- Tomaselli, R. (1981). Main physiognomic types and geographic distribution of shrub systems related to Mediterranean climates. In: F. Di Castri et al (Ed.) Ecosystems of the world II Mediterranean type shrublands (pp. 95-106). Elsevier.
- Tompkins, S., J.F. Mustard, C.M. Pieters, D.W. Forsyth (1997). Optimization of endmembers for spectral mixture analysis. Remote Sensing of Environment 59: 472-489.

- Ustin, S.L., D.A. Roberts, J.A. Gamon, G.P. Asner and R.O. Green (2004). Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. Bioscience 54.6, 523-534.
- Van Aardt, J.A.N. and R.H. Wynne (2001). Spectral Separability among Six Southern Tree Species. Photogrammetric Engineering & Remote Sensing 67.12, 1367-1375.
- Van der Meer, F.D., and S.M. de Jong (2002). Imaging Spectrometry. New York.
- Weiss, M., Baret, F., Smith, G.J. and Jonckheere, I (2004). Methods for in situ leaf area index measurement, part II: from gap fraction to leaf area index: retrieval methods and sampling strategies. Agriculture and Forest Meteorology 121: 17-53.
- Weiss. M., and F. Baret (2010). CAN-EYA V6.1 USER MANUAL. INRA, Avignon
- Youngentob, K.N., D.A. Roberts, A.A. Held, P.E. Dennison, X. Jia, D.B. Lindenmayer (2011). Mapping two Eucalyptus subgenera using multiple endmember spectral mixture analysis and continuum-removed imaging spectrometry data. Remote Sensing of Environment 115: 1115-1128.
- Xie, Y., Z. Sha and M. Yu. (2008). Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology 1.1: 9-23.

Appendices

Appendix A – plot description and canopy composition

	UTM		vegetation cover density	land cover class (Tomaselli 1981)	
code	*X (East)*	*Y (North)*			
A0101	521511	4826453	high	high matorral	dense
A0102	521554	4826426	high	high matorral	dense
A0103	521602	4826441	high	high matorral	dense
A0104	521601	4826394	high	high matorral	dense
A0105	521578	4826440	high	high matorral	dense
A0106	521549	4826400	high	high matorral	dense
A0107	521524	4826354	high	high matorral	discontinuous
A0108	521532	4826399	high	high matorral	dense
A0201	521616	4826064	middle	middle matorral	dense
A0202	521645	4826020	high	middle matorral	dense
A0203	521695	4826025	middle	high matorral	dense
A0204	521749	4826023	high	high matorral	dense
A0205	521774	4825981	high	high matorral	dense
A0206	521723	4825971	high	middle matorral	dense
A0207	521677	4825982	high	middle matorral	dense
A0208	521653	4826033	middle	middle matorral	dense
A0301	521873	4826382	middle	middle matorral	dense
A0302	521831	4826410	high	middle matorral	dense
A0303	521800	4826371	high	middle matorral	dense
A0304	521812	4826419	middle	middle matorral	dense
A0305	521786	4826461	middle	middle matorral	discontinuous
A0306	521808	4826430	high	middle matorral	dense
A0307	521821	4826383	middle	middle matorral	discontinuous
A0308	521774	4826372	high	high matorral	dense
A0401	521056	4826009	high	high matorral	discontinuous
A0402	521101	4826029	high	high matorral	dense
A0403	521095	4825980	high	high matorral	discontinuous
A0404	521121	4826022	high	high matorral	dense
A0405	521129	4826069	high	high matorral	discontinuous
A0406	521084	4826050	high	high matorral	discontinuous
A0407	521068	4826092	high	high matorral	dense
A0408	521108	4826063	high	high matorral	dense
A0501	521152	4826254	high	high matorral	dense
A0502	521129	4826209	high	high matorral	dense
A0503	521116	4826162	middle	high matorral	discontinuous
A0504	521086	4826199	middle	high matorral	discontinuous
A0505	521059	4826156	middle	high matorral	discontinuous
A0506	521017	4826189	high	high matorral	dense
A0507	520965	4826186	high	high matorral	dense
A0508	520962	4826241	middle	high matorral	discontinuous

A1001	521750	4826368	high	high matorral	dense
A1002	521739	4826381	high	high matorral	dense
B0101	522823	4826233	high	high matorral	dense
B0102	522844	4826195	middle	middle matorral	dense
B0103	522802	4826131	high	middle matorral	dense
B0104	522769	4826095	high	middle matorral	dense
B0105	522772	4826146	middle	middle matorral	dense
B0106	522799	4826195	high	middle matorral	dense
B0107	522752	4826187	high	middle matorral	dense
B0108	522771	4826233	high	middle matorral	dense
B0201	524243	4825795	high	high matorral	dense
B0202	524201	4825770	high	middle matorral	dense
B0203	524246	4825791	high	high matorral	dense
B0204	524288	4825753	high	high matorral	dense
B0205	524238	4825742	high	high matorral	dense
B0206	524243	4825695	high	middle matorral	dense
B0207	524194	4825682	high	middle matorral	dense
B0208	524213	4825638	high	middle matorral	dense
B0301	523203	4825977	high	high matorral	dense
B0302	523246	4825953	middle	middle matorral	dense
B0303	523211	4825993	high	high matorral	dense
B0304	523177	4826031	high	high matorral	dense
B0305	523131	4826054	high	high matorral	dense
B0306	523088	4826024	high	planted forest	\
B0307	523130	4825994	high	middle matorral	dense
B0308	523081	4826002	high	middle matorral	dense
B0401	523720	4825820	high	high matorral	dense
B0402	523770	4825830	high	middle matorral	dense
B0403	523738	4825861	high	middle matorral	dense
B0404	523784	4825836	high	middle matorral	dense
B0405	523784	4825787	high	middle matorral	dense
B0406	523740	4825762	high	middle matorral	dense
B0407	523788	4825739	high	middle matorral	dense
B0408	523826	4825709	high	middle matorral	dense
B0501	522966	4826294	high	high matorral	dense
B0502	523016	4826296	high	middle matorral	dense
B0503	523047	4826260	high	middle matorral	dense
B0504	522990	4826268	high	high matorral	dense
B0505	522943	4826291	high	high matorral	dense
B0506	522978	4826327	high	high matorral	dense
B0507	523024	4826339	high	middle matorral	dense
B0508	522992	4826373	high	high matorral	dense
C0101	524954	4823251	high	middle matorral	dense
C0102	524985	4823292	high	middle matorral	dense
C0103	524946	4823327	high	middle matorral	dense
C0104	524994	4823345	high	middle matorral	dense
C0105	525011	4823394	high	middle matorral	dense
C0106	525005	4823445	high	middle matorral	dense

C0107	524964	4823472	high	middle matorral	dense
C0108	525017	4823458	high	middle matorral	dense
C0201	525340	4823423	high	middle matorral	dense
C0202	525290	4823419	high	middle matorral	dense
C0203	525270	4823371	high	middle matorral	dense
C0204	525278	4823421	high	middle matorral	dense
C0205	525228	4823411	high	middle matorral	dense
C0206	525202	4823369	high	middle matorral	dense
C0207	525226	4823324	high	middle matorral	dense
C0208	525262	4823283	high	middle/high matorral	dense
C0301	524803	4823203	high	middle matorral	dense
C0302	524797	4823153	high	middle matorral	dense
C0303	524834	4823182	high	middle matorral	dense
C0304	524830	4823228	high	middle matorral	dense
C0305	524808	4823275	high	high matorral	dense
C0306	524764	4823254	high	middle matorral	dense
C0307	524814	4823263	high	high matorral	dense
C0308	524855	4823289	high	middle matorral	dense
C0401	525421	4823996	high	middle matorral	dense
C0402	525376	4824016	high	middle matorral	dense
C0403	525326	4824025	high	middle matorral	dense
C0404	525300	4823982	high	middle matorral	dense
C0405	525276	4823935	high	middle matorral	dense
C0406	525281	4823885	high	middle matorral	dense
C0407	525234	4823872	high	high matorral	dense
C0408	525276	4823902	high	middle matorral	dense
C0501	524880	4824025	high	middle matorral	dense
C0502	524834	4824004	high	middle matorral	dense
C0503	524837	4823955	high	middle matorral	dense
C0504	524832	4823904	high	high matorral	dense
C0505	524883	4823887	high	middle matorral	dense
C0506	524871	4823937	high	high matorral	dense
C0507	524822	4823925	high	high matorral	dense
C0508	524781	4823892	high	middle matorral	dense
D0101	523466	4823095	middle	middle matorral	discontinuous
D0102	523506	4823070	middle	middle matorral	discontinuous
D0103	523569	4823052	high	middle matorral	dense
D0104	523553	4823103	high	middle matorral	dense
D0105	523586	4823059	high	middle matorral	dense
D0106	523584	4823102	high	middle matorral	dense
D0107	523630	4823110	high	middle matorral	dense
D0108	523591	4823135	high	middle matorral	dense
D0201	523819	4823186	high	middle matorral	dense
D0202	523782	4823154	high	middle matorral	dense
D0203	523736	4823133	high	middle matorral	dense
D0204	523761	4823169	high	middle matorral	dense
D0205	523720	4823200	middle	middle matorral	dense

D0206	523706	4823154	high	middle matorral	dense
D0207	523663	4823139	high	high matorral	dense
D0208	523691	4823097	high	middle matorral	discontinuous
D0301	523427	4823410	high	middle matorral	dense
D0302	523429	4823460	high	middle matorral	dense
D0303	523390	4823500	high	middle matorral	dense
D0304	523436	4823488	high	middle matorral	dense
D0305	523482	4823476	high	middle matorral	dense
D0306	523433	4823497	high	middle matorral	dense
D0307	523432	4823548	high	middle matorral	dense
D0308	523397	4823583	high	middle matorral	dense
D0401	523532	4823793	high	middle matorral	dense
D0402	523572	4823822	high	middle matorral	dense
D0403	523521	4823826	high	middle matorral	dense
D0404	523472	4823836	high	middle matorral	dense
D0405	523496	4823792	high	middle matorral	dense
D0406	523448	4823793	high	middle matorral	dense
D0407	523432	4823841	high	middle matorral	dense
D0408	523381	4823844	high	middle matorral	dense
D0501	523451	4823650	high	middle matorral	dense
D0502	523410	4823685	high	middle matorral	dense
D0503	523386	4823726	high	middle matorral	dense
D0504	523344	4823753	high	middle matorral	dense
D0505	523386	4823781	high	middle matorral	dense
D0506	523431	4823805	high	middle matorral	dense
D0507	523449	4823759	high	high matorral	dense
D0508	523406	4823738	high	high matorral	dense
D1001	523837	4823072	middle	middle matorral	discontinuous
D1002	523850	4823050	middle	middle matorral	discontinuous
D1003	523824	4823117	high	middle matorral	dense
D1004	523866	4823139	high	middle matorral	dense
D1005	523856	4823170	high	middle matorral	dense
D1006	523549	4823289	high	high matorral	dense
D1007	523431	4823041	high	middle matorral	discontinuous

	OUT-CANOPY	**NON-**	**CANOPY**	**species abundances in the canopy (%)**						
code	**vegetation (%)**	**vegetation (%)**	**%**	*QIL*	*ARU*	*QPU*	*CAS*	*ERA*	*PHL*	*other*
A0101			100	100	0	0	0	0	0	
A0102			100	100	0	0	0	0	0	
A0103			100	90	0	10	0	0	0	
A0104			100	50	0	50	0	0	0	
A0105			100	100	0	0	0	0	0	
A0106			100	100	0	0	0	0	0	
A0107	20 - shrubs		80	65	0	35	0	0	0	
A0108			100	100	0	0	0	0	0	
A0201			100	70	15	0	0	0	15	

A0202			100	50	20	0	0	0	0	30
A0203			100	80	20	0	0	0	0	
A0204			100	20	0	80	0	0	0	
A0205			100	20	0	80	0	0	0	
A0206			100	50	0	40	0	0	10	
A0207			100	50	30	0	0	0	0	20
A0208			100	90	0	0	0	0	10	
A0301			100	95	0	5	0	0	0	
A0302			100	90	0	10	0	0	0	
A0303			100	75	25	0	0	0	0	
A0304			100	90	10	0	0	0	0	
A0305			100	70	30	0	0	0	0	
A0306			100	95	5	0	0	0	0	
A0307	10 - shrubs	20 - litter	70	100	0	0	0	0	0	
A0308			100	0	0	100	0	0	0	
A0401			100	30	0	70	0	0	0	
A0402			100	50	0	50	0	0	0	
A0403			100	0	0	100	0	0	0	
A0404			100	10	0	90	0	0	0	
A0405			100	20	0	80	0	0	0	
A0406			100	0	0	100	0	0	0	
A0407			100	10	0	90	0	0	0	
A0408			100	20	0	80	0	0	0	
A0501			100	20	0	80	0	0	0	
A0502			100	20	10	70	0	0	0	
A0503			100	0	0	100	0	0	0	
A0504			100	0	0	100	0	0	0	
A0505			100	0	0	100	0	0	0	
A0506			100	0	0	100	0	0	0	
A0507			100	90	0	10	0	0	0	
A0508			100	0	0	100	0	0	0	
A1001			100	0	0	100	0	0	0	
A1002			100	10	0	90	0	0	0	
B0101			100	100	0	0	0	0	0	
B0102			100	75	0	0	0	0	20	5
B0103			100	75	0	0	0	0	25	
B0104			100	15	50	25	0	0	10	
B0105			100	60	15	15	0	0	10	
B0106			100	40	10	40	0	0	10	
B0107			100	60	0	20	0	0	20	
B0108			100	90	0	10	0	0	0	
B0201			100	100	0	0	0	0	0	
B0202			100	100	0	0	0	0	0	
B0203			100	100	0	0	0	0	0	
B0204			100	100	0	0	0	0	0	
B0205			100	100	0	0	0	0	0	
B0206			100	100	0	0	0	0	0	
B0207			100	100	0	0	0	0	0	

B0208		100	100	0	0	0	0	0	
B0301		100	100	0	0	0	0	0	
B0302		100	10	0	0	90	0	0	
B0303		100	60	40	0	0	0	0	
B0304		100	60	15	0	10	0	0	15
B0305		100	75	25	0	0	0	0	
B0306		100	0	0	0	100	0	0	
B0307		100	20	30	50	0	0	0	
B0308		100	25	70	0	0	0	5	
B0401		100	100	0	0	0	0	0	
B0402		100	90	0	0	0	0	0	10
B0403		100	100	0	0	0	0	0	
B0404		100	85	0	0	0	0	0	15
B0405		100	100	0	0	0	0	0	
B0406		100	80	0	0	0	0	20	
B0407		100	80	0	0	0	0	20	
B0408		100	100	0	0	0	0	0	
B0501		100	20	0	80	0	0	0	
B0502		100	85	15	0	0	0	0	
B0503		100	80	0	20	0	0	0	
B0504		100	100	0	0	0	0	0	
B0505		100	100	0	0	0	0	0	
B0506		100	40	0	60	0	0	0	
B0507		100	75	0	25	0	0	0	
B0508		100	80	20	0	0	0	0	
C0101		100	100	0	0	0	0	0	
C0102		100	50	50	0	0	0	0	
C0103		100	70	30	0	0	0	0	
C0104		100	40	60	0	0	0	0	
C0105		100	30	65	0	0	5	0	
C0106		100	30	60	0	0	0	10	
C0107		100	55	30	0	0	5	5	5
C0108		100	30	50	0	0	10	0	10
C0201		100	30	70	0	0	0	0	
C0202		100	30	60	0	0	10	0	
C0203		100	60	30	0	0	10	0	
C0204		100	35	60	0	0	5	0	
C0205		100	50	50	0	0	0	0	
C0206		100	60	40	0	0	0	0	
C0207		100	35	60	0	0	5	0	
C0208		100	10	90	0	0	0	0	
C0301		100	50	50	0	0	0	0	
C0302		100	30	70	0	0	0	0	
C0303	10 - rocks	90	30	30	0	0	40	0	
C0304		100	60	30	0	0	10	0	
C0305		100	70	30	0	0	0	0	
C0306		100	75	25	0	0	0	0	
C0307		100	100	0	0	0	0	0	

C0308			100	100	0	0	0	0	0	
C0401			100	65	35	0	0	0	0	
C0402			100	25	75	0	0	0	0	
C0403			100	60	40	0	0	0	0	
C0404			100	50	30	20	0	0	0	
C0405			100	80	20	0	0	0	0	
C0406			100	80	20	0	0	0	0	
C0407			100	60	40	0	0	0	0	
C0408			100	70	30	0	0	0	0	
C0501			100	90	10	0	0	0	0	
C0502			100	55	45	0	0	0	0	
C0503			100	80	10	0	0	5	0	5
C0504			100	50	30	0	0	0	10	10
C0505			100	50	50	0	0	0	0	
C0506			100	30	70	0	0	0	0	
C0507			100	70	15	0	0	0	15	
C0508			100	50	40	0	0	0	5	5
D0101	10 - shrubs	5 - litter	85	40	0	0	0	60	0	
D0102	5 - shrubs	15 - litter	80	75	10	0	0	15	0	
D0103			100	65	30	0	0	5	0	
D0104			100	90	0	0	0	5	0	5
D0105			100	50	40	0	0	10	0	
D0106			100	40	40	0	0	20	0	
D0107			100	10	20	0	0	70	0	
D0108			100	75	5	0	0	0	20	
D0201			100	30	40	0	0	30	0	
D0202			100	90	0	0	0	10	0	
D0203			100	90	0	0	0	10	0	
D0204			100	35	0	0	0	65	0	
D0205			100	40	40	0	0	20	0	
D0206			100	50	0	0	0	45	0	5
D0207			100	75	25	0	0	0	0	
D0208	30 - shrubs		70	30	15	0	0	40	0	15
D0301			100	60	40	0	0	0	0	
D0302			100	10	80	0	0	10	0	
D0303			100	20	70	0	0	10	0	
D0304			100	50	50	0	0	0	0	
D0305			100	65	35	0	0	0	0	
D0306			100	70	30	0	0	0	0	
D0307			100	40	60	0	0	0	0	
D0308			100	30	70	0	0	0	0	
D0401			100	75	25	0	0	0	0	
D0402			100	65	35	0	0	0	0	
D0403			100	40	50	0	0	0	10	
D0404			100	80	20	0	0	0	0	
D0405			100	50	50	0	0	0	0	
D0406			100	70	30	0	0	0	0	
D0407			100	80	0	0	0	0	20	

D0408	100	65	30	0	0	0	5	
D0501	100	10	80	0	0	10	0	
D0502	100	15	70	0	0	15	0	
D0503	100	20	60	0	0	10	0	10
D0504	100	50	40	0	0	10	0	
D0505	100	60	30	0	0	10	0	
D0506	100	60	40	0	0	0	0	
D0507	100	60	35	0	0	0	5	
D0508	100	10	90	0	0	0	0	
D1001	100	20	0	0	0	80	0	
D1002	100	30	10	0	0	60	0	
D1003	100	30	0	0	0	70	0	
D1004	100	20	20	0	0	60	0	
D1005	100	10	50	0	0	40	0	
D1006	100	65	0	0	0	0	35	
D1007	100	10	30	0	0	60	0	

Appendix B – training plot data

homogeneous training plots	species percentage canopy cover					
100% QIL	QIL	ARU	QPU	CAS	ERA	PHL
A0101	100%					
A0102	100%					
A0106	100%					
A0307	100%					
B0101	100%					
B0202	100%					
B0204	100%					
B0206	100%					
B0207	100%					
B0301	100%					
B0401	100%					
B0405	100%					
B0504	100%					
C0101	100%					
C0308	100%					
100% ARU						
B0308	25%	70%				5%
C0201	30%	70%				
C0208	10%	90%				
C0302	30%	70%				
C0402	25%	75%				
C0506	25%	75%				
D0308	25%	75%				
D0508	10%	90%				
100% QPU						
A0308			100%			
A0403			100%			
A0404	10%		90%			
A0406			100%			
A0407	10%		90%			
A0503			100%			
A0504			100%			
A0505			100%			
A0506			100%			
A0508			100%			
A1001			100%			
A1002	10%		90%			
100% CAS						
B0302	10%			90%		
B0306				100%		

heterogeneous training plots	species percentage canopy cover					
100% QIL	QIL	ARU	QPU	CAS	ERA	PHL
A0101	100%					
A0102	100%					
A0106	100%					
A0307	100%					
B0101	100%					
B0202	100%					
B0204	100%					
B0206	100%					
B0207	100%					
B0301	100%					
B0401	100%					
B0405	100%					
B0504	100%					
C0101	100%					
C0308	100%					
75% QIL 25% ARU						
A0303	75%	25%				
A0305	70%	30%				
B0305	75%	25%				
C0103	70%	30%				
C0305	70%	30%				
C0306	75%	25%				
C0408	70%	30%				
D0207	75%	25%				
D0306	70%	30%				
D0401	75%	25%				
D0406	70%	30%				
50% QIL 50% ARU						
C0102	50%	50%				
C0205	50%	50%				
C0301	50%	50%				
C0502	55%	45%				
C0505	50%	50%				
D0304	50%	50%				
D0405	50%	50%				
75% ARU 25% QIL						
B0308	25%	70%				5%
C0201	30%	70%				
C0302	30%	70%				
C0402	25%	75%				

CAS0501*	100%	C0506	25%	75%			
CAS1001*	100%	D0308	25%	75%			
CAS1002*	100%						
		50% QIL **50% QPU**					
		A0104	50%		50		
		A0402	50%		50		
		B0506	40%		60		
		100% QPU					
		A0308			100%		
		A0403			100%		
		A0404	10%		90%		
		A0406			100%		
		A0407	10%		90%		
		A0503			100%		
		A0504			100%		
		A0505			100%		
		A0506			100%		
		A0508			100%		
		A1001			100%		
		A1002	10%		90%		
		75% ERA **25% QIL**					
		D0204	35%			65%	
		D1001	20%			80%	
		D1003	30%			70%	
		75% ERA **25% ARU**					
		D0107	10%	20%		70%	
		D1007	10%	30%		60%	
		75% QIL **25% PHL**					
		B0102	75%				20%
		B0103	75%				25%
		B0406	80%				20%
		B0407	80%				20%
		D0108	75%	5%			20%
		D0407	80%				20%
		100% CAS					
		B0302	10%			90%	
		B0306				100%	
*plots located outside study area		CAS0501*				100%	
		CAS1001*				100%	
		CAS1002 *				100%	

Appendix C – physical vegetation characteristics per plot

code	height (m)	vegetation cover	biomass (ton/ha)	heterogeneity (nr of species per plot)
A0101	6.0	80	230	1
A0102	7.0	89	202	1
A0103	8.0	86	170	2
A0104	12.0	86	325	2
A0105	8.0	89	150	1
A0106	6.0	69	171	1
A0107	4.0	74	51	2
A0108	6.0	70	131	1
A0201	6.0	68	136	3
A0202	5.0	82	108	2
A0203	9.0	83	332	2
A0204	12.0	67	392	2
A0205	17.0	85	102	2
A0206	6.0	89	47	3
A0207	4.0	87	75	2
A0208	6.0	85	84	2
A0301	4.5	62	177	2
A0302	6.0	88	151	2
A0303	6.0	57	89	2
A0304	6.0	77	161	2
A0305	6.0	82	101	2
A0306	6.0	54	113	2
A0307	5.0	76	34	1
A0308	12.0	81	2	1
A0401	12.0	73	219	2
A0402	9.0	72	35	2
A0403	12.0	85	267	1
A0404	11.0	79	96	2
A0405	12.0	86	181	2
A0406	15.0	72	384	1
A0407	14.0	81	140	2
A0408	12.0	87	214	2
A0501	14.0	70	136	2
A0502	12.0	71	89	3
A0503	10.0	67	49	1
A0504	11.0	61	228	1
A0505	12.0	62	207	1
A0506	13.0	72	175	1
A0507	13.0	90	228	2
A0508	14.0	73	111	1
A1001	13.0	54	186	1
A1002	15.0	64	399	2
B0101	7.5	88	176	1
B0102	4.0	76	133	2

B0103	5.0	85	106	2
B0104	6.0	89	131	4
B0105	6.0	86	134	4
B0106	6.0	82	103	4
B0107	6.0	83	161	3
B0108	7.0	87	155	2
B0201	8.5	75	148	1
B0202	5.0	79	95	1
B0203	10.0	70	154	1
B0204	10.0	80	287	1
B0205	9.0	75	180	1
B0206	4.0	83	79	1
B0207	6.0	93	145	1
B0208	6.0	71	171	1
B0301	9.0	67	171	1
B0302	12.0	78	184	2
B0303	7.0	68	137	2
B0304	8.0	65	75	3
B0305	7.0	69	95	2
B0306	17.0	82	1084	1
B0307	7.0	65	146	3
B0308	5.0	72	67	3
B0401	7.0	77	185	1
B0402	5.0	66	112	1
B0403	6.0	69	76	1
B0404	5.0	71	136	1
B0405	5.0	68	106	1
B0406	5.0	77	120	2
B0407	5.0	79	105	2
B0408	5.0	72	118	1
B0501	12.0	68	221	2
B0502	7.0	75	131	2
B0503	7.0	70	150	2
B0504	8.0	62	114	1
B0505	8.0	79	239	1
B0506	11.0	79	156	2
B0507	7.0	67	155	2
B0508	12.0	71	207	2
C0101	7.5	71	318	1
C0102	4.0	83	83	2
C0103	4.0	80	189	2
C0104	6.0	74	94	2
C0105	6.0	84	95	3
C0106	6.0	87	105	3
C0107	5.5	85	50	4
C0108	6.0	81	73	3
C0201	3.5	78	63	2
C0202	3.5	78	83	3

C0203	5.0	76	67	3
C0204	6.0	68	122	3
C0205	4.0	65	143	2
C0206	5.0	76	98	2
C0207	6.0	73	94	3
C0208	7.0	79	57	2
C0301	6.0	74	88	2
C0302	5.0	57	96	2
C0303	2.5	51	51	3
C0304	5.0	77	97	3
C0305	5.5	80	130	2
C0306	4.0	77	107	2
C0307	7.0	80	149	1
C0308	6.0	76	116	1
C0401	5.0	75	150	2
C0402	5.0	82	100	2
C0403	6.0	73	170	2
C0404	7.0	83	129	3
C0405	6.0	73	89	2
C0406	6.0	87	126	2
C0407	8.0	78	138	2
C0408	5.0	82	192	2
C0501	6.0	75	224	2
C0502	6.0	72	126	2
C0503	6.0	73	179	3
C0504	7.0	71	113	3
C0505	6.0	74	91	2
C0506	7.0	71	78	2
C0507	9.0	82	81	3
C0508	6.0	89	143	3
D0101	2.5	52	55	2
D0102	4.0	63	14	3
D0103	4.0	90	166	3
D0104	5.0	81	193	2
D0105	5.0	79	46	3
D0106	4.0	89	76	3
D0107	3.0	83	43	3
D0108	6.5	83	153	3
D0201	4.0	74	129	3
D0202	6.0	84	32	2
D0203	6.0	84	146	2
D0204	2.5	62	79	2
D0205	5.0	60	81	3
D0206	4.0	87	80	2
D0207	10.0	72	83	2
D0208	3.5	72	23	3
D0301	8.0	67	79	2
D0302	5.5	74	123	3

D0303	5.0	71	71	3
D0304	5.5	74	88	2
D0305	6.0	79	88	2
D0306	6.0	88	122	2
D0307	6.0	77	73	2
D0308	7.0	86	100	2
D0401	6.0	70	174	2
D0402	5.0	64	114	2
D0403	6.5	63	120	3
D0404	6.0	61	125	2
D0405	6.0	77	108	2
D0406	6.0	74	103	2
D0407	6.0	78	154	2
D0408	7.0	77	178	3
D0501	5.0	78	74	3
D0502	4.0	77	51	3
D0503	5.0	81	132	3
D0504	5.0	85	113	3
D0505	6.0	69	97	3
D0506	5.0	71	139	2
D0507	9.0	80	181	3
D0508	9.0	82	86	2
D1001	3.5	64	21	2
D1002	4.0	79	39	3
D1003	4.0	81	41	2
D1004	3.0	72	10	3
D1005	4.0	52	37	3
D1006	9.0	74	111	2
D1007	3.0	54	31	3