Utrecht University
Faculty of Geosciences
M.Sc. Physical Geography

GEO4-4415 MSc Thesis

Academic year 2011-2012

Spectrometry for plant species recognition in Mediterranean forest canopies

spectral separability and image unmixing studied on field and aerial data

Student: G. Buzzo
Supervision: E. A. Addink
S. M. de Jong

Summary

Mediterranean-type forests are formed by heterogeneous, dense communities. This condition represents a challenge for plant species discrimination using optical remote sensing. The detection of species based on top-canopy reflectance is hampered by the fact that spectral signatures of different crowns/species mix up in the spectrum sampled by the sensor.

The thesis develops a method which integrates field and aerial hyperspectral data together with a ground validation dataset. Object-based image analysis and spectral unmixing techniques are used to estimate the species fraction cover.

The study area is located in Hérault province – France; the hilly landscape of former farmlands presents a continuous forest cover. The research considers the six arboreal species which are regularly found in the forest canopies, throughout the whole study area.

The fieldwork component of the project was carried out in September-October 2011. Crown and leaf reflectance were measured in-situ with a spectrometer. The species composition and other characteristics of the canopy were recorded at sampling plots.

As an upgraded method for species mapping, the thesis investigates the viability of introducing "mixed" classes in the image spectral mixture analysis. The mixed classes are identified on the basis of the ground observations and are generated from the "pure" ones retrieved by the spectrometer.

A statistical analysis on the field spectra precedes the image analysis. Statistic tests serve to study the spectral discriminability of the species and their mixed combinations; the tests also isolate a subset of bands used in the image analysis.

The crown spectra produce the endmembers for the image spectral mixture analysis. Pure and mixed endmembers are both used. Linear Spectral Unmixing is applied to a previously segmented HyMap image.

The accuracy assessment is performed on the basis of the sampling plots. The species fraction cover observed on the ground is compared with that estimated on the reflectance.

The input of mixed endmembers in the image mixture analysis results in a smaller estimation error on average. The thesis concludes that using mixed endmembers is a valuable option to improve the species mapping in a heterogeneous forest.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

I.1 BACKGROUND

VEGETATION

The human society relies on plants as a source of food, fuel and raw material. Next to that vegetation detains key functions in the biosphere equilibrium for it regulates the energy and the mass exchange between the Earth surface and the atmosphere. Since plants are the almost unique sources of primary production, terrestrial ecosystems necessarily evolved in dependence of vegetation (Leuschner, 2005).

Vegetation regulates energy and mass fluxes and is at the basis of the ecosystem interrelations among organisms; the research interested in these services hence seeks an adequate knowledge of the floristic composition of the vegetation cover. Knowing the species distribution in the landscape may serve to optimize environmental models such as those related to water cycle (de Jong & Jetten, 2007). Forest species composition is precious information for natural resource management, environment protection, biodiversity and wildlife studies (Gong et al, 1997).

THE MEDITERRANEAN BIOME

Mediterranean bio-climatic zones are found north and south of the equator, between approximately 30° and 40° latitude. These zones correspond to five relatively small regions: the proper Mediterranean basin, California, Chile, South Africa, and the southern east and west edges of Australia (Di Castri, 1983).

These zones are associated by a similar climate with a seasonal pattern: warm-hot and dry summers alternate with cool and wet winters (Nahal, 1983). The climatic conditions are among the driving factors of the physical geography; the similarity of the climate explains some of the affinities found in the geomorphology, the pedology and more generally in the landscape (Bradbury, 1983).

In the proper Mediterranean zone vegetation formations are named differently according to the language: *matorral* in Spanish, *maquis* in French, *macchia* in Italian. Their structure and their floral composition depend on a series of factors both of natural and of human origin.

Mediterranean vegetation is adapted to sustain drought periods in summer and cold weather in winter. In the Mediterranean basin climate and soil have a gradient along a north-south direction as well as according to the altitude (Nahal, 1983). Such spatial variability is reflected by that of vegetation. Vegetation adapts its structure and composition according to the available resources (water, minerals, etc.).

Over the last millennia the human presence shaped the landscape in the Mediterranean region. The natural features of the vegetation cover have been deeply modified; the formations observed nowadays are the result of the complex interaction between man and the environment.

The composition of these forests is dominated by evergreen broadleaf vegetation of the *Fagaceae* family, mainly *Quercus ilex* and *Quercus suber* (Tomaselli, 1981). *Ericaceae* (including *Arbutus unedo* and *Erica arborea*) and gymnosperms (pines, junipers) are also abundant, but they are usually associated with a degraded stage of the natural Mediterranean forest (Debazac, 1983).

The regions bordering the Mediterranean attract the attention of geoscientists because of the rich diversity encountered in their natural ecosystems. Furthermore these regions are subject to major interests due to the high population and the relevance for national economies.
These peculiarities motivate the research in the Mediterranean area.

SPECTROMETRY

Applications of spectrometry rely on two basic facts: light is composed of different wavelengths; the interaction between light and matter varies according to the wavelength of the incoming radiation and the material properties of the object. To understand the two circumstances one has to consider the nature of light.
The electromagnetic radiation propagates in the form of waves of different length forming the so-called electromagnetic spectrum. An irradiating source emits energy discontinuously across the spectrum, according to its absolute temperature and its emissivity. The Sun is the main source of radiation on the Earth surface; what is commonly indicated as visible light corresponds to the radiation proceeding with wavelengths in the range of 0.4-0.7 μm (Lillesand et al, 2004).
When light encounters an object three processes may be observed: reflection, transmission, absorption. Materials behave differently on the basis of their chemical and physical conditions (Lillesand et al, 2004).
The term “imaging spectrometry” is used to indicate the techniques that analyze the high-resolution spectral signal reflected by a target illuminated by a natural or artificial source. The domain is often restricted to the shorter wavelengths, approximately from 0.4 to 3 μm, covering the visible, near- and mid-infrared bands (Kumar et al, 2002).

VEGETATION SPECTRUM

The typical features of the vegetation reflectance spectrum are easily recognizable in a leaf spectrum. A leaf spectrum has peculiar features different from non-photosynthetic matter. How the radiation is reflected, absorbed or transmitted depends on the characteristics of both the incident radiation and on the plant leaf surface roughness, the cells’ chemical components and structure (Kumar et al, 2002).

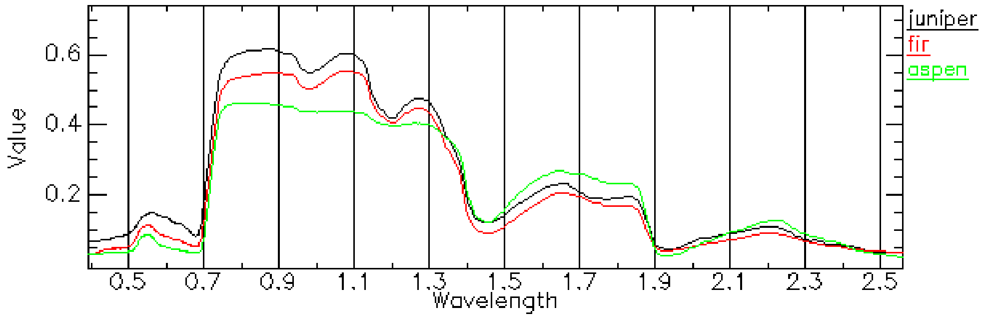


Figure 1. Examples of vegetation reflectance spectra.
The plots are generated in ENVI using data from the USGS library.

In the visible light wavelengths (0.4-0.7 μm), the region where the intensity of solar radiation is maximum, vegetation absorbs the radiation for the photosynthesis (Figure 1). The pigments (chlorophyll, carotene and xanthophyll) absorb light selectively because only with certain energy amounts, i.e. only with certain wavelengths, the electron transitions needed for the photosynthesis can take place. In result the vegetation spectrum shows deepest absorption features in the blue (~ 0.45 μm) and red (~ 0.65 μm) channels, separated by a relatively higher reflectance in the green one (~ 0.55 μm).

In the region between 0.69 and 0.72 μm the vegetation spectrum shows a characteristic feature known as "red-edge": in a relatively small channel, as the wavelength increases, the reflectance spectrum passes from minimum to maximum values (Figure 1). Since the absorption in the visible depends on the life biological processes of a plant, the slope of this edge is a useful indicator of the vegetation status.

The highest reflectance is located in the near-infrared region of the spectrum (0.7-1.3 μm). The spectral features are determined by the leaf cells structure, i.e. the cells arrangement, size and shape, and by the distribution of air spaces within the leaf (Figure 1).

Finally, the spectral signature of vegetation in the mid-infrared region (1.3-2.5 μm) is strongly influenced by the water content (Figure 1). Transitions in the vibrational and rotational states of the water molecules cause a general low reflectance with absorption features around at 1.45, 1.94 and 2.5 μm (Kumar et al, 2002).

Scale of observation and measurement devices

Measurements of vegetation reflectance can be performed at different scales of observation: leaf, branch, crown and canopy. The finer scales (leaf, branch) usually correspond to measurements form field or laboratory instrumentation; sensors operating from a remote position are powerful means to retrieve data at the coarser scales (crown, canopy) and to get an overview of a larger area.

Laboratory spectrometers (Clark et al, 2005; Zomer et al., 2009) and field spectrometers (Pu, 2009; van Ardt & Wynne, 2001; Zomer et al., 2009) are widely employed in vegetation studies. Compared to remote sensing devices, these instruments allow greater control in the measurement conditions. The sensors can be positioned close or in contact with the target;

other objects are thus excluded from the scene. Artificial light sources guarantee the precise repeatability of measurements.
Passive remote sensing of radiance provides a series of advantages: sensors retrieve continuous measurements of the Earth surface in a multi-scale and multi-temporal dimension. Remote sensing imaging spectrometry has numerous applications in vegetation science. The coverage of large areas becomes a necessity for analyses at a regional landscape level. As an example Ghiyamat & Shafri (2010) reviewed the literature on remote sensing imaging spectrometry for biodiversity assessment

I.2 SPECIES RECOGNITION IN HYPERSPECTRAL DATA

Studies confirm the utility of laboratory and field spectrometry to discriminate among vegetation species (Clark et al, 2005; Pu, 2009; van Ardt & Wynne, 2001; Zomer et al., 2009).
Species recognition via remote sensing is more problematic since different plants may have similar canopy-reflectance properties (Pu, 2009).
Remote sensors sample all radiation reflected from the scene; non-green parts of the plants, litter and soil cannot be avoided. The atmosphere is major source of noise in remote sensing, that on the contrary can be ignored for laboratory and field measurements. For hyperspectral images this noise may result in a loss of those spectrum features relevant for the recognition of species.
Generally the crown structural conditions and the way leaves are arranged are species-specific features; these are in effect the characteristics which commonly allow us to tell a tree from another one. Notwithstanding remote sensing spectrometers rarely measure single crowns, i.e. species. Airborne and spaceborne sensors sample vegetation reflectance on discrete portions of the Earth surface. This means that often one pixel synthesizes the reflectance of several plants, eventually of different species. So the identification of species is complicated by the "mixed" nature of the pixel spectrum. In Mediterranean evergreen forest, for instance, canopy has a great variability even over small distance. Different species thrive in dense formations; on average, crowns extend less than 10 mq. This size is much smaller than the spatial resolution usually operated by remote sensing spectrometers.
The quality of the image is affected by the lighting conditions which vary in space and time (Pu, 2009). Optimal illumination conditions, full Sun high on the horizon, are expected to improve image analysis, also when it comes to species recognition via remote sensing spectrometry. For instance, it is demonstrated that endmembers sampled on sunlit-canopy pixels give higher accuracy in species image classification (Clarks et al, 2005; Lucas et al, 2008; Youngentob, 2011).
The timing of the image acquisition is critical. The life cycle of Mediterranean vegetation has seasonal patterns; plants' crowns undergo seasonal changes such as the sprouting of new leaves, flowering and fruiting. Identification of species from remote sensing could clearly benefit from

phenology. To this regard constraints concern the temporal dimension of the datasets. The literature does not count examples using hyperspectral time-series; moreover the acquisition of the images does not necessary coincide with the time of the year when phenology differences are most evident.

The state-of-the-art regarding the species identification in hyperspectral data presents a series of benchmarks concerning both remote and non-remote sensing:

- It is recognized that the reflectance spectrum of vegetation contains information relating to taxonomy; spectral signatures of vegetation have been proven to be species-dependent (Lucas et al, 2008; Manevski et al, 2010; van Ardt & Wynne, 2001; Gong et al, 1997).
- The inter-species spectra divergence may be more clearly sensed in some regions of the spectrum (Manevski et al, 2010; Schmidt & Skidmore, 2003).
- When applied to remote sensing images covering a vegetated area, it is then assumed that it is possible to derive from hyperspectral dataset information related to canopy composition in terms of vegetation types and eventually of vegetation species (Clark et al, 2005; Youngentob et al, 2011).
- The choice on image analysis method is not neutral; when tested on the same data different techniques can produce significantly different accuracies (Buddembaum et al, 2005; Clark et al, 2005).

Some of the gaps in the current knowledge are targeted by this thesis research.
Species recognition in heterogeneous Mediterranean vegetation represents in effect a rather new study case for an application of remote sensing imaging spectrometry. Sluiter 2005 investigates the same study area, yet using hyperspectral imagery for the recognition of vegetation types and not of species. Manevski et al (2010) dealt with species recognition in Crete vegetation, but only concerning field hyperspectral data.
Most importantly image classification is done on the basis of heterogeneous endmembers. Heterogeneous endmembers are here generated *ad hoc* by artificially mixing pure species spectra. There are no examples of spectral library built in this way.

I.3 OBJECTIVE

The overall objective is the identification of species in the forest canopy via the analysis of hyperspectral data. The thesis's ambition is to provide an original contribution to the study heterogeneous Mediterranean forest and in particular to classification of hyperspectral aerial dataset. Statistic analyses, image spectral unmixing and segmentation techniques are employed to achieve this objective.

Specific objectives of the research are formulated into the following questions:

1. Are the dominant vegetation species of the study area spectrally separable?
2. Is it possible to derive the composition (in terms of species) of Mediterranean forest from remote sensing imagery using endmember generated from field spectra?
3. Does the mixture analysis improve if, instead of homogeneous ones, it uses heterogeneous vegetation classes, i.e. composite endmembers generated as a combination of (pure) field spectra?
4. Do adjustments such as continuum-removal and/or data resizing improve the species composition mapping?
5. Do quantitative or qualitative aspects of the vegetation cover influence the classification accuracy?

General objective	Species identification in hyperspectral data for the description of the vegetation composition on a landscape scale.
Study area	Middle catchment of the River Peyne, in southern France.
Techniques	Field and remote sensing spectrometry. Field sampling of vegetation.

	Questions		**Activities**
1	Are the dominant vegetation species of the study area spectrally separable?	▶	Separability among species field spectra is investigated by the means of statistic analyses (Wilcoxon test).
2	Is it possible to derive the species composition of Mediterranean forest from remote sensing imagery using endmember generated from field spectra?	▶	Image linear spectral unmixing; retrieval of species estimations; validation against the site observations.
3	Does mixture analysis improve if it uses composite endmembers generated as a combination of (pure) field spectra?	▶	Building of mixed-spectra spectral library; iteration of image spectral unmixing using pure and mixed reference spectra; accuracy assessment.
4	Do adjustments such as continuum-removal and/or data resizing improve the species composition mapping?	▶	Iteration of image spectral unmixing using different inputs.
5	Do quantitative or qualitative aspects of the vegetation cover influence the classification accuracy?	▶	Analysis of the correlation between classification accuracy and the information on the vegetation cover collected in the field.

Table 1. Summary table of the thesis' contents.

II. Data and methods

II.1 Study area

The environment

The study area is located in southern France, about 60 kilometers from the Mediterranean. It is situated in the Hérault department of the Languedoc-Roussillon province, in the vicinity of Montpellier and Béziers; the area is next to the small town of Vailhan (43°33'9.44"N, 3°18'9.79"E). The area is in the central part of the catchment of the river Peyne (Figure 2).

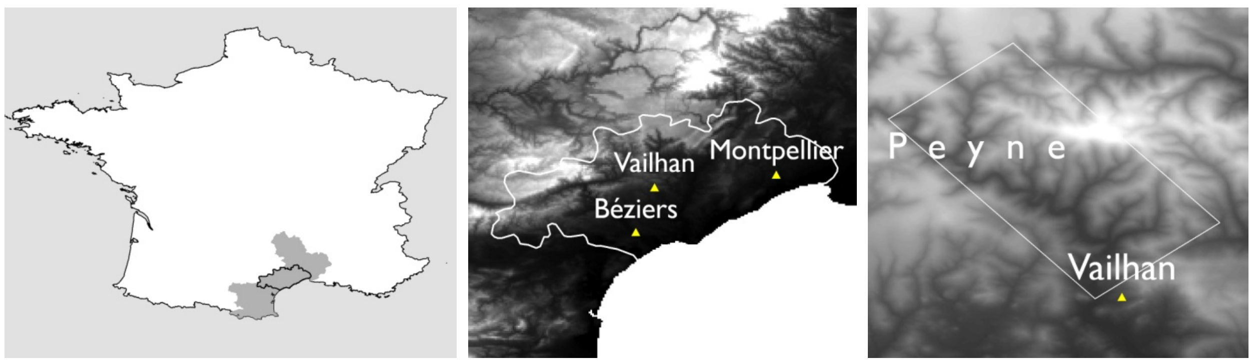


Figure 2. Location of the study area.
The map on the left displays Languedoc-Roussillon province in grey shading; within the province, the boundary of Hérault department is enclosed in a black contour. In the central map are the Hérault department and the localities mentioned in this chapter. The map on the right shows the location of the study area to the north of Vailhan, along the course of the Peyne.

The area lies on the southern limits of the Massif Central, a region characterized by the co-presence of many different geological substrates. In particular, the study area geology is formed by flysch, limestone and basalt (Sluiter, 2005).

The morphology is that of a transition between the plains stretching along the Gulf of Lion shores and the hilly hinterland. The elevation ranges between 100 and 500 meters a.m.s.l. The river Peyne crosses the area following a NW-SE course being interrupted by an artificial dam; the Lac des Olivettes occupies the center of the study area (Figure D.1; Appendix D).

The region is characterized by a sub-humid Mediterranean climate with hot, dry summers and cool winters (Sluiter, 2005).

The anthropic factor

During the last century, the temporal evolution of human occupation in the study area followed a trend that can be generalized to much of the South-Europe rural areas. Traditional socio-economic models kept some vitality until the early XX century; in the contemporary economy

originated after WW2 these models have been quickly abandoned. As a result of the loss of economic relevance rural areas generally suffered substantial depopulation.
In the study area, the traditional grazing economy has widely disappeared and rural communities shrunk as inhabitants moved to bigger urban centers. Grazing open pastures together with much of former farmland have reverted into natural shrub- and woodland. More recently, say in the last thirty years, depopulation has stopped; benefiting from the renovate transport network, the demography has new inputs from tourists and from residents working outside the area. This revival of the human presence has already determined modifications in the land-use and sometimes changes in the vegetation cover (Sluiter, 2005).

FOREST FORMATIONS

The various combinations of local climate, elevation, lithology and human influence explain the wide range of growing conditions and the variety of vegetation formations observed in the study area. In the Mediterranean biome, vegetation communities are classified according a varied methodology and hitherto terminology; simplifying the issue, it is possible to say that the study area is entirely within the region of the *temperate broad-leaved evergreen forest* (Debazac, 1983).

Despite the limited extension, two vegetation formations are recognized in the study area.
The southern portion of the study area is occupied by *matorral*. According to Tomaselli (1981) the matorral is a degraded stage of evergreen forest; he defines it as "a formation of woody plants, whose aerial parts are not differentiated into trunk and leaves, because they are much ramified from the base, and are of shrubby habit"; based on canopy's height one can distinguish *tall matorral* (2-5 meters), *middle matorral* (0.6-2 meters) and *short matorral* which is dominated by herbaceous species. More subgroups can be distinguished based on density and species composition (Tomaselli, 1981). According to this classification, the forest to the south of the study area can be described as *dense tall matorral*.
In the northern hilly reliefs the matorral leaves the pace to formations approximating the climax formation, i.e. the *mixed deciduous oak forest*.

The area is indeed rich in different formations and succession stages. In terms of forest type the study area represents a transition from the Mediterranean sclerophyll forest to the temperate deciduous forest, the latter being regularly encountered northward in the Orb valley and the surrounding reliefs.
The south-north gradient of the tree species composition parallels the one of elevation, microclimatic conditions and of human pressure. The elevation increases gradually from south to north sustaining wetter and colder meteo-climatic conditions in the northern part; in the southern part of the study area the proximity to human settlements and activities is likely to have prevented the formation of that type of late-stage forest which is more easily observed to the north.

SPECIES COMPOSITION OF THE FOREST CANOPY

The chief characteristic of the forests of the area, and in particular of the matorral, is the dense association of small-to-medium size plants belonging to different species. This results in highly heterogeneous canopies. A varied composition does not imply a high diversity of species; the forest canopies include a rather limited number of species, a number that sensibly decreases if we discard the rare species. Table 2 lists the arboreal species recurrent in the study area; the information is derived both from literature and from in-situ observations.

group	**family**	**species**
angiosperms	Aceraceae	*Acer monspessulanum*
	Anacardiaceae	*Pistacia lentiscus*, *Pistacia terebinthus*
	Buxaceae	*Buxus sempervirens*
	Caprifoliaceae	*Viburnum tinus*
	Ericaceae	*Arbutus unedo, Erica arborea*
	Fabaceae	*Spartium junceum*
	Fagaceae	*Castanea sativa*, *Quercus coccifera*, *Quercus ilex*, *Quercus pubescens*
	Oleaceae	*Phillyrea angustifolia*, *Phillyrea latifolia*
	Rosaceae	*Crataegus azarolus*
gymnosperms	Cupressaceae	*Juniperus oxycedrus*, *Juniperus phoenicea*
	Pinaceae	*Pinus halepensis*

Table 2. Species recurrent in the forest canopies of the study area.

The dense-tall-matorral canopies are dominated by an association of *Quercus ilex* and *Arbutus unedo* with a more rare presence of *Phillyrea latifolia*. *Erica arborea* is also frequent yet in younger-stage successions. The other species are all regularly encountered though in significantly smaller proportions. *Viburnum tinus* is regularly present in the understory which is generally poorly developed. Ground and standing litter is often abundant.

The gymnosperms species also make part of the matorral communities; they usually have a scattered distribution occupying areas with lower or absent arboreal vegetation.

In the mixed-deciduous-oak-forest the deciduous species *Acer monspessulanum* and *Quercus pubescens*, rare in the matorral, become frequent. This formation is dominated by an association of *Quercus ilex* and *Quercus pubescens*. The understory is richer than the one of matorral, counting a number of species among which *Buxus sempervirens* is predominant.

II.2 IMAGERY

This study uses HyMap imagery acquired on the 23rd of July 2008. The HyMap system has an IFOV (Instantaneous Field Of View) of 2.5 mr along track and 2.0 mr across track. The FOV

(Field Of View) is 61.3 degrees (512 pixels). With usual flight conditions the GIFOV (Ground Instantaneous Field Of View) is about 3-10 m (HyVista, 2011). The sensor is an assemblage of four instruments each operating in 32 spectral bands (Table 3).

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
SWIR1	1.40 – 1.80 μm	15 – 16 nm	13 nm
SWIR2	1.95 – 2.48 μm	18 – 20 nm	17 nm

Table 3. HyMap sensor technical details (HyVista, 2011).

The HyMap image consists of 126 bands covering the spectrum from 0.45 to 2.48 μm. It spans over an elongated strip along the Peyne River with a pixel size of 5 meters.
Raw remote sensing data require an articulated sequence of pre-processing steps; these are needed to adjust both the spatial and the spectral information. Such corrections determine the quality of the derived reflectance image, so affecting the overall results of subsequent analyses. Nonetheless an appropriate discussion of this influence would require an attention outside the interest of this work. The conversion from radiance to reflectance and the orthorectification had already been done using the methods described in Schläpfer & Richter (2002) and Richter & Schläpfer (2002).
Only the central portion of the image is considered; the full extent is reduced to the area with dense forest (Figure D.2; Appendix D). Furthermore the image analyses adopt a mask to avoid non-vegetated pixels. Bands 13 (0.633.9 μm) and 21 (0.7494 μm) are used to compute the NDVI; pixels with an index below 0.5 are masked out. The choice of the threshold is based on a visual interpretation.

II.3 IN-SITU DATA ACQUISITION

Fieldwork is an essential component of this thesis. The general goal is to consolidate the knowledge about the study area; autoptic observation can indeed help to evaluate the elaborations' outputs and eventually avoid rough mistakes.
Two specific tasks are related to the thesis research objectives: field spectra collection, site sampling. The first task provides the input for the separability analyses; moreover field spectra serve to generate the endmembers for image classification. The purpose of the second task is to acquire the validation dataset used by the accuracy assessment of image classification.

Fieldwork took place in September and October 2011. Within the geographic boundaries given by the HyMap dataset, only forested portions have been considered. The survey area covers

approximately 16 km^2 and is located in the surroundings of the Lac des Olivettes (Figure D.1; Appendix D).
The observations and the spectra measurements were done at a number of locations following the sampling scheme and methodology explained in more detail in the following paragraphs.
The town of Neffiès served as base point. The eight weeks campaign was conducted by a team of two surveyors; a car was used to approach the neighborhood of the sites which were successively reached on foot.

FIELD-SPECTRA COLLECTION

Reflectance spectra were obtained with an ASD FieldSpec FR spectrometer. The instrument is made up of three spectrometers each operating in a different region of the spectrum: the first in the visible and the near infrared (350-1000 nm), and the other two in the short wave infrared (SWIR1 1000-1800 nm, SWIR2 1800-2500 nm). The sampling interval is 1.4 nm for the region 350 - 1000 nm and 2 nm for the region 1000 - 2500 nm. The spectral resolution is 3 nm for the region 350 - 1000 nm and 10 nm for the region 1000 - 2500 nm (ASD, 1999). The FieldSpec FR retrieves the values of 2151 bands spanning from 350 to 2500 nm.
When in function, the instrument measures continuously; at the user's command one spectrum is saved as the average of 50 successive measurements. At the beginning and regularly during the FieldSpec measurements session, calibration of the instrument was performed using a reference white target.
Two optical arrangements have been used. Using the bare one-meter optic fiber, handled with the help of a pistol grip, it is possible to obtain the maximum FOV, 25 degrees. Measurements were performed also with a leaf-clip probe that uses an artificial light source (Figure 3).
The adoption of the first of the two arrangements is meant to retrieve the reflectance at the branch and crown spatial scale, assumed to be better comparable with aerial imagery such as the HyMap dataset.
The spectra collected with the leaf-clip probe provide substantial arguments to the research interests. The controlled measurement conditions are expected to guarantee reliable results on the separability of different vegetation species.

25° optic

leaf clip

Figure 3. ASD FieldSpec FR operative scales: 25° optic, leaf clip.

The task considered the six main species: *Arbutus unedo, Castanea sativa, Erica arborea, Phillyrea latifolia Quercus ilex, Quercus pubescens*.
The suitable locations for the FieldSpec measurements had adequate presence of the target species and easy accessibility. Spectral signature of vegetation shows intra-species variability due to multiple factors depending on the local context where the single plant is located: soil type, light and water availability among others. Minimizing the effect of these local factors is fundamental to retrieve a reference average spectrum; sampling sites were therefore appropriately spread all across the study area.
Three measurements were repeated for each individual plant; in the successive phase of data preparation these three measurements were averaged to produce one single-plant spectrum. Each of the six species has on forty samples: twenty spectra for both scales, each one generated as an average of three FieldSpec measurements.

SITE SAMPLING

The forest sampling generated a consistent dataset to be used for the validation of the image analysis. In-situ observations evaluated the canopy composition and its height; these variables are relevant for the variability of the spectral signature. The operations performed at the plots are:

1. species identification (the options were limited to the same six species considered for the spectrometer measurements);
2. estimation of the relative abundance in the canopy of each identified species;
3. estimation of the canopy height.

The site sampling proceeded according to a stratified-random-cluster scheme (Nijland et al, 2009). The approach is meant to achieve a compromise between the advisability of concentrating the efforts on "promising" subsets, and the necessity to minimize bias.
Four strata were chosen on the basis of multiple criteria: distribution across the whole area, presence of continuous forest and proximity to a suitable access point (Figures D.1-3; Appendix D). Within the strata, eight clusters are formed by eight sampling plots.
The first sampling site of a cluster was identified on a 1:25.000 topographic map; hills tops or flatter surfaces were preferred as staring node. Once the observations at one site were completed, the following one was identified in a random direction at a distance of 50 meters from the previous. The random component is included in the sampling scheme to ensure greater statistical value.
All positioning operations were made using a Garmin GPS. Positional uncertainty ranged between 3-5 meters.
A total of 169 sampling sites were visited. Of these, 160 are equally subdivided among strata so that each includes five clusters made up of eight sites. Nine extra sites were directly identified on the ground; they contained species that seemed under-represented by the stratified-random-cluster sampling scheme.

The plots were squares of 5 m side. This revealed to be the most appropriate operative method. With such a size a surveyor, standing at the plot center, is able to observe its whole canopy and estimate the composition with the check of the second surveyor. Adopting a larger plot size in the forest of the study area is likely to complicate significantly this task.

II.4 DATA PREPARATION

FIELD SPECTRA

Operations are performed mainly in MATLAB and for some steps in ENVI. FieldSpec FR measurements can be exported singularly from the ASD format into *.txt files which easily provide the input to the MATLAB functions.

As mentioned in the previous paragraph, multiple spectra were collected for each individual plant; therefore the first operation obtains one average spectrum from the single measurements.

Furthermore the hyperspectral data require additional treatment before proceeding with the spectral analyses. Truncation, re-sampling, transformations are adopted for a series of purposes: noise removal, reduction of the data dimensionality, enhancement of the variability of the informative content.

The two optical configurations produced spectra with different characteristics in terms of noise/signal rate.

When measuring the reflectance using the 25° optic, the signal is affected by the medium; the moisture present in the air is responsible for noise of variable magnitude in the water absorption bands, i.e. around 1.4 and 1.8 μm. The noisy bands of the water absorption regions as well as those of both spectrum edges are removed. In result, from the original 2151, the spectra produced with 25° optic have 1817 bands.

The leaf-clip probe operates with much greater control of the target illumination and of the registration of the reflectance signal. The noise concentrates in the shorter wavelength; the first hundred bands are removed so that 2052 are left.

Spectra are resized and re-sampled to the shorter spectrum range and to the lower spectral resolution of the HyMap image.

The re-sampling operates by averaging the values within an interval. Intervals are identified on the FieldSpec data by taking the HyMap bands as centers. The size of the interval is regulated according to the spectrum region so to simulate the spectral resolution of the HyMap ("Bandwidth across module", Table 2).

The 25° optic data, which had originally 1817 bands, have 114 bands in the re-sampled spectra. The 2052 bands of leaf-clip data becomes 126 in the re-sampled spectra.

Continuum-removal is applied. The field spectra already cleansed from the noisy bands, are imported in ENVI as a Spectral Library and continuum-removed. Then the data is exported to ASCII format and re-imported in MATLAB. Re-sampling as previously described is hence repeated for the continuum-removed data.

The continuum removal (CR) transforms a reflectance spectrum to a range 0:1, by normalizing the overall brightness and enhancing the absorption features.

A continuum, the so-called "convex hull", is drawn so to connect with a straight line the spectrum maxima (Figure 4). The spectrum maxima matching the continuum assume a value of one, while for the other wavelengths the continuum-reflectance depends on the distance between the two curves. To obtain a continuum-removed spectrum, the original reflectance is divided by the continuum-reflectance (Youngentob et al, 2011).

The adoption of CR is meant to enhance the visibility of those spectral features which characterize the vegetation species; indeed it removes the disturbing effect of the overall albedo which is regulated by other factors such as soil and canopy (Schmidt & Skidmore, 2003).

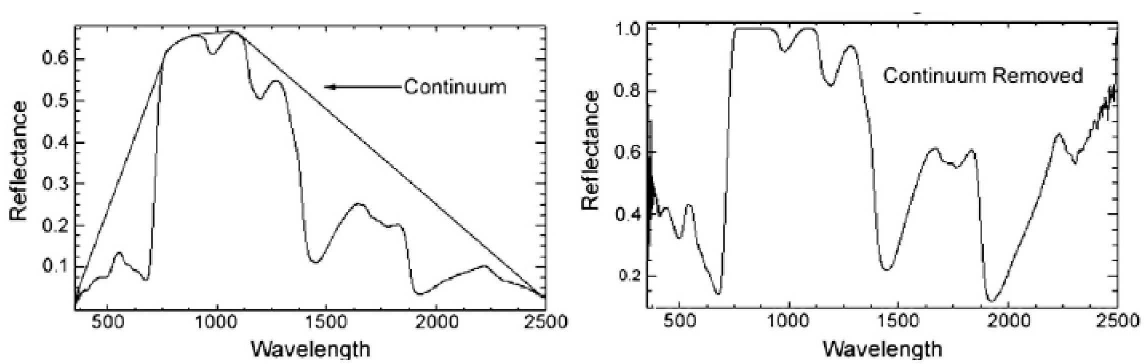


Figure 4. Continuum removal for a vegetation spectrum (Youngentob et al, 2011).

From the re-sampled field spectra mixed spectra are generated.

One of the specific objectives of the thesis concerns image analysis, i.e. spectral unmixing on the basis of vegetation species classes. The question is whether endmembers composed by a mixture of species spectra perform better than species pure endmembers. By "mixture of reference spectra" it is meant an artificial composition of pure spectra.

As the mixed spectra are meant to be used as endmembers in the HyMap unmixing, only the re-sampled spectra are taken into account.

Mixed spectra are obtained only by the combination of two species which in turn can assume only relative weights of 25%, 50% or 75%.

The choice of the significant combinations is done on the basis of the reality verified on the ground; the effectual species combination and the weight to be assigned to each member of the pairing is limited to a set of ten mixed spectra. They are all formed by *Quercus ilex* and a second species (Table 4).

Quercus ilex	75	25	*Arbutus unedo*
	50	50	
	25	75	
	25	75	*Castanea sativa*
	50	50	*Erica arborea*
	25	75	
	75	25	*Phillyrea latifolia*
	75	25	*Quercus pubescens*
	50	50	
	25	75	

Table 4. Mixed classes.
The outer columns indicate the species, while the inner columns specify the proportions used to generate the mixed spectrum (values in percentage %).

Each of the six species has twenty samples for either one of the two scales. If all possible combinations are used to produce the mixture of, for instance, the twenty spectra of *Quercus ilex* and the twenty of *Quercus pubescens*, a useless overabundance would occur. The reduction of the data is achieved by calculating five average spectra for each species. In this way the generation of the mixed classes can be done on two classes each made up of five samples. The resulting mixed classes have twenty-five samples.

The calculation of the five average spectra followed the same steps for the six species:

1. Proceeding band-by-band, the twenty spectra are ranked on their value, i.e. the reflectance;
2. The "values rank" of one spectrum is calculated as the average of its bands' ranks;
3. Spectra are sorted according to their "values rank";
4. Sorted spectra are grouped into five non-overlapping sets;
5. The average spectrum of each of the five sets is calculated.

The five average spectra obtained in the last step are used as input for the mixing. Twenty-five samples can be obtained for each mixed class; the combinations are those displayed in Table 4.

The steps just described produce a total of twelve data configurations, six for each of the two scales of observation (Figure 5). Configurations differ on the FieldSpec FR optical device (25° optic or leaf-clip), on the type of data (original reflectance or continuum-removed), on the number of bands (FieldSpec FR or HyMap resolution), on the number of classes (six or sixteen). Classes have about twenty samples. Such size seems to provide a sufficiently consistent dataset. Iterations are necessary to represent the inner variability of one class and so to sustain the statistical reliability of the analyses. In Schmidt & Skidmore (2003) twenty is the minimum amount of spectra per class.
The twelve data configurations serve for the spectral separability analyses later discussed.

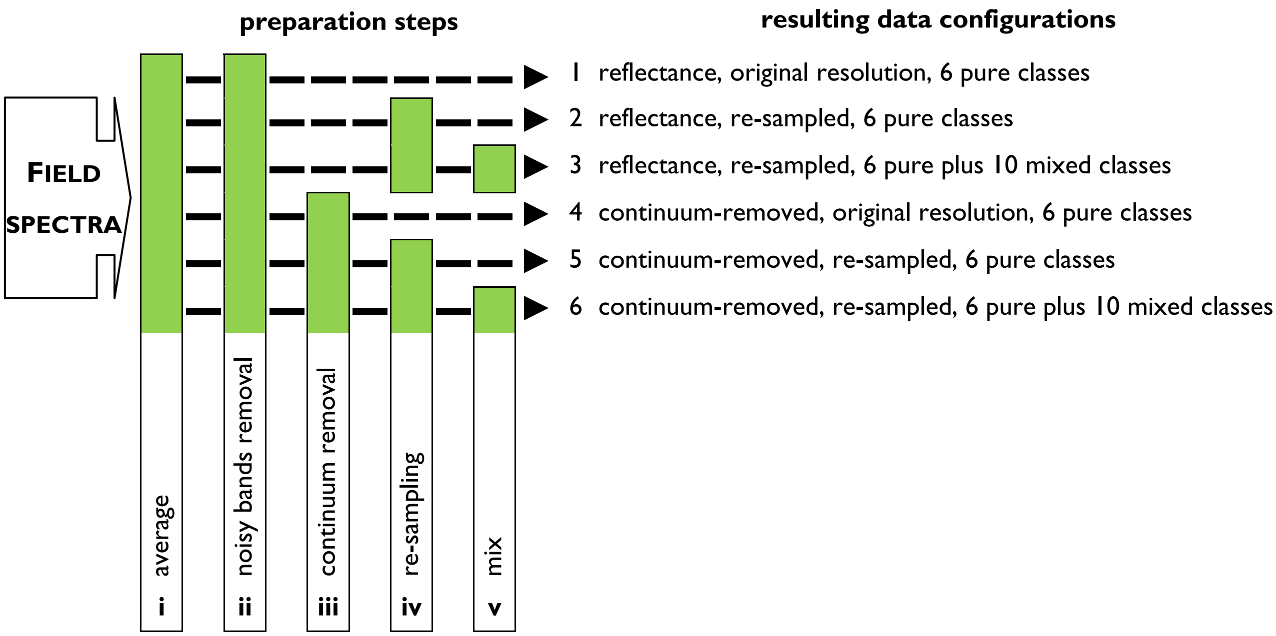


Figure 5. Preparation of ASD FiledSpec Fr data.
Spectra collection adopted two scales of observation: crown (25° optic) and leaf (leaf clip probe).
A species spectrum is obtained as average of three measurements (i). The spectra are cleaned from noisy bands (ii). The spectra form directly a spectral class otherwise they pass through other preparation steps (iii-v).
In the end twelve spectra collections are generated, six for each of the two scales of observation.

A portion of the dataset is further processed; field spectra are needed to create the endmembers for image classification. ENVI Spectral Libraries are built. Only the 25° optic re-sampled data are considered since this scale and this spectral resolution are more compatible with those of the aerial imagery. The samples of each class are averaged to obtain one endmember. The procedure is repeated for both original and continuum-removed reflectance.

HYMAP IMAGE SEGMENTATION

"Object-based image analysis" has become a major trend in remote sensing since about year 2000; objects group several pixels on the basis of spectral values and, most importantly, of contextual information such as texture and spatial arrangement (Blaschke, 2010).
The adoption of the technique in this work is motivated by a series of considerations.
The main goal is tackling the positional error, which originates from the HyMap orthorectification, and from the uncertainty associated with the GPS georeferencing of the sampling sites. It is assumed that the exact spatial match of a sampling site with a pixel is highly unlikely, thus undermining the accuracy assessment of pixel spectral unmixing against the validation dataset. On the contrary the objects retrieved by the HyMap segmentation group together a variable number of pixels basing on their reflectance. Due to their similar top-canopy spectra, pixels included in one object are assumed to have a similar species composition; their mean spectrum can be used in the unmixing analysis and whichever of them is suitable to be validated against the field data.

Tracing objects with a common canopy reflectance seems to approximate the natural occurrence of characteristic species associations. The field experience taught that the forests of the study area have usually a heterogeneous canopy composition; the presence and relative abundance of species tend to vary quickly when observed at the high resolution of the in-situ observations. Nonetheless one may observe also larger homogeneous patches consisting of associations of two/three species, for instance that formed by *Quercus ilex* and *Arbutus unedo* or the one including *Quercus ilex*, *Arbutus unedo* and *Erica arborea*. The segmentation is expected to recognize such patches and group together the corresponding pixels.

Operating on objects rather than on single pixels simplifies the HyMap data variability; it allows easing the computational effort of the spectral unmixing and producing more readable outputs which in a pixel-based unmixing would usually require further post-processing.

Segmentation of the HyMap imagery is performed in eCognition software environment using the *multiresolution segmentation* algorithm; the scale parameter is set to 25, color to 0.2, and shape to 1. The spectral unmixing analyses (§ II.6 Spectral unmixing) are performed in another software environment. This requires the segmentation results to be first exported and then adapted through some transformations.

The objects generated by eCognition are exported into ESRI shapefile. The shapefile has 126 attributes (each for one HyMap band) and as many shapes as the number of objects. The value assigned to one attribute of one shape is derived from the respective band, from the mean of the pixels included in the respective object.

The shapefile is then exported into raster maps attribute by attribute, i.e. band by band. These 126 maps have a cell size of 5x5 meters, like the original HyMap. Objects are represented by areas with pixels of equal value.

All 126 raster maps are finally re-assembled into one multidimensional image.

The new “object-based HyMap” is the input of all successive analyses. It is important to note that objects are defined only on the spectrum of the original image. The transformations, continuum-removal and band selection, are applied afterwards, to the outputs of the segmentation procedure.

II.5 STATISTICS FOR SPECTRAL SEPARABILITY

Spectral separability among vegetation classes can be assessed by the means of statistic tests. The hypothesis on the separability of a pair of spectra is verified at a certain significance level; the tests verify if the among-species is larger than the within-species variability. A prerequisite for these statistical analyses is the availability of iterated reflectance measurements for each single sample. The methods published in Schmidt and Skidmore (Schmidt & Skidmore, 2001;

Schmidt & Skidmore, 2003) serve as a basis for the analysis carried out here. The Wilcoxon sum-of-ranks test (W-test) is used to perform a pair-wise comparison of the field spectra.

The W-test is a non-parametric method, which bases on two assumptions: 1) samples are independent; 2) samples are from continuous populations (Gibbons, 1971).
Two samples *X* and *Y* are arranged together and sorted. The sum-of-ranks of *X* in the combined ordered arrangement is calculated.
On the basis of the samples' sizes, the null probability distribution is derived. This is a normal distribution function defined by all possible sum-of-ranks.
According to the desired significance level, the rejection regions are identified on this theoretical normal distribution. Both one sided and two-sided alternative are possible.
If *X* sum-of-ranks falls within the rejection regions, the null hypothesis is rejected (Gibbons, 1971).

The test retrieves binary outputs. "One" stands for the rejection of the null hypothesis: reflectance values belong to different population, i.e. the two vegetation classes are statistically different. "Zero" stands for the acceptance of the null hypothesis: for the tested band the two species are not significantly different.
For all possible pairings of classes, for each individual band, the test verifies whether two classes are significantly different. As the test assesses the hypothesis band-by-band, it gets to emerge those bands more effective to the aim of discrimination. The information is in the utility of the thesis which aims to determine if these bands could enhance the image classification.

The test is performed in MATLAB using the function *runksum*. The significance level is set to 0.05. The W-test is repeated for all twelve configurations previously generated (§ II.4 Data preparation). The raw results of the tests are 1/0 outputs for all bands in use, for each pairing of classes (1 = rejection of the null hypothesis = spectra are significantly different). These values are further processed in two ways:

1. The "net separability" of a class is retrieved by detecting the bands where statistical separability is verified for all combinations of a class. This provides insight in the overall performance of the single class.
2. The test non-null scores are summed band-by-bands to retrieve the overall frequency of statistically separable pairings. High rejection frequencies identify wavelengths with pronounced spectral separability.

Seven bands are isolated on the basis of the rejection frequencies retrieved by the configurations using sixteen classes; the larger number of classes eases the isolation of the maxima. Bands are identified so that they evenly distribute across the spectrum range; the amount, seven, is arbitrary (in Schmidt & Skidmore 2003 the selected bands are six). The operation is repeated for both scales of observation and for both data types.

The identification of the seven bands is a requirement for the following work phase. Spectral unmixing is performed also on the basis of a spectrum subset consisting of these seven bands.

II.6 SPECTRAL UNMIXING

The elaboration of remote sensing hyperspectral images needs to consider the *scene model* resolution, which is regulated by the sensor resolution and the scale of the scene objects (Strahler, 1986). In the case of a sensor with a spatial resolution lower than the scale of the targets, a pixel will synthesize the signal of an initially undefined number of unknown objects. That is what verifies in the case under discussion; the reason is twofold and due to the specificity of the Mediterranean forests under discussion: first, crown dimensions are usually much smaller than the HyMap pixel size; second, plants form dense communities characterized by species heterogeneity, and not large homogeneous patches.
It is assumed that the mixing occurs linearly: the relative weight of a single object/canopy in the pixel spectrum depends on its relative abundance, i.e. fractional area. According to this model, objects are optically separated and multiple scattering is not considered. The assumption, though simplifying the radiative transfer phenomenon, seems valid in the context of this application.

Linear spectral unmixing (LSU) is a supervised method for spectral mixture analysis; it is "supervised" as it assumes a preliminary knowledge, the field endmembers, which is used as reference to process "knowledge-less" data, the segmented HyMap. The technique enables us to quantify the contribution of one endmember/class to the reflectance, and therein to estimate the class's fractional area. The analysis is performed in ENVI software environment.

The imagery obtained after the pre-processing phase serves as LSU input, while the spectral library is the collection of crown spectra derived from the re-sampled, 25° optic data. The whole LSU process is equally repeated for the original and the continuum-removed dataset. LSU is iterated for three endmembers collections: 1) six "pure" spectra, 2) ten "mixed" spectra, 3) sixteen pure and mixed spectra.

Both full-range as well as subset spectra are used. The spectral subset is the one identified on the basis of the W-test.
Some authors regard data reduction as beneficial for image mixture analysis. It can shorten processing time; it serves to tackle disproportion between image dimensionality and the training set when the training set is made up of few samples (Clark et al, 2005; Schmidt & Skidmore, 2003).

In the end a total of twelve ENVI LSU rule images are obtained (Figure 6).

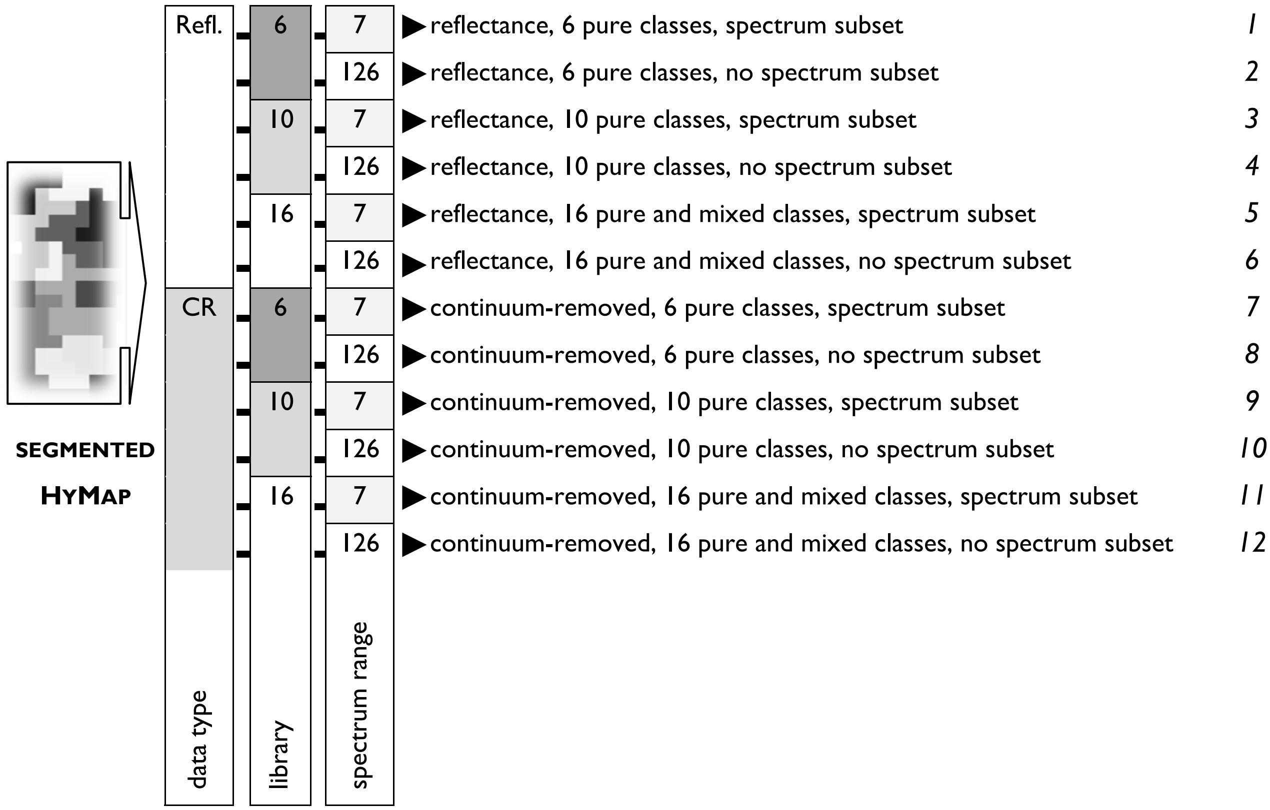


Figure 6. Spectral unmixing procedure.
In ENVI, using the Linear Spectral Unmixing tool, twelve outputs ("rule images") are generated with a different combination of data type (*Refl.* = reflectance, *CR* = continuum-removed), spectral library (*6* = six pure-spectrum classes, *10* = ten mixed-spectrum classes , *16* = sixteen pure and mixed-spectrum classes), spectrum range (7 = spectrum subset, 126 = full spectrum range).

II.7 ACCURACY ASSESSMENT

The data collected by direct inspection in the study area provide the validation dataset for the spectral unmixing results. 169 sample sites are used for the validation.
In a few steps, a set of 169 pixels is extracted from each of the rule images: in ENVI, the GPS coordinates of the 169 sampling sites are used to create a point-type "region of interest" (ROI). The ROI, composed of one-pixel sized parts, is overlaid subsequently to the various LSU rules images and each time the values of the 169 pixels covered by the ROI are exported into an ASCII format file.

Further compatibility with the validation data is needed. First all negative values are set to zero; then positive ones are converted into percentage and treated according to the type of the respective class.

Concerning the results obtained by LSU using only the ten mixed classes, the percentages of the mixed classes have to be decomposed into single pure components and reassemble to retrieve the total percentage of each of the six species.
Similar procedure is followed for the results obtained by LSU using the sixteen pure and mixed classes; the contributions of the "mixed" classes are added to percentages of the pure ones.
Concerning the LSU of six pure classes, once the positive values are re-scaled to percentages, they are directly comparable to the field observations.
In the end, all original LSU estimations at the 169 pixel locations are converted into percentage abundances of the six species.

The LSU outputs, re-elaborated as just described, are compared with the observed canopy compositions. This comparison is implemented in such a way that it is possible to assess the overall accuracy of one configuration, but also to verify if it changes for the different variables in play. The goal is to determine whether the accuracy depends on factors such as the species composition, the degree of heterogeneity of the canopy, the location of sampling site.
The error is calculated as the difference between the estimations and the observations. Out of these raw errors, the median roughly indicates whether the mismatch is due to overestimation or underestimation.
The Root Mean Square Error (RMSE) of one site as well as of one species is obtained as the square root of the average of the squared errors. As a result, for each of the twelve LSU configurations it is possible to derive a set of 169 site-RMSEs and another of 6 species-RMSEs. The comparison among the different configurations can thus proceed either per-site or per-species.
A configuration's average RMSE is calculated for both the procedures; the value serves as an indication of the overall performance of the single configuration.

The dependency of the error on species-specific or site-specific factors is formulized by the Pearson's correlation coefficient (ρ). The coefficient is calculated between observations and average RMSEs. Average RMSEs are per-species or per-site averages of the twelve configurations RMSEs (third to last column in Table C.3-4; Appendix C). The factors considered are occurrence and variability/heterogeneity.
Concerning the per-species assessment, ρ is calculated between the six species-RMSEs and the six sums of species site-occurrence (range 0-169). A second ρ is calculated between the six species-RMSEs and the six standard deviations of the species observations.
Concerning the per-site assessment, ρ is calculated between 169 site-RMSEs and the 169 sums of the site species-occurrence (range 0-6). A second ρ is calculated between the 169 site-RMSEs and a measure of heterogeneity. "Species heterogeneity" is calculated as the inverse of the product of all non-zeros observations; this measure ranges from a minimum of 1corresponding to a 100% cover, to a maximum of 2222 corresponding to a site composition of 30%, 5%, 5%, 60% (1 / [0.3 x 0.05 x 0.05] x 0.6 = 2222).

The procedures described till here are meant to assess quantitatively the fit among estimations and observations. From a qualitative point of view it is assessed the match between the per-site heterogeneity of estimations and that of observations.
The same "species heterogeneity" generated for observations is calculated also for the estimations. Sites are ranked successively according to the heterogeneity of estimations and then of observations. The mismatch between rankings is formulized by the RMSE calculated on the errors between estimations' and observations' ranks.

II.8 SPECIES MAPPING

The accuracy assessment supports the decision on the species mapping. The configuration associated with the lowest average RMSE is used to produce raster maps of the species composition in the study area.
The LSU rule images are converted into six species-layers on a percentage scale (negative values are excluded). In each of the six layers pixels represent estimations of the fraction cover of a species.

III. Results

III.1 Statistics for spectral separability

Spectra separability is verified to a good extent. Leaf spectra have greater separability than crown spectra. Continuum-removal enhances the discriminability among classes.
The results of the W-tests are displayed in Appendix A with graphic and tabular layouts. The statistics are calculated for twelve data configurations; configurations differ in terms of:

1. Spectral resolution (FieldSpec resolution; re-sampled)
2. Scale of observation/optic device (leaf/leaf clip; crown/25°optic)
3. Data type (reflectance; continuum-removed)
4. Classes number and type (6 pure classes; 16 pure and mixed classes)

These four variables influence differently the statistics.

Concerning the spectral resolution, there is a substantial similarity of the results. Higher resolution provides greater definition of the separability. This is particularly clear in the NIR range of leaf clip continuum-removed data (graph *b*, Figure A.10-11; Appendix A); some of the non-separability clusters (black shading in the graphs) are "broken" into a sparser pattern of separable and non-separable bands.

Concerning the scale of observation, there is an overall similarity though with appreciable differences.
For 25° optic reflectance data, a high frequency of rejections is observed in VIS with peaks at 450 and 650 nm, in the red-edge region (700 nm), in NIR with a maximum in the shorter wavelengths, in SWIR1 in correspondence of the water reflectance absorption. SWIR2 has a greater occurrence of rejections than the other regions.
For reflectance data, differences emerge in VIS and NIR. Leaf spectra separability decreases at the extremes of VIS region, while is higher in the central part. The two scales clearly differ in NIR: leaf spectra separability has a maximum in the central NIR region, around 1100 nm, and not in NIR shorter wavelengths as the crown spectra.
The differences between the two scales are far more evident if instead of the cumulated rejections we compare in detail the behavior of single pairings. It is frequently observed that a pairing can maintain the overall net separability, while the rejections occur for completely different bands (nr. 3, Table A.1a; Appendix A). Bands that allow a high separability at crown scale can be insignificant for the leaf scale, and the other way round (nr. 9, Table A.1a; Appendix A).

Concerning the data type, the normalization of the data by continuum-removal enhances the smaller variability against the overall reflectance. The pattern of the rejections is modified into a more dispersed one (graphs *a*, Figure A.7-12; Appendix A).

In the NIR region it is observed a switch of the rejection/non-rejection frequency between absorption and higher-reflectance bands. Indeed the continuum-removal normalizes the NIR maximum reflectance to one, leveling out differences of high-reflectance NIR bands; the frequency of rejections in this region undergoes a general reduction; on the contrary the rejection frequencies of the absorption bands, centered on 1000 and 1100 nm, rise. Such circumstance is more evident for the leaf clip spectra.
Generally, the performance of a pairing relatively to the others tends to be unchanged. It is observed that a pairing with a high/poor separability obtained with reflectance data tends to maintain its rank also with continuum-removed data.

Concerning the number of classes, a larger number diminishes the chance that a band shows a good separability for all classes. It is observed that the more the classes the less the "net separability" (graphs *c*, sub-tables *b*; Appendix A). There are indeed less chances that the W-test rejection occurs for all possible pairings. In result, as the number of classes is enlarged, the bands retrieving a rejection for all pairings diminish or disappear.
The number of pairings multiplies when sixteen instead of six classes are used. The advantage of a larger number of per-band results is evident looking at the bar charts. A much smoother profile characterize the charts related to the set of sixteen classes (graph *a;* Fig. A.3, 6, 9, 12; Appendix A); local maxima are better isolated and serve to localize higher separability (Table 5. Most discriminative bands.).

	Reflectance (nm)							Continuum-removed (nm)						
25° optic	472	663	735	890	1476	1774	2156	516	648	835	1223	1490	1725	2139
Leaf clip	575	692	1064	1292	1433	1798	2297	531	706	778	1195	1545	1725	2347

Table 5. Most discriminative bands.
The bands are selected on the basis of the W-test done on the set of sixteen re-sampled spectra.

Concerning the classes' type, within the pure-spectrum classes *Erica arborea* has the highest average "net separability" while *Phillyrea latifolia* detains the lowest. All species but *Arbutus unedo* and *Phillyrea latifolia* retrieve greater net separability for the leaf scale of observation. All species, except *Erica arborea* retrieve greater net separability when using continuum-removed data instead of reflectance (Table A.1b, A.2b; Appendix A).
The mixed spectra perform accordingly to spectra from which they are originated. In the cases of high similarity of the classes' composition, it is observed that for 25° optic reflectance separability is generally low, while on the contrary continuum-removed and especially leaf scale enhance the separability (nr. 38, 56, 75; Table A.3a; Appendix A). The "net separability" of mixed spectra is evaluated only on the basis of the W-test rejections obtained with the complete set of sixteen classes. As mentioned above, the greater number of classes lessens significantly the net separability, i.e. the number of bands statistically different in all the possible pairings. For two of the mixed classes there are no such bands (classes 13, 14; table A.3b; Appendix A).

III.2 Spectral unmixing

Summary statistics are calculated on the estimations of the species fraction cover (Table B.1; Appendix B). The products of the image elaborations are fully appreciated in the two spatial dimensions. For simplicity, the results have been checked at the 169 locations corresponding to the sampling sites; the information derived at the 169 locations can eventually be extended to the rest of the image.

It emerges that continuum-removed data always retrieves an overall higher variability of estimations.

The estimations' variability depends mostly on which of the endmember collections is used. The estimations' variability is flattened by the inclusion of the mixed classes all generated on the basis of the *Quercus ilex* spectrum. The estimations' variability is indeed lowest for the library of ten mixed classes and highest for the library using only the six pure classes; this is verified for both data types. As the variability decreases, the estimations for *Quercus ilex* rise steadily.

Concerning reflectance data, the lowest variability of *Quercus ilex* estimations (0%) corresponds to the highest average estimation (63%), whereas the highest variability (10%) corresponds to the lowest average estimation (10%). *Quercus ilex* continuum-removed data has a similar situation.

For reflectance data, *Phillyrea latifolia* is recognized rarely. The species is not detected at any of the sites for the library of ten mixed reflectance classes. Higher estimations are obtained with the endmember collection including sixteen pure and mixed spectra, especially when continuum-removed data are used.

III.3 Accuracy assessment

RMSEs calculated out of the LSU results are noticeably different depending on which of the twelve input/training-set configurations is used. Furthermore, the per-species and the per-site procedures result also in slightly different RMSEs, though of a similar magnitude (Table C.3-4; Appendix C).

The use of mixed classes lessens the error. The lowest average RMSE (19%) is retrieved by the configuration using sixteen pure and mixed classes, the full-spectrum range and continuum-removed reflectance; calculated per-species. The determination of the best-performing configuration is not as univocal as that of the worst-performing one; it has to be pointed out that both the per-species and the per-site average RMSEs indicate the same set of four top-configurations with lowest RMSEs (19-20%).

The highest average RMSEs are retrieved by the configurations using six pure classes, the full spectral range and the original reflectance data (32%).

AVERAGE RMSE (%)

data type	library	spectrum range		PER-SPECIES	PER-SITE
Refl.	6	7	▶reflectance, 6 pure classes, spectrum subset	25	26
		126	▶reflectance, 6 pure classes, no spectrum subset	27	<u>32</u>
	10	7	▶reflectance, 10 pure classes, spectrum subset	20	20
		126	▶reflectance, 10 pure classes, no spectrum subset	20	20
	16	7	▶reflectance, 16 pure and mixed classes, spectrum subset	23	24
		126	▶reflectance, 16 pure and mixed classes, no spectrum subset	21	21
CR	6	7	▶continuum-removed, 6 pure classes, spectrum subset	25	26
		126	▶continuum-removed, 6 pure classes, no spectrum subset	21	22
	10	7	▶continuum-removed, 10 pure classes, spectrum subset	21	22
		126	▶continuum-removed, 10 pure classes, no spectrum subset	22	25
	16	7	▶continuum-removed, 16 pure and mixed classes, spectrum subset	20	20
		126	▶continuum-removed, 16 pure and mixed classes, no spectrum subset	<u>19</u>	20

Figure 7. Accuracy assessment. Per-species and per-site average RMSE.
The scheme recapitulates the results of the accuracy assessment. Minimum and maximum RMSEs are underlined. The lowest four average RMSEs are framed.
Average RMSEs is calculated on RMSEs displayed in Table C.3-4, Appendix C.

PER-SPECIES ASSESSMENT

The lowest average RMSE is 19% and is obtained by using the 16-classes library, the full spectral range and the continuum-removed data. Besides the same library and input data type retrieve comparable results (20%) when associated to the spectral subset of seven bands. Average RMSEs of 20% result also from both the configurations using reflectance data and an endmember collection of only ten mixed classes.
The highest per-species average RMSE, 27%, is generated by the LSU when reflectance full-spectrum range data are classified with only the six pure endmembers.

Concerning the average RMSE of the single species, *Quercus ilex* and *Phillyrea latifolia* are the two species with highest (37%) and lowest (7%) errors, respectively (Table C.3; Appendix C).

Per-species RMSEs have a significant variability depending on the configuration, i.e. depending on which combination of endmembers, spectral range, and data type is used (Figure 8-9).

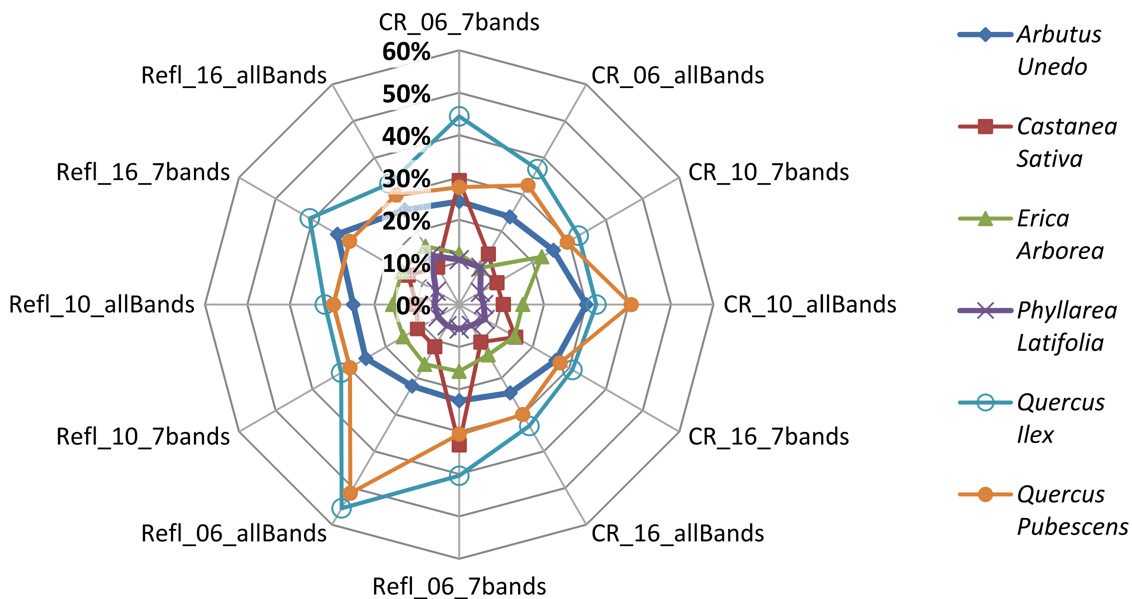


Figure 8. Average per-species RMSE.
The axis names abbreviate the configuration of data type (*Refl* = reflectance, *CR* = continuum-removed), spectral library (*06* = six pure-spectrum classes, *10* = ten mixed-spectrum classes , *16* = sixteen pure and mixed-spectrum classes), spectrum range (*7bands* = spectrum subset, *allBands* = full spectrum range).

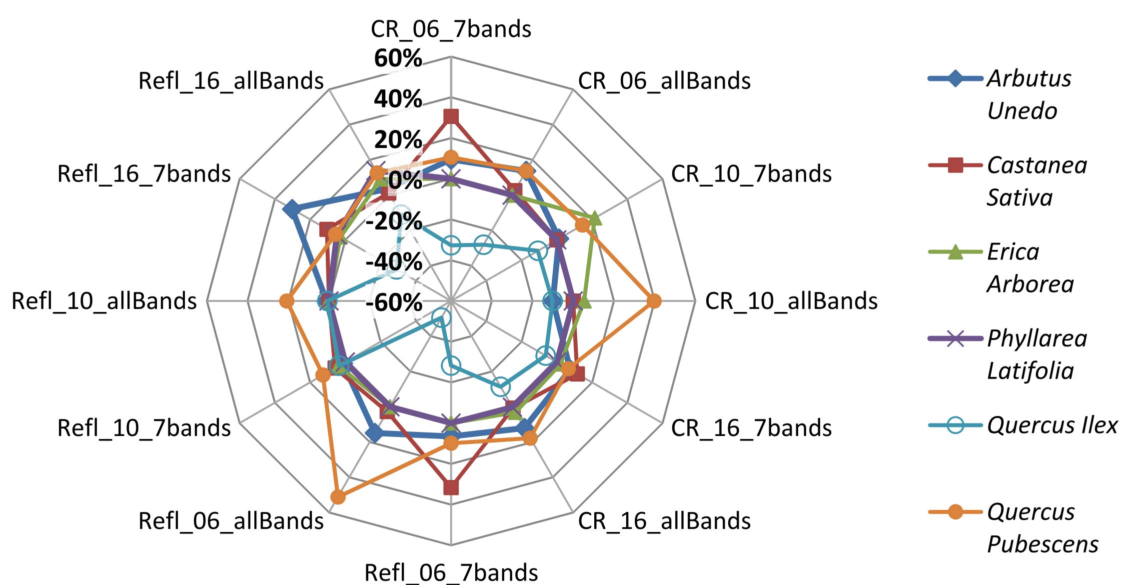


Figure 9. Median of the per-species raw errors.
The axis names abbreviate the configuration of data type (*Refl* = reflectance, *CR* = continuum-removed), spectral library (*06* = six pure-spectrum classes, *10* = ten mixed-spectrum classes , *16* = sixteen pure and mixed-spectrum classes), spectrum range (*7bands* = spectrum subset, *allBands* = full spectrum range).
The graph provides indication on the estimation and in particular shows whether a species tends to be over or under-estimated.

The radar graph of average RMSE shows that species tend to maintain their reciprocal magnitude in changed configurations; in the graph, shapes tend to be concentric. Looking at the median radar graph, it is possible to get an indication on whether the error is due to either to under- or overestimation.

Quercus ilex, the species with the highest average RMSE, is generally underestimated (negative median). *Quercus pubescens* on the contrary is generally overestimated.
The smaller RMSE of *Phillyrea latifolia* goes together with medians of the error (raw difference between estimation and observation) equal to zero.

Concerning the three different libraries, there is an overall reduction of the RMSE when the mixed spectra are used. This general result applies to *Castanea sativa* and *Quercus ilex*, but is denied in the case of *Arbutus unedo*.
For reflectance data, ten mixed classes retrieve lower RMSEs than sixteen pure and mixed classes. The continuum-removed data show an opposite tendency.

Concerning the spectral range, the use there is a balance between the two full-range and the subset. Nonetheless the per-species RMSE is more often higher for the spectrum subset.
Quercus pubescens shows a clear improvement with the spectrum subset.

Concerning the data type, continuum-removal is associated to the absolute lowest average RMSE. Nonetheless for 10 on 36 occurrences the two data types retrieve the same RMSE. Species-RMSE is lower for continuum-removed almost as many times as reflectance data (14/36 times against 12/36).

Correlations coefficients confirm a direct and positive relation between the occurrence and the variability of a species and its average RMSE. The average per-species RMSE is the average of the twelve configurations RMSEs (third to last column in Table C.3; Appendix C).

	Phillyrea latifolia	*Castanea sativa*	*Erica arborea*	*Arbutus unedo*	*Quercus pubescens*	*Quercus ilex*
average RMSE	7,3%	15,0%	15,2%	25,7%	32,3%	37,0%
● **speciesCount**	23	3	37	93	38	159
x **speciesStDev**	0,05	0,10	0,16	0,25	0,30	0,32

Table 6. RMSE correlation with species-specific factors.
"speciesCount" indicates the total occurrences of one species, i.e. the total number of plots in the validation dataset where a species is observed. "speciesStDev" is the standard deviation calculated on a species observations The table relates to the graph displayed in
Figure 10.

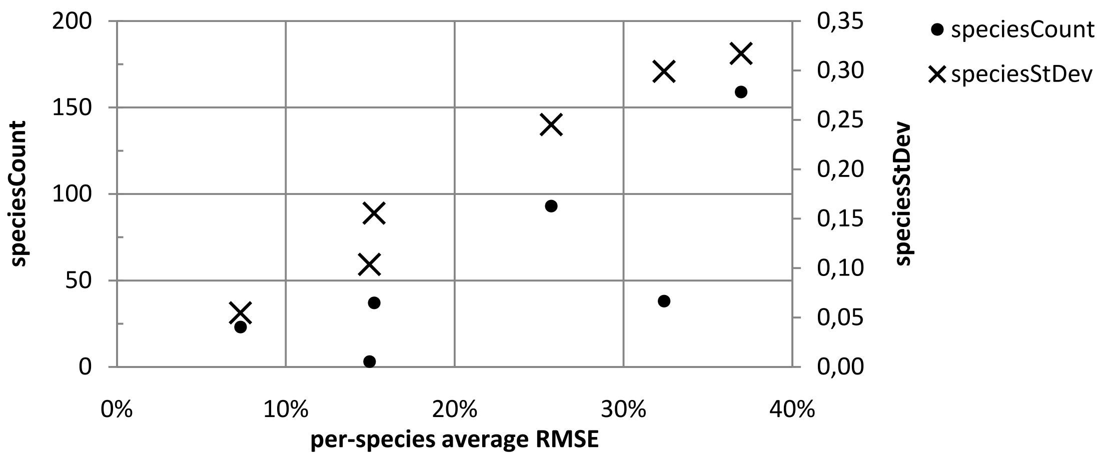


Figure 10. RMSE correlation with species-specific factors.
On the horizontal axis is the per-species average RMSE. The left vertical axis indicates the total occurrences of one species, i.e. the total number of plots in the validation dataset where a species is observed; the right vertical axis represents the standard deviation calculated on a species observations. Pearson's ρ coefficients are 0.75 and 0.98 for the first and the second data series, respectively. Refer to
Table 6 for details.

PER-SITE ASSESSMENT

Concerning the configurations, the lowest average RMSE is 20% (average RMSE of 169 site-RMSEs) is equally obtained by four different configurations (Table 4, Appendix C). These are the same four retrieving the lowest per-species average RMSEs (Table 3, Appendix C).
The highest per-site average RMSE is 32%; this value derives from a LSU which uses six pure classes, full-spectrum range and reflectance data, i.e. from the same configuration said to have the highest per-species average RMSE.

The performance of the single sites is assessed on the average site RMSE. The average per-site RMSE is the average of the twelve configurations RMSEs (third to last column in Table C.4; Appendix C). Values range between 12% and 43%.
The lowest average site RMSE (12%) refers to site C44; here a 1% RMSE is obtained with the 16-classes library, full-spectrum range, continuum-removed data. The canopy composition observed at sampling site C44 is Arbutus unedo 30%, *Quercus ilex* 50%, *Quercus pubescens* 20%.
The highest average per-site RMSE (43%) belongs to site B36, fully covered by *Castanea sativa*.

Spectral libraries with mixed classes have lower RMSEs than those with only pure classes.
The difference in performance between the 16-classes and the 10-classes library is less marked; it is noted that more than half of the sites have lower RMSE when the 16-classes library is used.
Concerning the spectral range and the data type, it results that more than half of the sites has RMSE lower for a spectral subset and/or continuum-removed data.

There is a negative correlation between the magnitude of the error (average per-site RMSE) and the heterogeneity of the canopy composition.

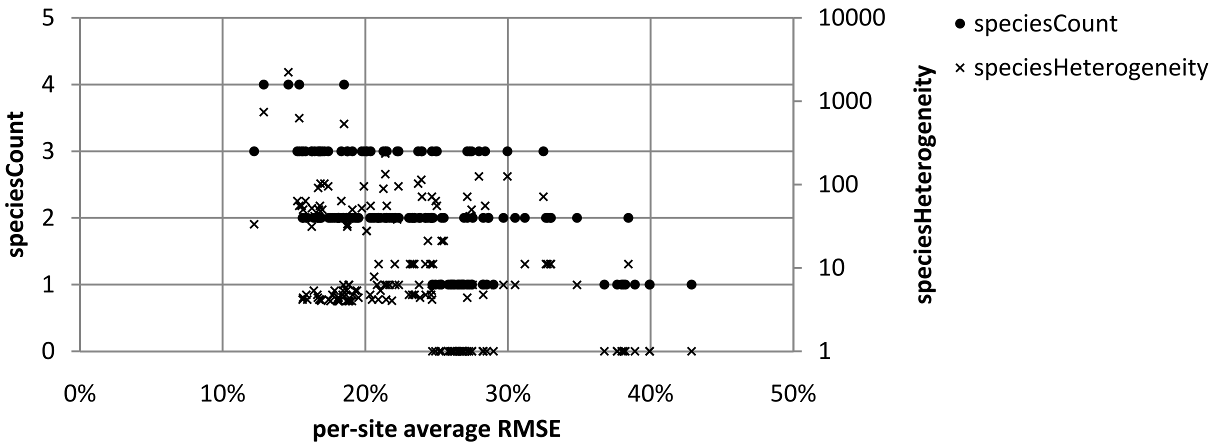


Figure 11. RMSE correlation with site-specific factors.
On the horizontal axis the per-site average RMSE. On the left vertical axis is the number of the species observed at the site. The right vertical axis (log10 scale) represents the “species heterogeneity” calculated as the inverse of the product of all non-zeros observations. Pearson's ρ coefficients are -0.54 and -0.21 for the first and the second data series, respectively.

Correlations to other factors, such as canopy height and the location of the site in the study area, have also been calculated. It results that they have no relation with the classification error.

OBSERVATIONS VS ESTIMATIONS

The per-site procedure checked the correspondence of the heterogeneity of estimations with that of the observations.

It emerges that there is a significant mismatch between the two rankings built on the measure of “species heterogeneity”. Furthermore, reflectance data has generally a smaller rank-RMSE. For both data types the lowest rank-RMSE is calculated on the sixteen-classes library, with spectrum subset in the case of continuum-removed data.

As mentioned above, the lowest average RMSE (19%) is obtained by using the 16-classes library, the full spectral range and the continuum-removed data. The results of this configuration are described for the single species in the graphs of Figure 12. Observed species composition (horizontal axis) is related to estimations derived from LSU outputs (vertical axis). Each point represents a sampling site.

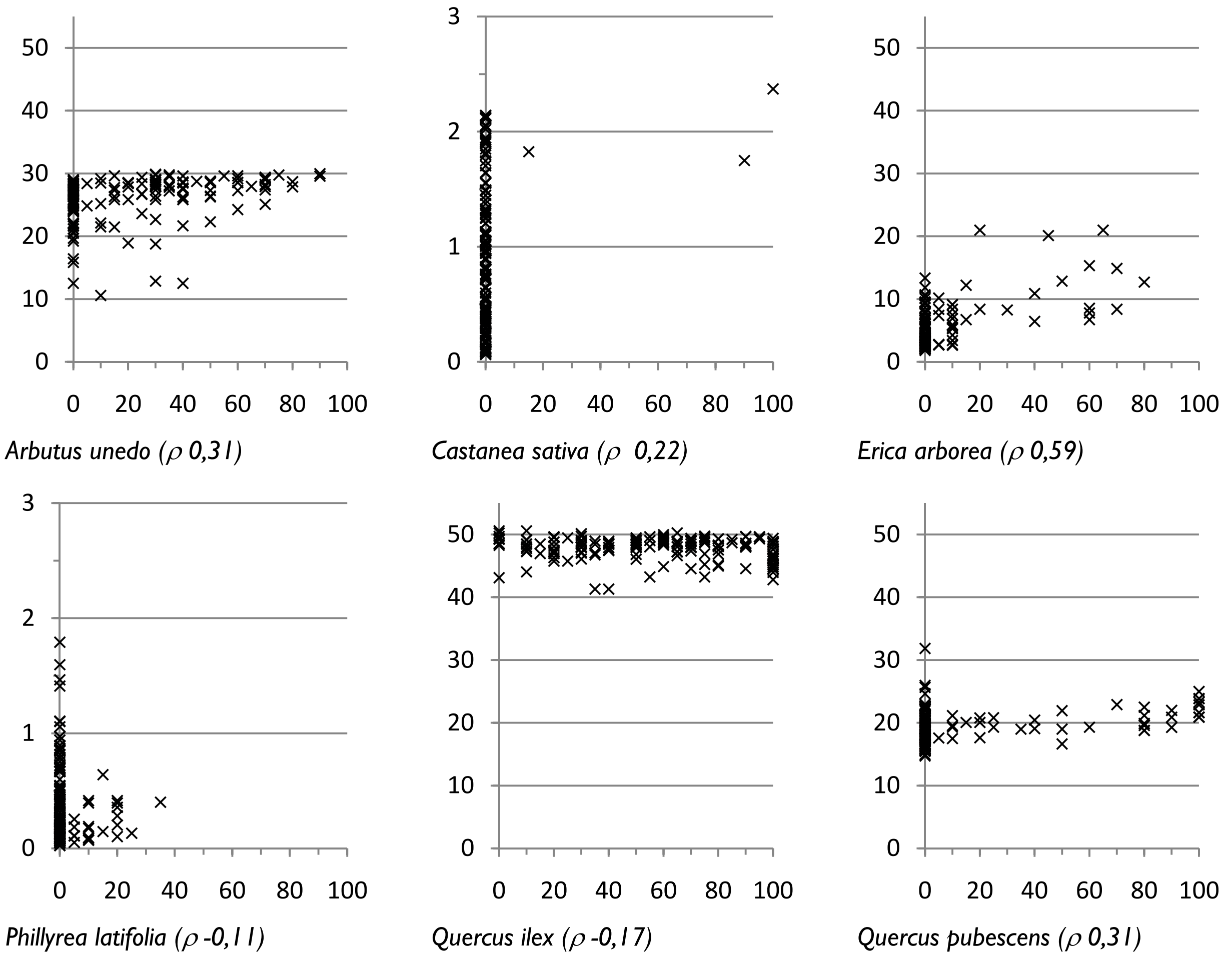


Figure 12. Scatter graphs of observed and estimated percentage cover.
On the horizontal axis is the observed percentage cover. On the vertical axis is the percentage cover as estimated by the top-performing configuration (sixteen-classes library, full spectrum range, continuum-removed).
Pearson's coefficients (ρ) are indicated in parenthesis.

Quercus ilex, *Arbutus unedo*, *Quercus pubescens* are, in this order, the species assigned the highest fraction cover.

Erica arborea has the highest positive correlation coefficient, while *Quercus ilex* the lowest negative.

Castanea sativa is observed only at three sites; low estimations are derived at all sites.

Phillyrea latifolia is observed at a discrete number of sites though in small percentages. On average, estimations are lower than *Castanea sativa*.

III.4 SPECIES MAPPING

The configuration chosen for the species composition mapping has the lowest average RMSE (19%) and is obtained by using the 16-classes library, the full spectral range and the continuum-removed data.

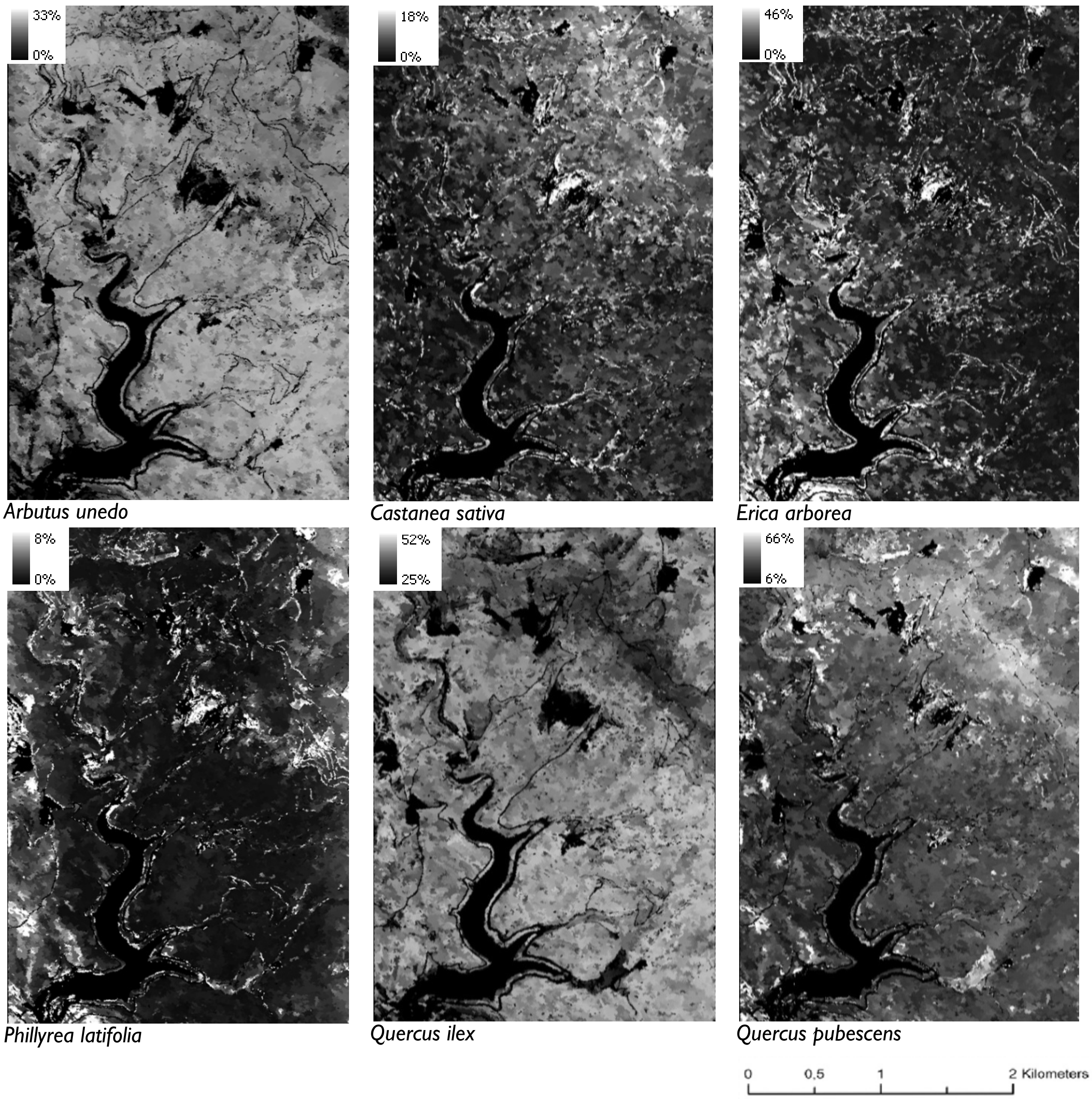


Figure 13. Species maps.
The six maps display the canopy composition in terms of percentage fraction cover of the species. The species composition of one location/pixel is given by its percentages in the six maps. The estimations are done on the basis of the spectral unmixing procedure retrieving the lowest average RMSE.

IV. DISCUSSION

IV.1 SPECTRAL SEPARABILITY

Spectral separability is highly dependent on the scale of observation and the data type.
Crown and leaf scale determine different results for the same class; in general terms, leaf clip data enhance the separability of spectra.
W-test results are significantly modified by continuum-removal; it is observed that, with few exceptions, classes of transformed data have a greater net separability.
As a result of these two circumstances, the net separability is usually highest for leaf scale, continuum-removed data.

Concerning the scale of observation, leaf spectra retrieve a much greater spectral separability than the crown scale. The use of the leaf clip probe ensures a much greater control in measurement's conditions. A constant, uniform illumination for all leaf samples minimizes the within-class variability which is much greater using the 25° optic. The reduction of the within-class variability means that the sum-of-ranks test is more likely to verify the separation between values of two classes.

Concerning the data type, the continuum-removal does not increase the total of the per-pairing rejections, while on the contrary it increases the "net separability" of a class. It can be observed on the b/w W-test grids, though not more abundant, are much more fragmented. This fragmentation results in a higher probability that separability occurs for all pairings.
Even if not univocal, these indications sustain the use of continuum-removed data in the successive spectral analyses. As single classes show an overall higher net separability, the transformation can potentially ease the image spectral unmixing.

While scale of observation and data type determine the quality of the results, the number of classes determines the quantity of the results and it is a relevant variable to the aim of the subsequent data reduction.
The highest recurrences of W-test positive rejections, i.e. the local maxima of the bar charts, identify the wavelengths where separability is greater. The identification is done on the results obtained using the set of sixteen pure and mixed classes. Though not broadening the set of species, in this way it is still possible to simulate the statistical analysis of a larger set of species, i.e. of a much larger number of pairings.
The identification of the spectral regions of highest separability is clear for the reflectance data and less for the continuum-removed data.

The results of the W-test depend on the specific set of spectra in use. The separability emerged from the analyses has not an absolute significance. The availability of more species and more samples per class would surely increase the reliability of the results.

In any case the results are consistent within the dataset. Spectral separability is verified among the classes here in use, whereas a comparison with other works is not direct.

The statistical analyses allow a characterization of the six dominant species in terms of reflectance. There is evidence that *Erica arborea* is well spectrally separable, while *Phillyrea latifolia* is not. Conclusions on the discriminability of the other species are not so straightforward depending much on the scale and the data type.

IV.2 IMAGE ANALYSIS

The variety of responses obtained by the twelve configurations can be considered a relevant outcome itself. The empirical method of iterating the work-flow proved to be a valid approach to assess the goodness of a given combination of data type, spectral subset, endmembers collection.

PURE VS MIXED CLASSES

The poor accuracy associated with pure classes might be taken as a suggestion to explore other approaches.

The six pure-spectrum endmembers are generated as an average of the class samples, which in turn are averages of ASD FieldSpec FR measurements. The average spectrum is assumed to represent more consistently the reflectance properties characteristic of a species.

In recent studies, image species classification resulted more successful when the training set is built on image-spectra from sun-lit portions of the canopy (Clarks et al, 2005; Lucas et al, 2008; Youngentob, 2011). Though data and methods are different from ours, those conclusions may apply also here. If not by averaging all the samples, the class could be built on the average of the most intense ones.

In one case the pure-spectrum endmembers are associated with lower RMSE. In about twenty sites in strata A and B the lowest RMSE corresponds to the configuration using six full-spectrum pure classes, reflectance data. This contrasts with almost all other 140 sites where the same configuration retrieves the highest RMSE. The configuration is the one retrieving the highest estimations for *Quercus pubescens* and the lowest for *Quercus ilex*. In the particular case of these twenty sites, the error is minimized since *Quercus pubescens* observed fraction cover is more than 50%, the rest being *Quercus ilex*.

The smaller errors obtained with mixed classes depend on how they are defined. The mixed classes are created on the basis of the ground observations: only species associations and only

relative abundances that have been effectively found in the field were considered. All observations were used to determine which mixed classes were to be generated as well as to assess the classification accuracy.
The quality of the training set is limited so to more closely resemble the validation set. This of course reduces the error probability, a reduction that is achieved by appealing to subsidiary knowledge, i.e. the field data.

The apparent excellent accuracy related to *Phillyrea latifolia* depends on a combination of facts:

- the species is systematically under-estimated;
- the species is in effect rarely observed (23/169 sites);
- if present, it occupies always a small fraction of the canopy (average of 15%, maximum 35%).

The systematic under-estimation depends on the characteristics of the species spectrum and on the endmembers in use. Firstly, the spectral separability showed that *Phillyrea latifolia* has generally a poor discriminability with *Quercus ilex*, whose estimation is greatly advantaged by the mixed classes here in use. Moreover, the species has a small weight in the unique mixed class (*Quercus ilex* 75%, *Phillyrea latifolia* 25%) which prevents to retrieve estimations higher than 25%. In result of these circumstances the difference between the estimations and observations are small, thus explaining the low RMSE of *Phillyrea latifolia*.

TRANSFORMATION AND REDUCTION OF THE DATA

Generally the average RMSEs are lower for continuum-removed than reflectance data; the lowest average RMSE is indeed obtained with continuum-removed data. The outcomes of the image unmixing parallel that of the spectral separability. The statistical analyses indicate that the discriminability among classes is enhanced by continuum-removal. This result well compare with other studies (Schmidt & Skidmore, 2003; Youngentob, 2011).

A reduction of the data dimensionality is here achieved by selecting a subset of seven bands on the basis of the W-test. The accuracy assessment shows that the reduced dataset does not outclass the full-spectrum one; the two are essentially the same in terms of estimation error (Figure 7).
The relative balance between the spectrum ranges actually provides some interesting evidence. It indicates that, even after drastically reducing the computational effort, the accuracy of the estimations maintains comparable. This might turn important when dimensionality represented a serious limitation. Further evidence could be searched by experimenting endmembers composed of series of samples, instead of a single average spectrum.

CORRELATIONS

There is a clear correlation between the overall accuracy and *Quercus ilex* cover. The species is observed in 159/169 plot and in two third of them with 50% or more canopy cover. It is easily understood that the method which estimates the species more often and more abundantly is the one achieving the smallest average RMSE. The spectral unmixing outputs show that the use of mixed classes (all generated with *Quercus ilex* spectrum) results in more generous estimation of this dominant species. Configurations adopting mixed classes are in effect those with lowest error.

The more frequently a species is observed the higher the RMSE; moreover higher variability of the species' observations corresponds to a higher error. This can be explained by considering the functioning of the unmixing in ENVI. The range of the estimations is sensibly narrower than that the observations; the LSU indeed tends to retrieve a rather constant species estimation across the whole area. In the case the species' observations are very varied this consequently increases the species error.

Concerning the effect of site-specific factors on the error, there is a negative correlation between RMSE and the heterogeneity of the canopy composition. ENVI LSU tends systematically to produce estimations for all the six species; per-site estimations are generally heterogeneous and never monospecific (100% cover). It necessary results that the site observations better fitting these general circumstances are those with more species. Indeed the lowest site-RMSE is calculated on a species composition of 30%, 50%, 20%, whereas the highest belongs to a site with 100% cover.

OBSERVATIONS VS ESTIMATIONS

Three conclusions are drawn on the heterogeneity of observations and of estimations:

- there is an overall poor match;
- the mismatch is less for reflectance data;
- the closest convergence is reached with the endmember collection including sixteen pure and mixed classes.

The motivation of this mismatch could be related to scale issues or positional uncertainty. A plot observation might not effectively be representative of the composition of the surrounding area, i.e. of the composition that in the end will produce the average spectrum of the segmentation object. That means that even if a site is observed to have a high "species heterogeneity", this might not be preserved by the object. As estimations are calculated on the object-based imagery, their heterogeneity is representative of the larger forest area including the sampling site, not only of the specific site.

Graphs in the previous chapter give insight in the quality of the validation dataset (Figure 12). Species are not equally represented in the field, but most interestingly this does not disrupt the accuracy. Indeed the presence verified on the ground does correspond to the overall estimation of the species. In other words, species rarely observed (*Castanea sativa*, *Phillyrea* *latifolia*) have very low estimations, while frequent/abundant observations correspond to high estimations (*Quercus ilex*, *Arbutus unedo*).

Castanea sativa is recurrent in the study area and that is why it has been included in the spectral library. The species is almost invisible in the sampling scheme. The spectral unmixing does actually detect constantly *Castanea sativa*, but only in very small proportions. It is noteworthy that the highest estimation (2.37%) does correspond to the highest observation (100%).

The case of *Phillyrea latifolia* is more cumbersome. It is regularly encountered in the study area and has a fairly good representation in the validation dataset. Differently from *Castanea sativa*, the maximum estimation and observation do not coincide. The unmixing seems indeed to malfunction (ρ -0.17) which might be explained by the poor discriminability with the overabundant *Quercus ilex*. There could be yet another explanation related to the characteristics of this species, i.e. its average crown-size.
In site observations *Phillyrea latifolia* has never above 35% canopy cover; those observed fractions represent in most of the cases also the total extent of the crowns. In the study area plants of *Phillyrea latifolia* have an elongated structure; stems with few branches reach the highest canopy with compact leafy tips. Though small patches of *Phillyrea latifolia* are encountered, the species is never dominant, always mixed with others.
As a result of the small crown-size, *Phillyrea latifolia* is likely to have a minor weight in the object spectrum; even if the observed fraction is great its influence on the object spectrum may be insignificant. This in turn could explain the revealed misclassification.
For the same arguments, the extensive crowns of *Castanea sativa* could instead explain why observations and estimations converge in localizing the highest cover.

The graphs also show that the estimations tend to a constant so that ρ is generally close to zero; the same conclusion emerged from the standard deviations calculated on the estimations. *Quercus ilex*, for instance, has roughly a constant estimation of 48%. This can be attributed to the ENVI classifier and to the segmentation.
The spectral unmixing has to work on very subtle differences among spectra. Endmembers are all generated from crown average spectra whose spectral separability is not always verified. Due to use of the mixed classes, the classifier has to operate on even more limited margins.
Moreover the occurrence of disparate estimations is restrained by the object-based analysis. Spectral unmixing is performed on averaged pixel-spectra, the objects, which are more likely to be similar; sites with different observed composition not infrequently ended up in the same object. These circumstances would explain why sites with different canopy composition result in similar or equal estimations.
Constant estimations are not a failure of the unmixing; rather on the contrary they can be (in part) recognized as a result of the object-based analysis. Classes are unmixed on object-scale and not on pixel-scale; so the estimations have to be considered species composition at the

neighborhood and not at the site level. Well, such "neighborhood-based" species composition coming out of the image analysis fits the field experience. The three main species are correctly pointed out (*Quercus ilex*, *Arbutus unedo*, *Quercus pubescens*), in their correct hierarchical order in terms of abundance in the canopies.

A last comment on the graphs concerns *Erica arborea*.
Among the six species *Erica arborea* has the best convergence between the site observations and estimations. This can be taken as an indication of the good discriminability of its spectral signature. In turn the good spectral discriminability can be attributed to a combination of species-specific and environmental factors.
Earlier in this report, it has been noted that its crown reflectance has good spectral separability; the leaves size and arrangement, the crown structure are a peculiarity of this species. When it comes to image classification, these characteristics are likely to play in favor of the correct estimation of *Erica arborea*.
The presence of the species tends to be correlated to surface conditions that enhance the reflectance from the surface. The species thrives in low canopy stands, whereas it suffers the competition of other tree species in tall forest. *Erica arborea* is easily found where the forest opens. Openings of the canopy may occur along roads and trails, at the margins of fields, in logged plots. Due to the proximity of open spaces and the thinner vegetation cover, the reflectance of *Erica arborea*–rich areas is affected by the background more significantly than the closed forest.
These environmental factors are paired uniquely with *Erica arborea*. The environmental factors complement the species proper spectral signature. As a result they do not false, but on the contrary they support the better estimations of *Erica arborea*.

SPECIES MAPPING
The maps produced with the best performing model can be regarded as the final products of this work. The results are encouraging. The situation depicted by the maps resembles the one observed in the field with main species *Quercus ilex, Arbutus unedo*, *Quercus pubescens* followed by the others.

The empirical approach of iterating the image analysis with different configurations allows searching for the method that minimizes the error at the sampling sites. The sites are in effect our only proofs on the unmixing performance; their representativeness of the situation on the ground is therefore crucial. Here it is ensured by the reasonable amount of sampling sites.
Even if the perfect correspondence of estimations and observations at the sites is not reached, the average accuracy is assumed to be extendible to the whole image. The method minimizing the error at the validation sites is the one that retrieves the more reliable information also for the rest of the image.

V. CONCLUSION

The thesis undertook the study of Mediterranean forests by the means of imaging spectrometry. It investigates the possibilities to use imaging spectrometry for species identification and species mapping of Mediterranean vegetation. It consists of a sequence of analyses of hyperspectral data both from remote sensing and field measurements. The study concerns a rural area in Southern France.

Counting on the experience of previous research projects, fieldwork was carried out between September and October 2011.

Ground observations acquired information on forest vegetation, focusing on six arboreal species. The presence of these species and their relative abundance in the canopy were assessed and registered at 169 sample sites distributed over the study area.

A field spectrometer was used to measure crown and leaf reflectance. The crown scale serves to generate the endmembers for the image analysis, while the leaf spectra are a solid term of comparison in the spectral separability.

A series of preparatory phases precedes the data analysis. In particular, the spectrometer measurements require the removal of noise, the adaptation of the spectral resolution to that of the aerial dataset, the generation of mixed spectra.

Vegetation species can be separated by the means of the reflectance spectrum. Discriminability is higher among leaf spectra than among crown spectra. Compared to leaves, crowns have higher spectral separability in the visible spectrum and in the short-wavelength infrared.

The continuum-removal transformation generally determines the increase of the overall spectral separability.

The statistic tests are a pre-requisite to isolate the most distinctive bands across the whole spectrum range. Seven of these bands (Table 5) are those forming the spectrum subset used in the image spectral unmixing. In this way the spectra separability analysis supports the image analysis.

An optimization of the procedure is realized by iterating the mixture analysis according to twelve different input configurations. In this way some significant conclusions emerge from the comparison of the results.

The thesis sustains the use of endmembers generated as a mixture of pure crown spectra measured in the field. The reduction of the error is achieved by limiting the endmembers to a set of classes more similar to the observations.

There are indications that continuum-removal has a positive effect in reducing the error of the estimated species fraction cover. Nonetheless there is not a dramatic improvement.

Full-range and subset spectra retrieve essentially similar accuracies. An improvement in the identification of the most distinctive bands is expected to refine the conclusions on the appropriateness of adopting a spectrum subset.
Quantitative and qualitative aspects of the vegetation cover influence the classification accuracy. Classification accuracy is correlated with canopy species heterogeneity. Furthermore one species, *Quercus ilex*, detains a determinant weight in regulating the error.

The utility of species mapping motivates this research.
Information about the quality of the vegetation can be supplied to research on carbon fluxes and the water cycle. Forest species composition is an advantage in biodiversity and wildlife studies, natural resource management, and environment protection.
Species may not distribute randomly in the area. They may instead respond to soil chemistry, water availability, or reflect the history of the forest management. Mapping their occurrence and abundance can in the end serve to trace other landscape phenomena.
The final outcomes of the thesis could represent the starting point of other studies.

Automation is sought to improve the work efficiency. Programming is used to perform a part of the elaborations. The good realization of this task is considered among the achievements of the thesis.
The thesis does not benefit from the dependency on different software. Further integration of the work phases into more automated blocks is required. This could effectively optimize time. Automation would allow enlarging the number of solutions by considering other classes, different methods for image analysis and so forth. This would expand the basis for the thesis conclusions.
This report describes a path directed to a set of research questions; all choices taken by the path are not neutral: a selection is realized on the available dataset, which is rich in quality and quantity; the research touches diverse topics adapting to its goals a selection of methods from a series of disciplines such as statistics and image processing. Actually, along the study, not few options have been excluded for they were not useful to the economy of the work. The analyses here presented do not exclude possible alternatives or improvements.

General objective	Species identification in hyperspectral data for the description of the vegetation composition on a landscape scale.
Study area	Middle catchment of the River Peyne, in southern France.
Techniques	Field and remote sensing spectrometry. Field sampling of vegetation.

	Questions	Activities	Results
1	Are the dominant vegetation species of the study area spectrally separable?	▶ Separability among species field spectra is investigated by the means of statistic analyses (Wilcoxon test).	▶ Species discrimination is verified both at the leaf and, most interestingly, also at the crown scale.
2	Is it possible to derive the species composition of Mediterranean forest from remote sensing imagery using endmember generated from field spectra?	▶ Image linear spectral unmixing; retrieval of species estimations; validation against the site observations.	▶ The classification of a HyMap dataset using endmembers generated form ASD FieldSpec Fr measurements retains some criticality.
3	Does mixture analysis improve if it uses composite endmembers generated as a combination of (pure) field spectra?	▶ Building of mixed-spectra spectral library; iteration of image spectral unmixing using pure and mixed reference spectra; accuracy assessment.	▶ RMSE is effectively diminished by employing mixed-spectra classes generated on the basis of the information collected in-situ.
4	Do adjustments such as continuum-removal and/or data resizing improve the species composition mapping?	▶ Iteration of image spectral unmixing using different inputs.	▶ The improvement is fair. It verifies more clearly for continuum-removal than the spectrum subset.
5	Do quantitative or qualitative aspects of the vegetation cover influence the classification accuracy?	▶ Analysis of the correlation between classification accuracy and the information on the vegetation cover collected in the field.	▶ Due to the functioning of the spectral unmixing, higher heterogeneity of an observation usually results in a lower RMSE.

Table 7. Summary table of the thesis’ contents.

VII. REFERENCES

- ASD (1999), Analytical Spectral Devices, Inc. (ASD) Technical Guide 4th Ed. Boulder.
- BLASCHKE T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 2-16.
- BRADBURY, D. E. (1981). The Physical Geography of the Mediterranean lands. In: F. Di Castri, D. W. Goodall, R. L. Specht (Eds.), *Ecosystems of the world II - Mediterranean-type shrublands* (pp. 53-62). Elsevier.
- CLARK, M. L., D. A. ROBERTS, D. CLARK (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing for Environment* 96, 375-398.
- DEBAZAC, E. F. (1981). Temperate broad-leaved evergreen forests of the Mediterranean region and Middle East. In: J. D. Ovington (Ed.), *Ecosystems of the World 10 – Temperate Broad-leaved evergreen forests* (pp. 107-123). Elsevier.
- DE JONG, S. M., V. G. JETTEN (2007). Estimating spatial patterns of rainfall interception from remotely sensed vegetation indices and spectral mixture analysis. *International Journal of Geographical Information Science* 21, 529–545.
- DI CASTRI, F. (1981). Mediterranean-type shrublands of the world. In: F. Di Castri, D. W. Goodall, R. L. Specht (Eds.), *Ecosystems of the world II - Mediterranean-type shrublands* (pp. 1-52). Elsevier.
- GIBBONS, J. D. (1971). Nonparametric Statistical Inference. McGraw-Hill.
- GONG, P. R. PU, B. YU (1997). Conifer Species Recognition: An Exploratory Analysis of In Situ Hyperspectral Data. *Remote Sensing for Environment* 62, 189-200.
- GHIYAMAT, A., H. Z. M. SHAFRI (2010). A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing* 31, 1837–1856.
- HYVISTA (2012), HyVista Corporation website. http://www.hyvista.com (viewed on January 9, 2012).
- KUMAR, L., K. SCHMIDT, S. DURY, A. SKIDMORE (2002). Imaging spectrometry and vegetation science. In: F. D. van Der Meer, S. M. de Jong (Eds.), *Imaging spectrometry - Basic Principles and Prospective Applications* (pp. 111-155). Springer.
- LEUSCHNER C. (2005). Vegetation and ecosystems. In: E. van der Maarel (Ed.), *Vegetation ecology* (pp. 85-105). Blackwell.
- LILLESAND, T. M., R. W. KIEFER, J. W. CHIPMAN (2004). Remote sensing and image interpretation. Wiley.
- LUCAS R., P. BUNTING, M. PATERSON, L. CHISHOLM (2008). Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sensing for Environment* 112, 2088-2103.
- MANEVSKI, K., I. MANAKOS, G. P. PETROPOULOS, C. KALAITZIDIS (2011). Discrimination of commom Mediterranean plant species using field spectroradiometry. *International Journal of Applied Earth Observation and Geoinformation* 13, 922-933.
- MC COY, R. M. (2004). Field methods in remote sensing. New York.
- NAHAL, I. (1981). The Mediterranean climate from a biological viewpoint. In: F. Di Castri, D. W. Goodall, R. L. Specht (Eds.), *Ecosystems of the world II - Mediterranean-type shrublands* (pp. 63-86). Elsevier.

- 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 for Environment* 113, 771-780.
- PU, R. (2009). Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing* 30 (11), 2759-2779.
- RICHTER, R. & D. SCHLÄPFER (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part II: atmospheric/topographic correction. *International Journal of Remote Sensing* 23, 2631−2649.
- SCHLÄPFER, D., & R. RICHTER (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part I: parametric orthorectification. *International Journal of Remote Sensing* 23, 2609−2630.
- SCHMIDT, K.S., 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., A. K. SKIDMORE (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment* 85, 92-108.
- SLUITER, R. (2005). Mediterranean land cover change - Modelling and monitoring natural vegetation using GIS and remote sensing. *Netherlands Geographical Studies* 333. Utrecht.
- STRAHLER A. H. (1986). On the Nature of Models of Remote Sensing. *Remote Sensing of Environment* 20, 121-139.
- THOMAS, V., P. TREITZ, D. JELINSKI, J. MILLER, P. LAFLEUR, J. H. MCCAUGHEY (2002). Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment* 84, 83-99.
- TOMASELLI, R. (1981). Main physiognomic types and geographic distribution of shrub systems related to Mediterranean climates. In: F. Di Castri, D. W. Goodall, R. L. Specht (Eds.), *Ecosystems of the world II - Mediterranean-type shrublands* (pp. 95-106). Elsevier.
- VAN AARDT, J. A. N., R. H. WYNNE (2001). Spectral Separability among Six Southern Tree Species. *Photogrammetric Engineering & Remote Sensing* 67 (12), 1367-1375.
- YOUNGENTOB, K. N., D. A. ROBERTS, A. A. HELD, P. E. DENNISON, X. JIA, D. B. LINDENMAYER (2011). Mapping *Eucalyptus* subgenera using multiple endmember spectral mixture analysis and continuum-removed imaging spectrometry data. *Remote Sensing of Environment* 115, 1115-1128.
- ZOMER R. J., A. TRABUCCO, S.L. USTIN (2009). Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. *Remote Sensing of Environment* 90, 2170-2177.

VII. APPENDICES

The Appendices group relevant contents in graphic, tabular and cartographic layout.
Appendix A concerns the statistical separability of crown and leaf spectra; the outputs of Wilcoxon sum-of-ranks test are displayed in graphics and tables.
Appendix B concerns the image spectral mixture analysis. Results are summarized through statistics calculated at image locations corresponding to the sampling sites.
Appendix C pertains to the accuracy assessment.
Appendix D collects cartography concerning the study area and the fieldwork.

In the frame below, the short forms used in the Appendices.

aru	*Arbutus unedo*
cas	*Castanea sativa*
era	*Erica arborea*
phl	*Phillyrea latifolia*
qil	*Quercus ilex*
qpu	*Quercus pubescens*

> the same abbreviations combined with numbers indicating the percentages identify the mixed classes

classes	number of classes used for the W-test. **6** stands for the set of six pure classes; **16** stands for the collection of sixteen pure and mixed-spectra classes.
data type	indicates whether the **reflectance** or **continuum-removed** data are used.
library	spectral library used for spectral unmixing. The numbers **6**, **10**, **16** indicate the endmembers collection of six pure-spectra classes, ten mixed-spectra classes, sixteen pure and mixed-spectra classes, respectively.
resolution	spectral resolution/number of bands: original ASD FieldSpec FR resolution (**ASD**), re-sampled to the imagery specifications (**HyMap**).
scale	scale of observation indicated with the optical configuration adopted for the ASD FieldSpec FR measurements: **25° optic** stands for crown scale, **leaf clip** for leaf scale.
spectrum range	numbers **7** and **126** specify whether inputs are subset or full-range data, respectively.

APPENDIX A – STATISTICS FOR SPECTRAL SEPARABILITY

Graphs and tables refer to the Wilcoxon sum-of-ranks tests carried out on the field spectra. The test outputs 1/0 results where 1 = rejection of the null hypothesis = spectra are significantly different.

Graphs *a* have wavelength (nm) on the horizontal axis. The bar chart displays the number of rejections totalized by each band (left vertical axis). A reflectance spectrum of *Quercus ilex* is overlaid as a reference (right vertical axis).
In graphs *b* the horizontal axis represents wavelength (nm) while pairings stay on the vertical axis. White shading stands for rejection, i.e. significant separability. Graphs using 16 classes omit the parings' names; the numeric reference on the vertical axis relates to that of Table A.3a.
Graphs *c* display the "net-separability"; wavelength (nm) and classes stay on the horizontal and vertical axis, respectively. Markers symbolize the bands where separability occurs for all possible pairings. Below the graph, the number indicates the total amount of such bands.

Tables *a* concern the test pairings; the displayed values are sums of test rejections.
Tables *b* concern the "net-separability" of the classes. The numbers represent the amounts of bands where rejection occurs for all possible pairings of a class.
The values in the tables refer to six spectrum ranges. The frame below displays the total number of bands for each of the ranges, for the two scales of observation (in italics the values for re-sampled spectral resolution).

SPECTRUM RANGE			TOTAL NUMBER OF BANDS			
			25° OPTIC		LEAF CLIP	
whole spectrum	350-2500 (nm)	**full**	1817	*114*	2052	*126*
visible	350-700	**VIS**	350	*17*	252	*17*
red-edge	690-720	**red**	31	*3*	31	*3*
near-infrared	700-1370	**NIR**	670	*45*	600	*42*
shortwave-infrared	1400-1800	**SWIR1**	411	*31*	628	*35*
shortwave-infrared	1940-2500	**SWIR2**	386	*21*	572	*32*

Fig. A.1

Scale **25° optic**
Data type **reflectance**
Resolution **ASD**
Classes **6**

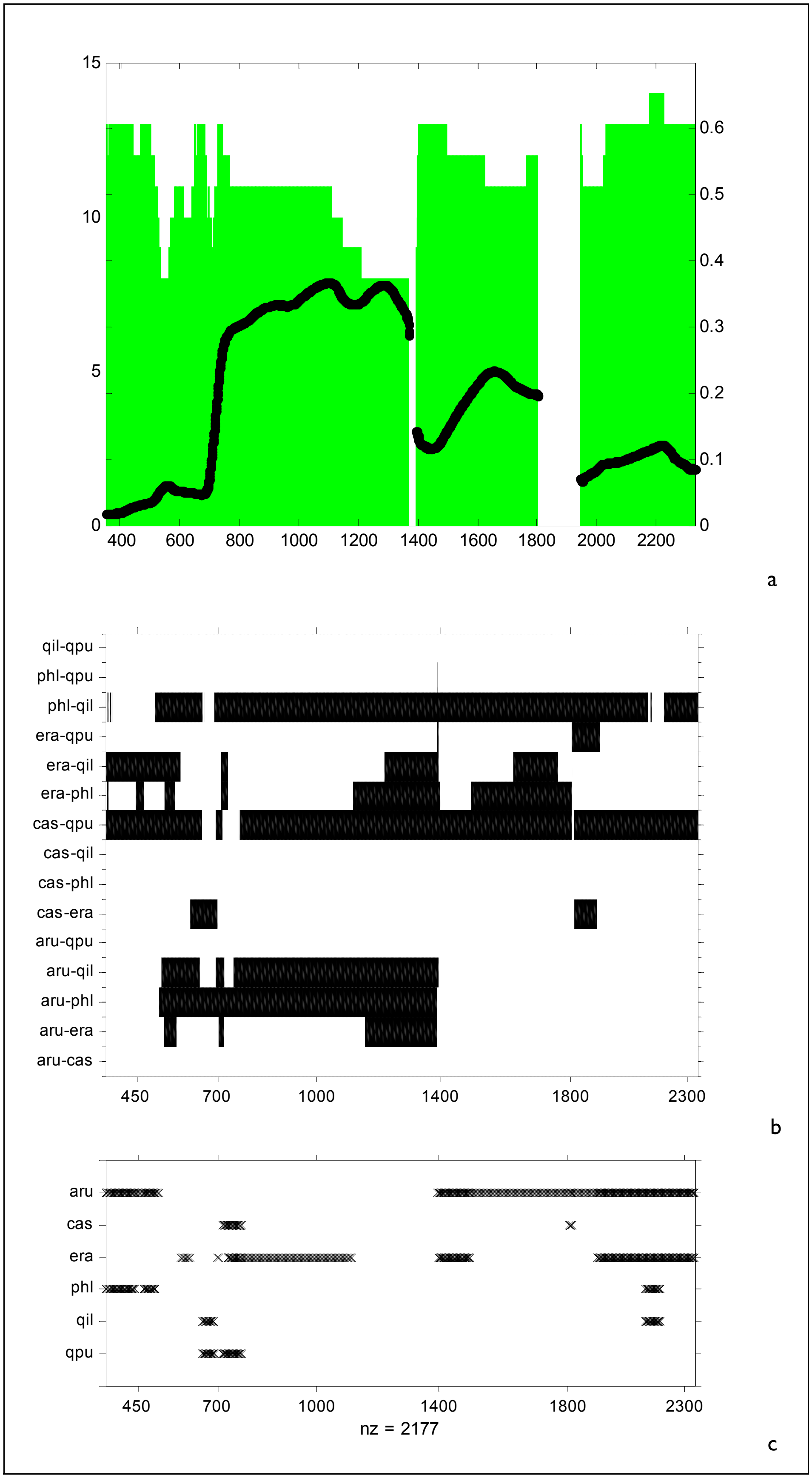


FIG. A.2

Scale **25° optic**
Data type **reflectance**
Resolution **HyMap**
Classes **6**

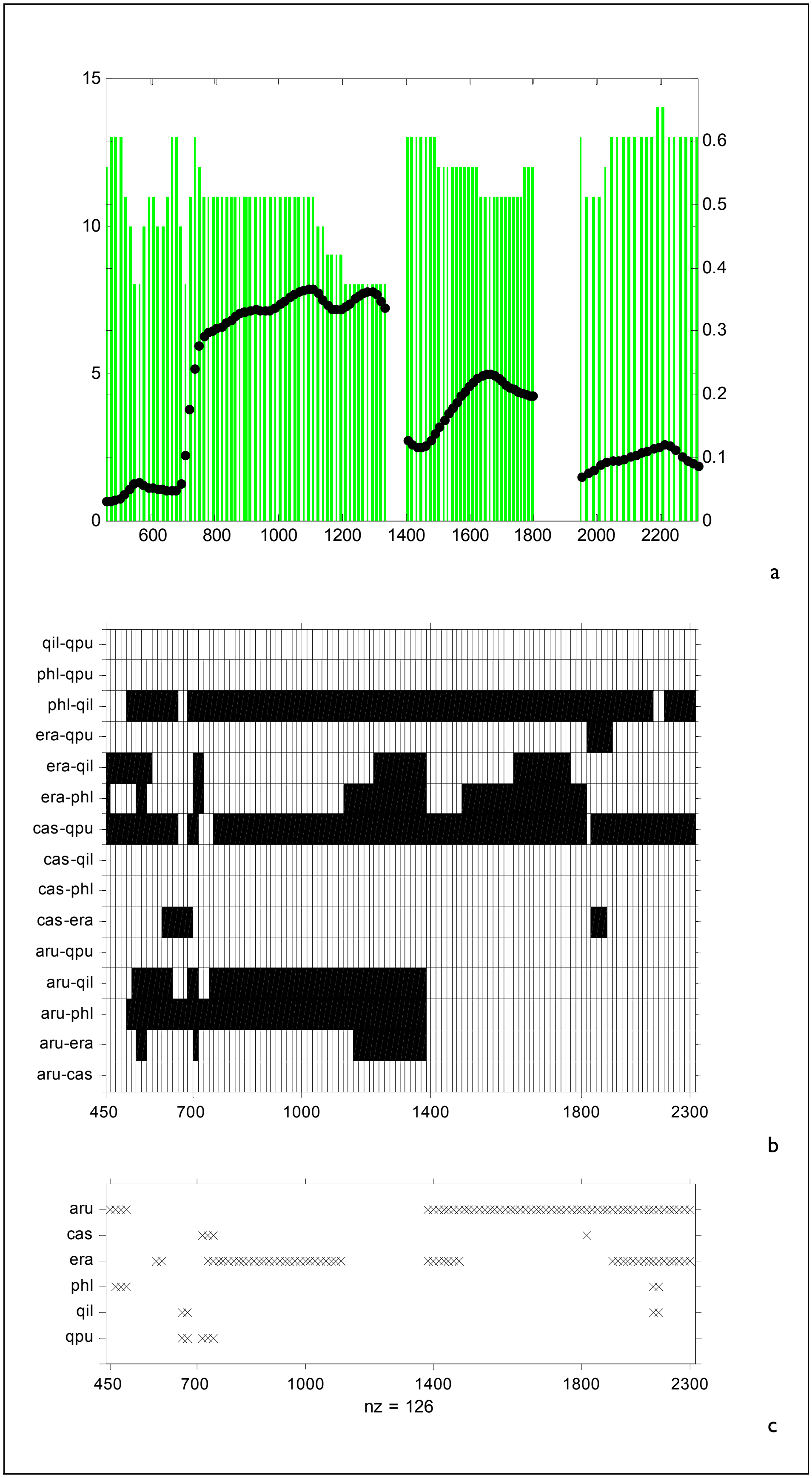


Fig. A.3

Scale **25°optic**
Data type **reflectance**
Resolution **HyMap**
Classes **16**

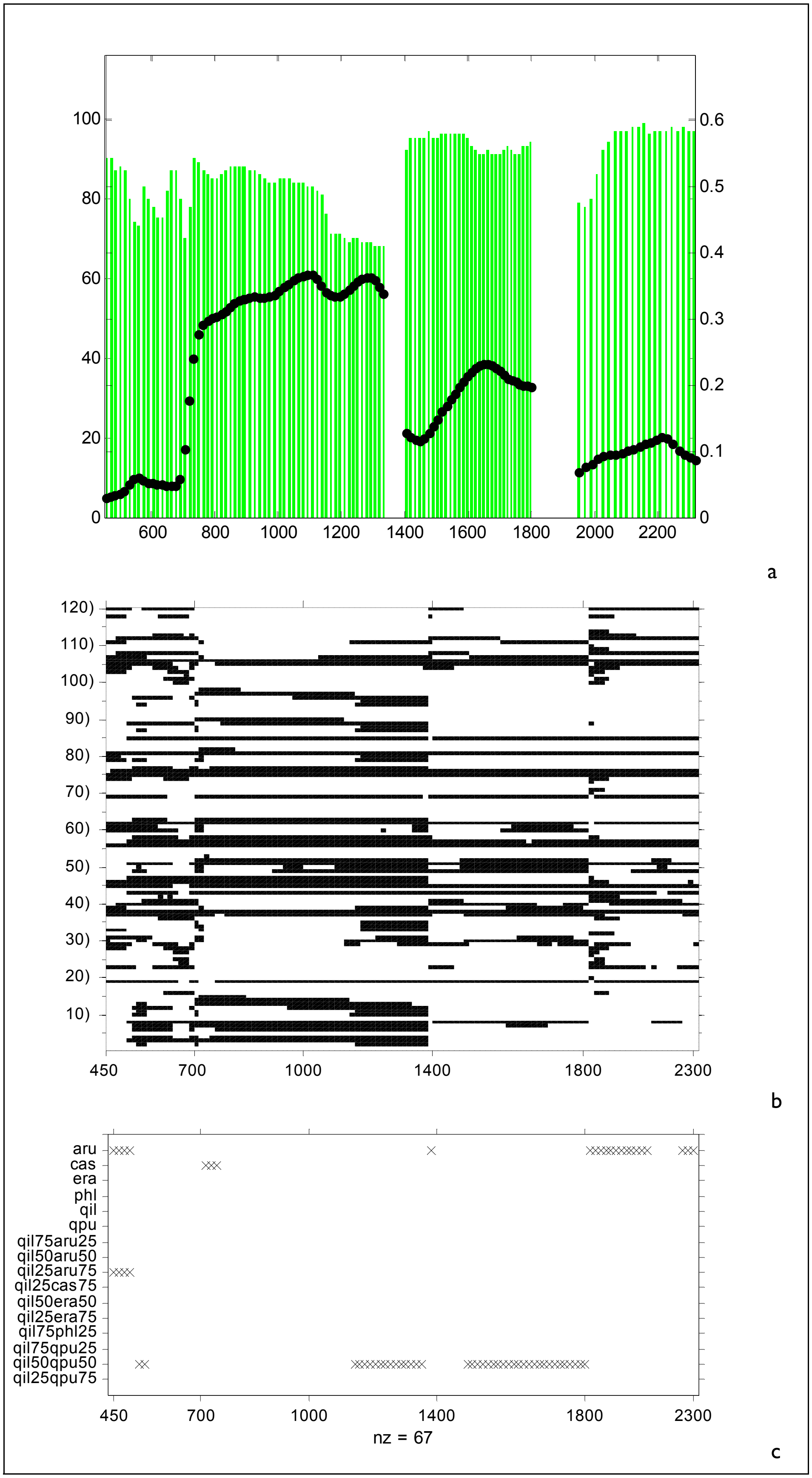


Fig. A.4

Scale **leaf clip**
Data type **reflectance**
Resolution **ASD**
Classes **6**

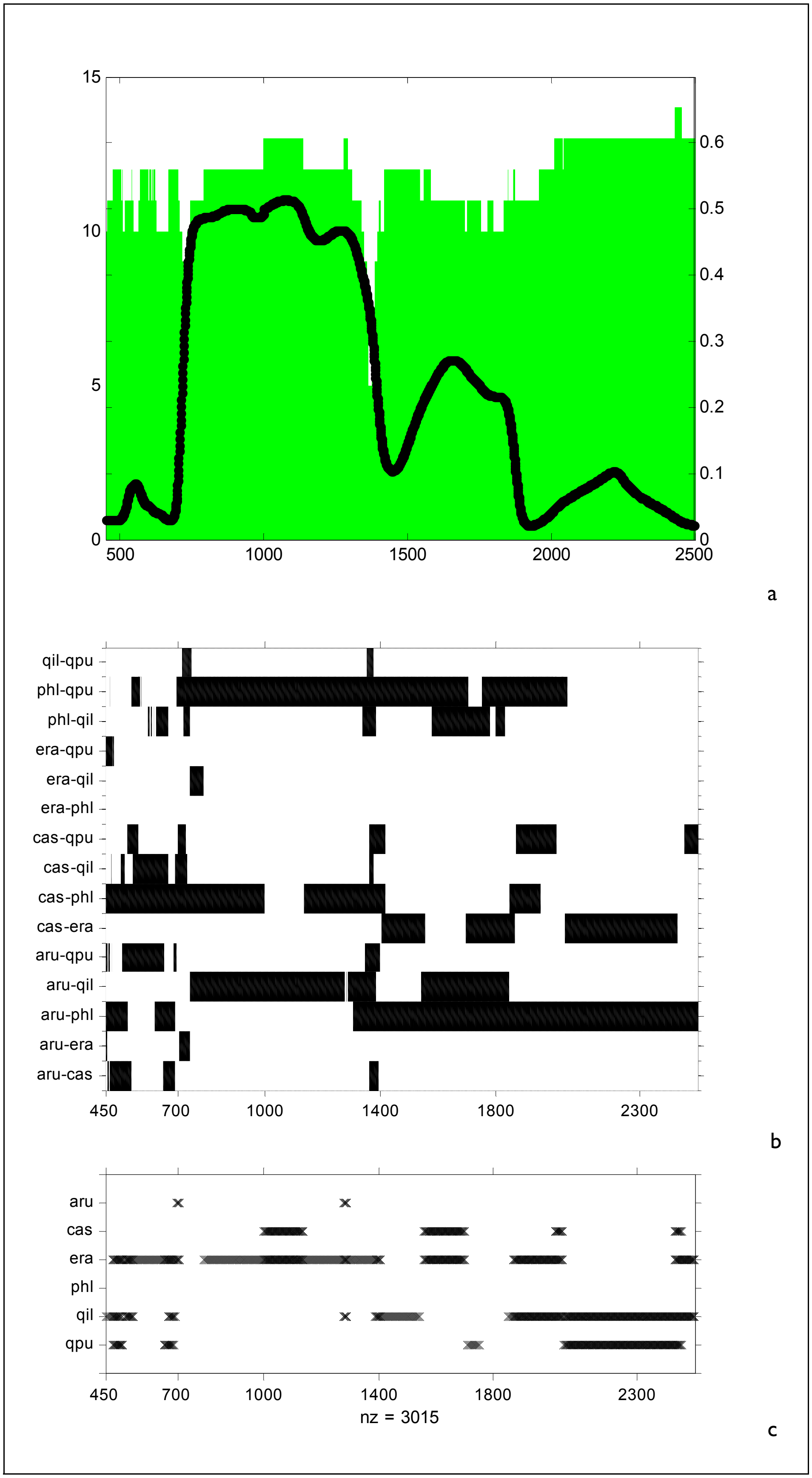


FIG. A.5

Scale **leaf clip**
Data type **reflectance**
Resolution **HyMap**
Classes **6**

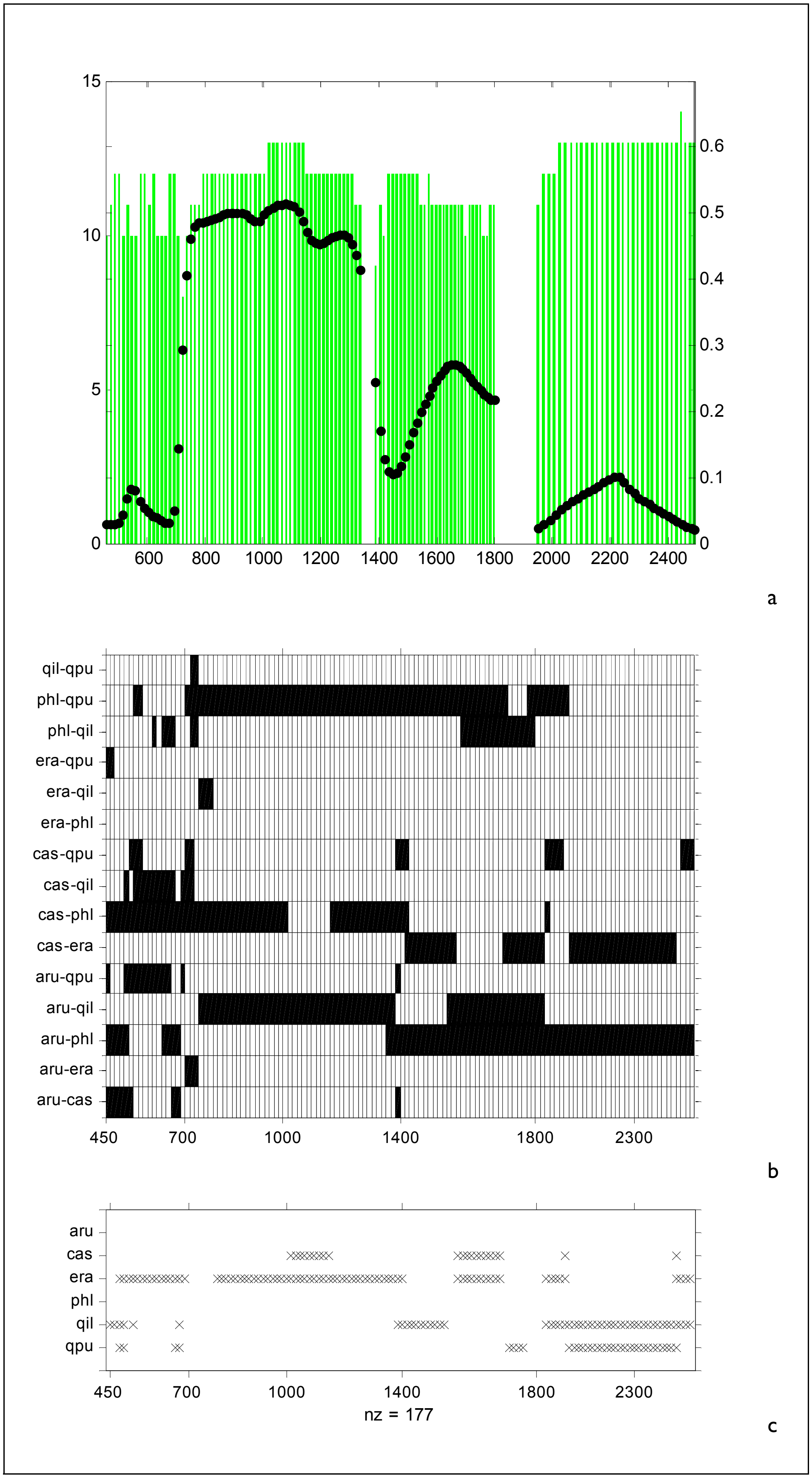


FIG. A.6

Scale **leaf clip**
Data type **reflectance**
Resolution **HyMap**
Classes **16**

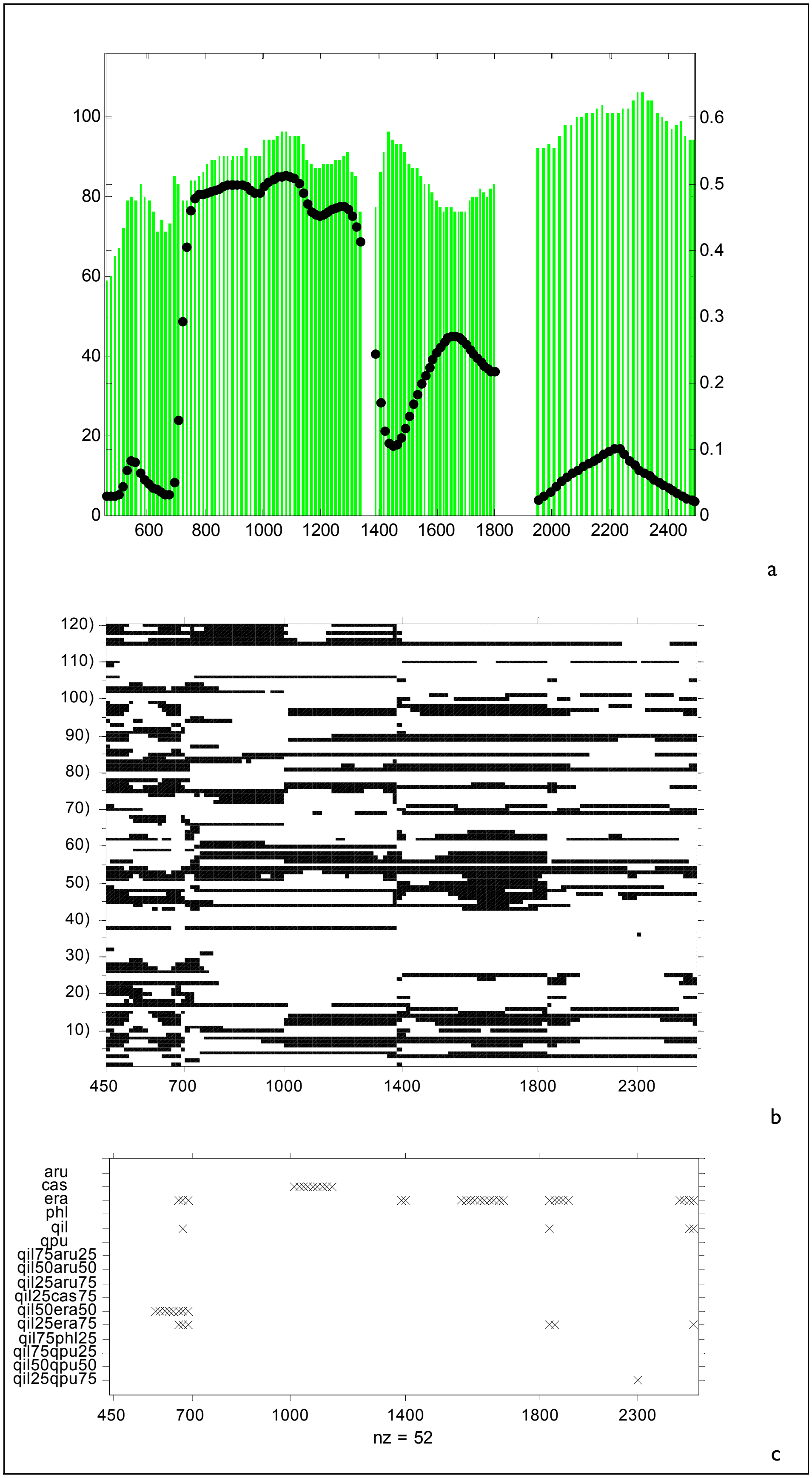


Fig. A.7

Scale **25°optic**
Data type **continuum-removed**
Resolution **ASD**
Classes **6**

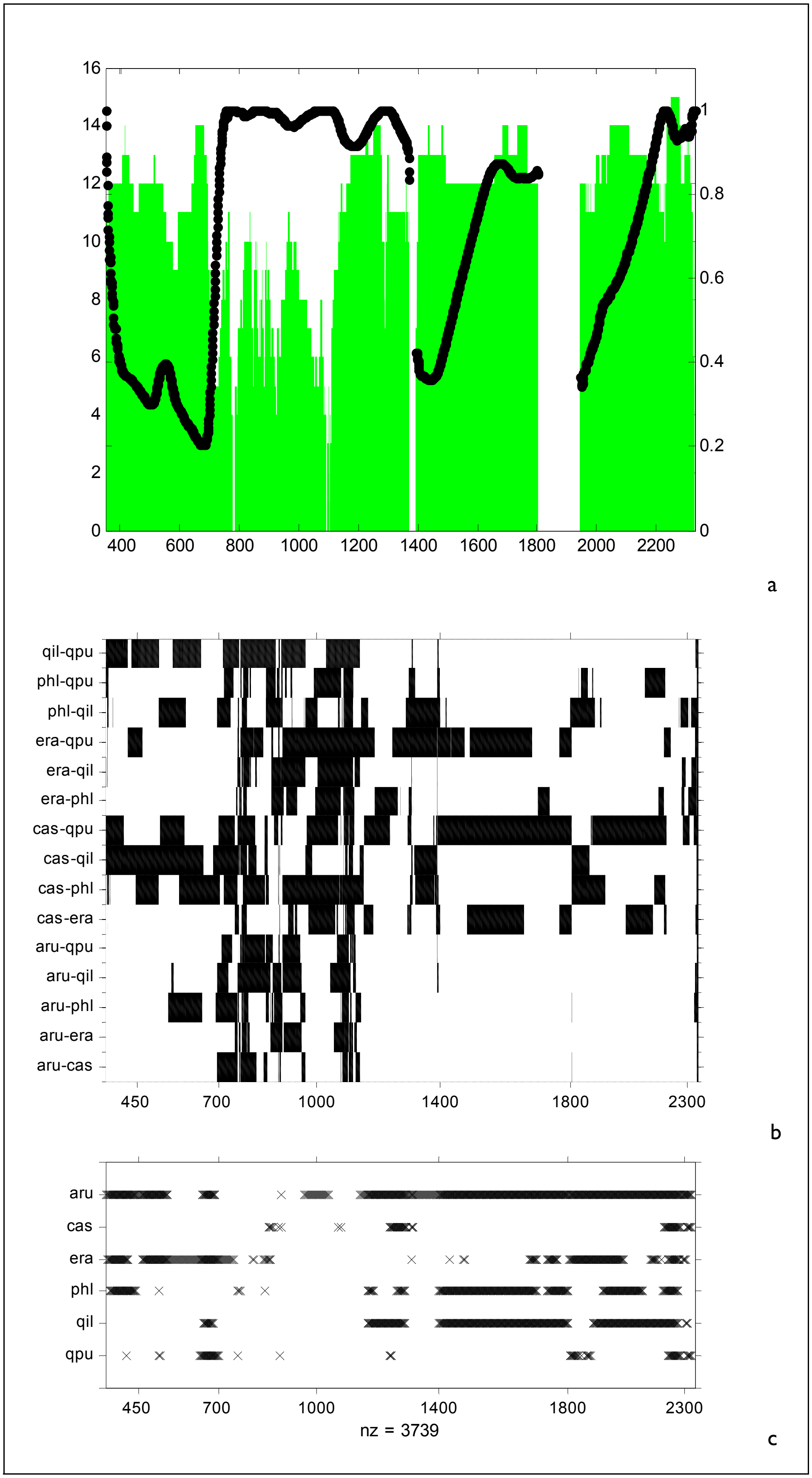


Fig. A.8

Scale **25°optic**
Data type **continuum-removed**
Resolution **HyMap**
Classes **6**

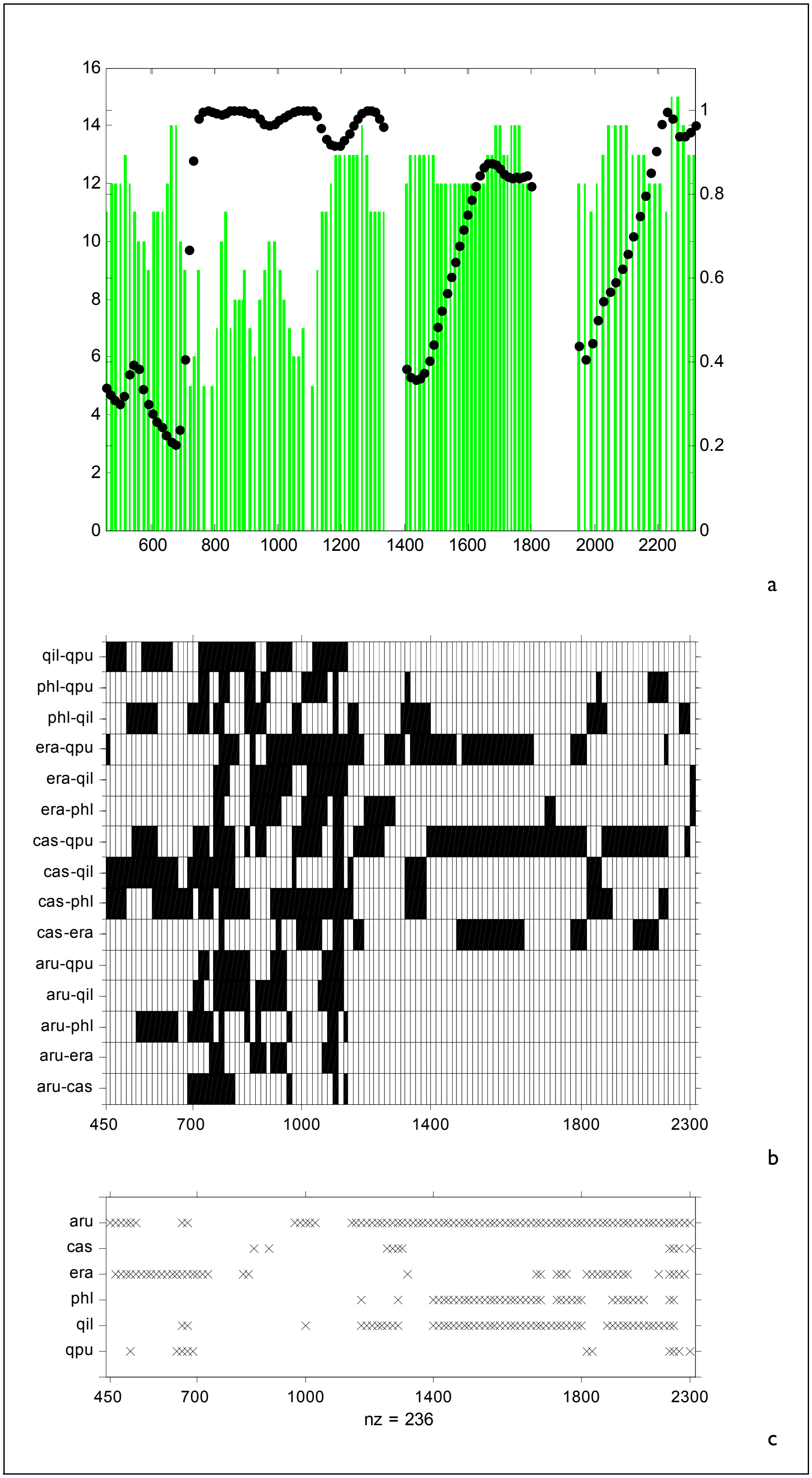


Fig. A.9

Scale **25°optic**
Data type **continuum-removed**
Resolution **HyMap**
Classes **6**

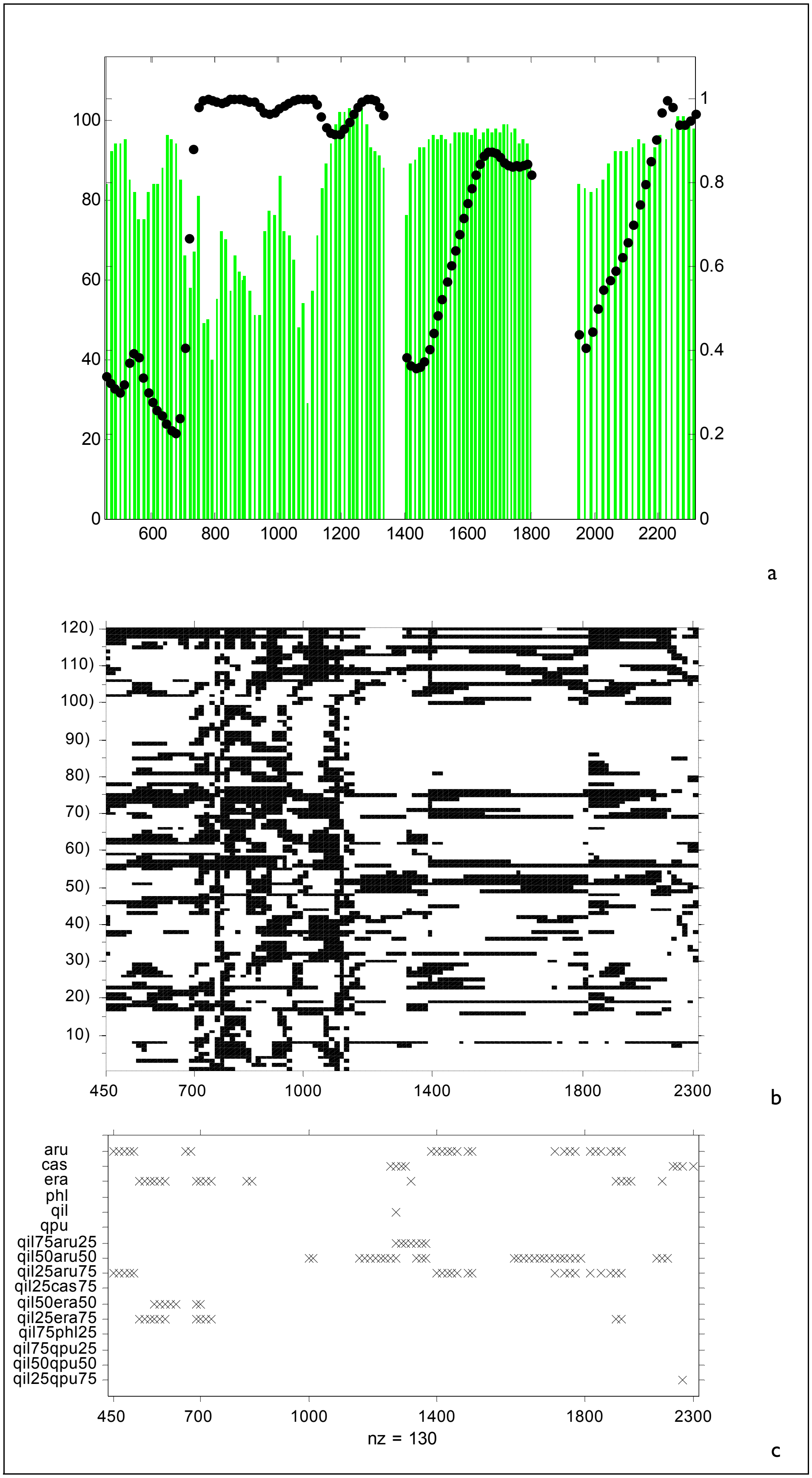


Fig. A.10

Scale **leaf clip**
Data type **continuum-removed**
Resolution **ASD**
Classes **6**

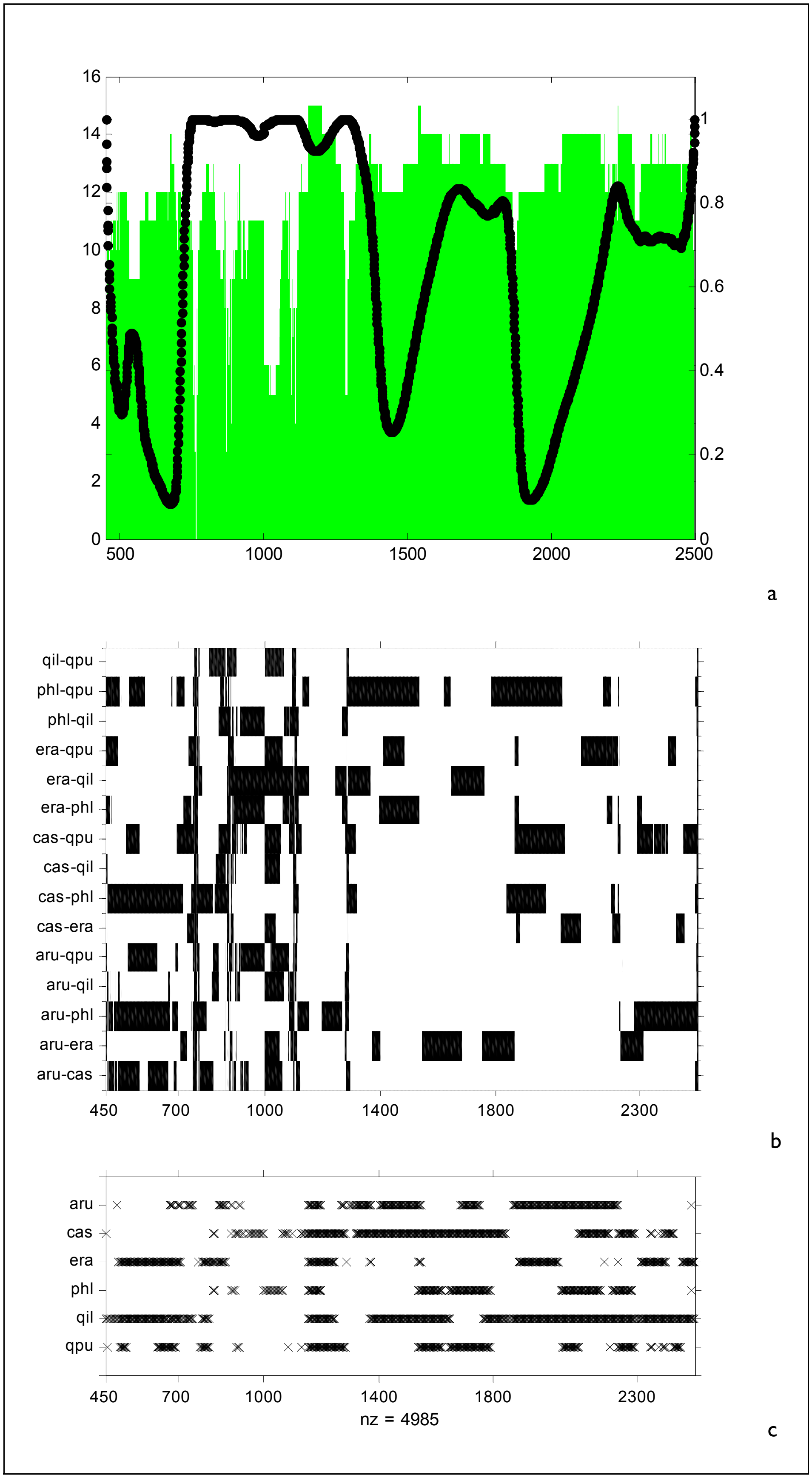


FIG. A.11

Scale **leaf clip**
Data type **continuum-removed**
Resolution **HyMap**
Classes **6**

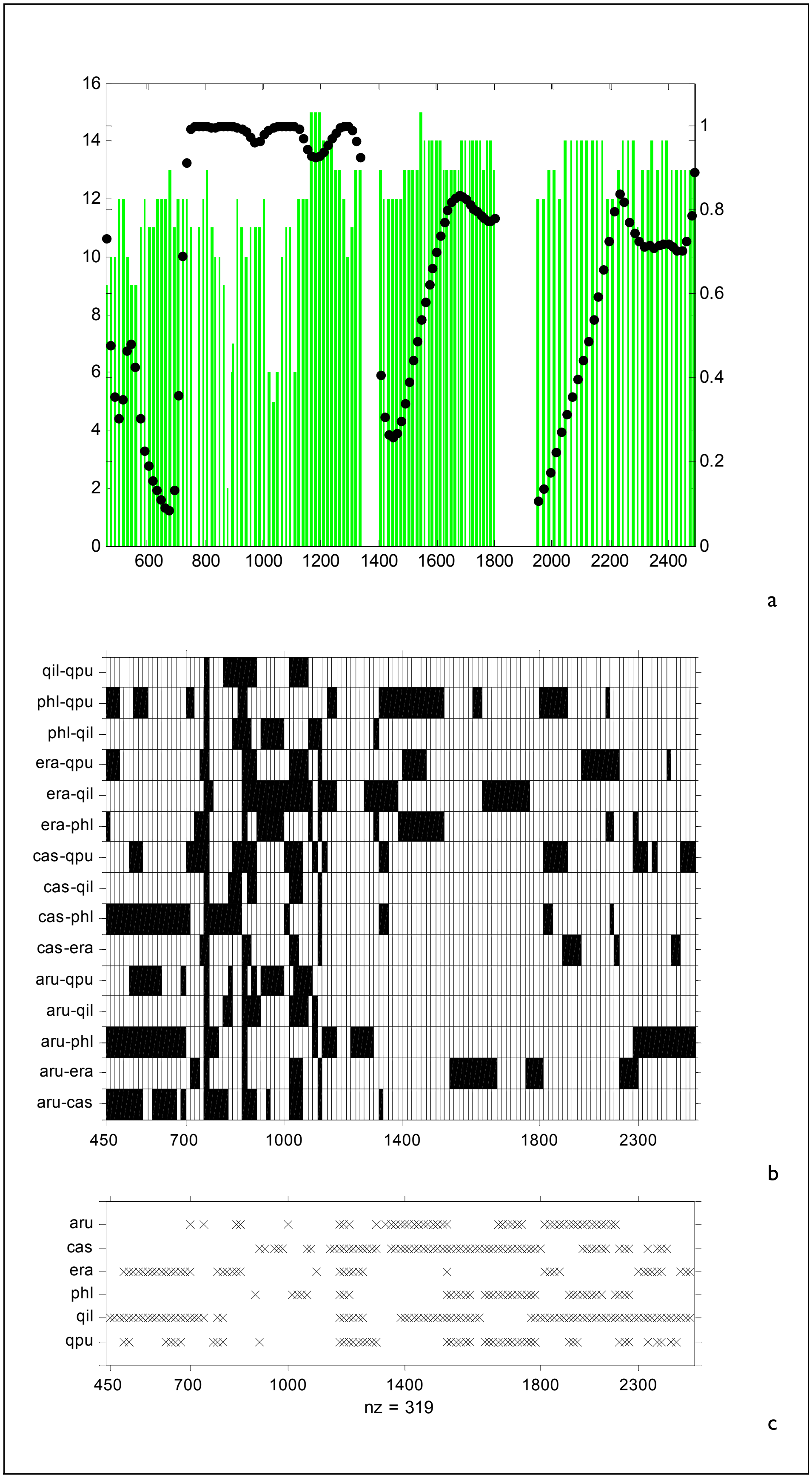


FIG. A.12

Scale **leaf clip**
Data type **continuum-removed**
Resolution **HyMap**
Classes **16**

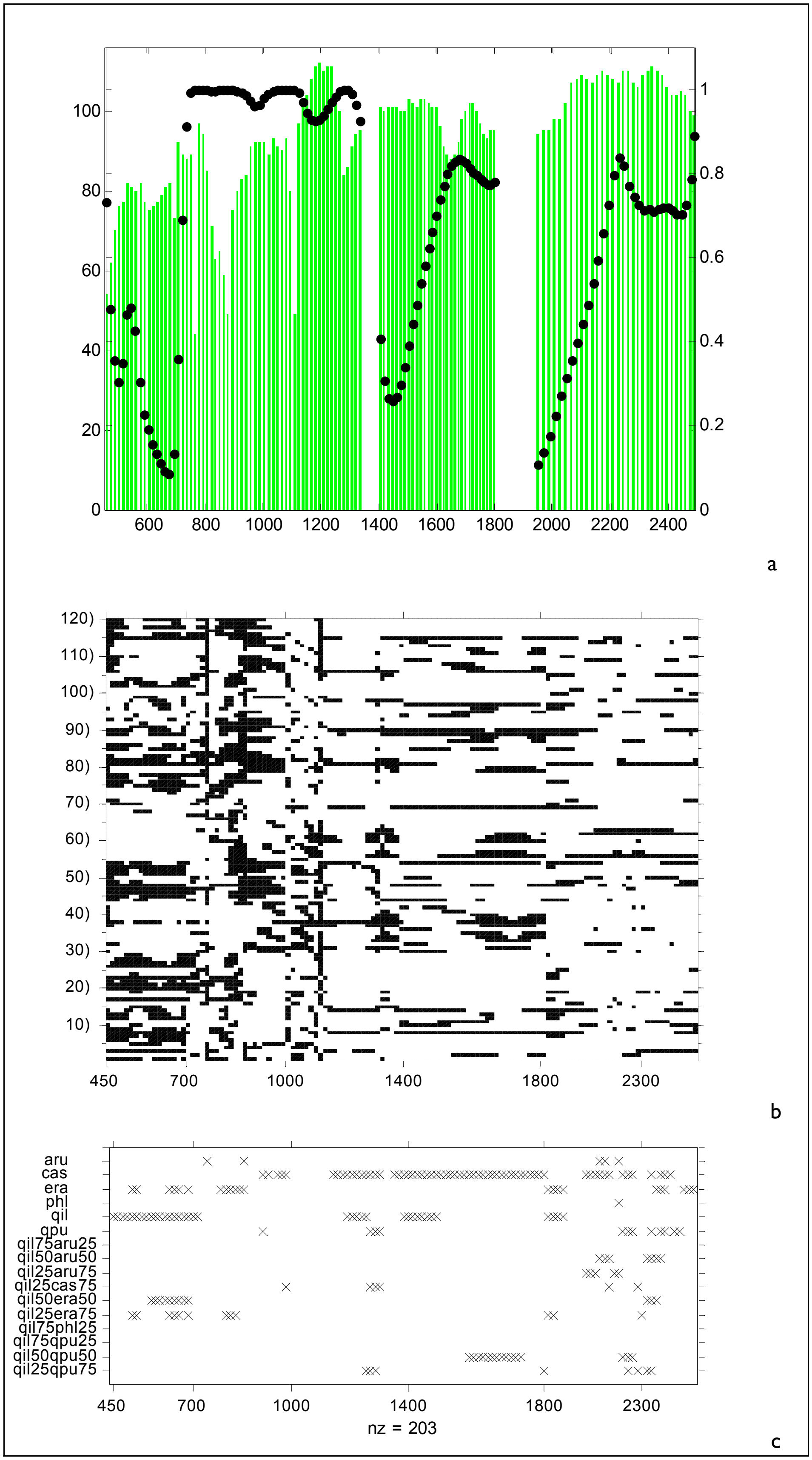


Table A.1

Resolution **ASD**
Classes **6**

	25° optic												**Leaf clip**											
	reflectance						**continuum-removed**						**reflectance**						**continuum-removed**					
	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
a																								
1 **aru-cas**	1817	350	31	670	411	386	1629	341	2	494	411	383	1899	129	31	600	598	572	1656	65	25	398	628	565
2 **aru-era**	1540	311	15	432	411	386	1632	349	31	487	411	385	2010	248	16	562	628	572	1530	246	20	448	344	492
3 **aru-phl**	963	166	0	0	411	386	1504	230	0	488	411	375	710	102	29	600	8	0	1406	17	21	409	628	352
4 **aru-qil**	1042	221	5	28	407	386	1503	337	4	380	407	379	1116	252	31	54	238	572	1859	233	31	428	628	570
5 **aru-qpu**	1817	350	31	670	411	386	1577	349	18	433	410	385	1829	83	24	600	574	572	1694	139	22	361	628	566
6 **cas-era**	1665	267	26	670	411	317	1281	346	31	458	189	288	1338	252	31	600	305	181	1800	248	31	494	614	444
7 **cas-phl**	1817	350	31	670	411	386	1028	145	11	227	408	248	1110	0	0	137	431	542	1457	6	3	429	521	501
8 **cas-qil**	1817	350	31	670	411	386	1191	30	0	424	409	328	1852	103	1	566	611	572	1907	247	31	462	628	570
9 **cas-qpu**	104	41	11	56	0	7	665	216	8	320	1	128	1742	212	11	572	517	441	1392	197	8	345	548	302
10 **era-phl**	1158	286	16	389	97	386	1443	348	31	380	376	339	2052	252	31	600	628	572	1607	233	31	371	471	532
11 **era-qil**	1261	118	16	487	270	386	1506	343	31	392	410	361	2004	252	31	552	628	572	1510	251	31	244	444	571
12 **era-qpu**	1727	350	31	670	407	300	928	303	31	168	101	356	2026	226	31	600	628	572	1631	209	31	465	542	415
13 **phl-qil**	228	182	0	0	0	46	1276	251	3	366	396	263	1688	194	31	577	345	572	1824	250	31	376	628	570
14 **phl-qpu**	1816	350	31	670	410	386	1474	341	25	437	404	292	713	214	6	0	47	452	1268	136	6	477	223	432
15 **qil-qpu**	1816	350	31	670	410	386	1232	114	21	330	407	381	2001	252	25	569	608	572	1850	250	31	401	628	571
b																								
1 **aru**	959	166	0	0	407	386	1324	230	0	312	407	375	20	4	9	16	0	0	763	11	8	105	344	303
2 **cas**	63	0	11	56	0	7	146	0	0	82	0	64	327	0	0	137	140	50	972	1	0	251	521	199
3 **era**	814	34	4	383	97	300	654	300	31	73	55	226	1227	226	16	515	305	181	675	208	20	184	54	229
4 **phl**	165	119	0	0	0	46	656	79	0	47	362	168	0	0	0	0	0	0	546	0	0	120	223	203
5 **qil**	79	33	0	0	0	46	803	29	0	115	396	263	921	103	1	11	235	572	1415	230	31	171	444	570
6 **qpu**	97	41	11	56	0	0	156	60	8	10	0	86	520	67	0	0	47	406	614	80	3	164	223	147

TABLE A.2

Resolution **HyMap**
Classes **6**

a

	25° optic												**leaf clip**											
	reflectance						**continuum-removed**						**reflectance**						**continuum-removed**					
	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
1 **aru-cas**	114	17	3	45	31	21	102	16	0	34	31	21	117	9	3	42	34	32	97	3	2	28	35	31
2 **aru-era**	97	15	2	30	31	21	102	17	3	33	31	21	123	17	1	39	35	32	99	17	2	34	21	27
3 **aru-phl**	56	4	0	0	31	21	94	8	0	34	31	21	51	8	3	42	1	0	82	0	2	29	35	18
4 **aru-qil**	62	8	1	2	31	21	94	17	1	25	31	21	63	17	3	3	11	32	113	17	3	30	35	31
5 **aru-qpu**	114	17	3	45	31	21	98	17	2	29	31	21	113	5	2	42	34	32	104	9	2	29	35	31
6 **cas-era**	105	11	2	45	31	18	82	17	3	34	15	16	83	17	3	42	15	9	111	17	3	35	35	24
7 **cas-phl**	114	17	3	45	31	21	66	5	1	16	31	14	69	0	0	9	29	31	92	0	1	30	33	29
8 **cas-qil**	114	17	3	45	31	21	80	2	0	29	31	18	113	6	0	40	35	32	115	17	3	32	35	31
9 **cas-qpu**	6	2	1	3	0	1	40	12	1	21	0	7	111	14	1	40	32	25	92	14	1	25	33	20
10 **era-phl**	69	14	1	27	7	21	90	17	3	24	29	20	126	17	3	42	35	32	98	16	3	29	25	28
11 **era-qil**	82	8	1	33	20	21	94	17	3	26	31	20	123	17	3	39	35	32	87	17	3	17	22	31
12 **era-qpu**	109	17	3	45	31	16	58	16	3	14	8	20	124	15	3	42	35	32	98	14	3	32	30	22
13 **phl-qil**	8	6	0	0	0	2	81	10	0	26	30	15	104	13	2	40	19	32	111	17	3	28	35	31
14 **phl-qpu**	114	17	3	45	31	21	94	17	2	30	31	16	46	15	1	0	4	27	89	11	1	34	18	26
15 **qil-qpu**	114	17	3	45	31	21	81	7	2	22	31	21	124	17	2	40	35	32	113	17	3	30	35	31

b

	25° optic												**leaf clip**											
	reflectance						**continuum-removed**						**reflectance**						**continuum-removed**					
	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
1 **aru**	56	4	0	0	31	21	80	8	0	20	31	21	0	0	0	0	0	0	45	0	1	9	21	15
2 **cas**	4	0	1	3	0	1	10	0	0	6	0	4	21	0	0	9	10	2	64	0	0	18	33	13
3 **era**	52	2	0	27	7	16	41	16	3	6	5	14	75	15	1	36	15	9	42	14	2	14	2	12
4 **phl**	5	3	0	0	0	2	39	0	0	2	28	9	0	0	0	0	0	0	38	0	0	8	18	12
5 **qil**	4	2	0	0	0	2	55	2	0	9	30	14	49	6	0	0	11	32	82	17	3	12	22	31
6 **qpu**	5	2	1	3	0	0	11	5	1	0	0	6	32	4	0	0	4	24	48	6	0	13	18	11

TABLE A.3

Resolution **HyMap**
Classes **16**

a		25° optic												Leaf clip											
		reflectance						continuum-removed						reflectance						continuum-removed					
		full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
1	aru-cas	114	17	3	45	31	21	102	16	0	34	31	21	117	9	3	42	34	32	97	3	2	28	35	31
2	aru-era	97	15	2	30	31	21	102	17	3	33	31	21	123	17	1	39	35	32	99	17	2	34	21	27
3	aru-phl	56	4	0	0	31	21	94	8	0	34	31	21	51	8	3	42	1	0	82	0	2	29	35	18
4	aru-qil	62	8	1	2	31	21	94	17	1	25	31	21	63	17	3	3	11	32	113	17	3	30	35	31
5	aru-qpu	114	17	3	45	31	21	98	17	2	29	31	21	113	5	2	42	34	32	104	9	2	29	35	31
6	aru-qil75aru25	64	8	2	4	31	21	96	16	2	28	31	21	80	8	3	21	21	30	108	4	2	38	35	31
7	aru-qil50aru50	52	8	0	0	23	21	90	15	1	25	31	19	38	4	2	16	0	18	101	1	2	38	34	28
8	aru-qil25aru75	20	4	0	0	1	15	35	9	0	8	12	6	4	0	2	4	0	0	43	0	1	22	2	19
9	aru-qil25cas75	114	17	3	45	31	21	105	17	1	36	31	21	117	10	3	42	33	32	100	5	2	34	34	27
10	aru-qil50era50	95	14	2	29	31	21	106	17	3	37	31	21	77	11	2	23	11	32	96	13	2	38	24	21
11	aru-qil25era75	97	15	2	30	31	21	108	17	3	39	31	21	116	15	1	34	35	32	104	17	3	40	22	25
12	aru-qil75phl25	78	8	2	18	31	21	103	17	2	34	31	21	65	9	3	21	9	26	103	4	2	38	32	29
13	aru-qil75qpu25	70	15	1	3	31	21	105	17	1	36	31	21	48	7	3	21	3	17	104	6	2	39	31	28
14	aru-qil50qpu50	84	17	1	15	31	21	104	17	1	35	31	21	28	6	3	22	0	0	63	8	2	27	14	14
15	aru-qil25qpu75	105	17	2	36	31	21	104	17	1	35	31	21	99	4	2	42	21	32	105	9	2	35	32	29
16	cas-era	105	11	2	45	31	18	82	17	3	34	15	16	83	17	3	42	15	9	111	17	3	35	35	24
17	cas-phl	114	17	3	45	31	21	66	5	1	16	31	14	69	0	0	9	29	31	92	0	1	30	33	29
18	cas-qil	114	17	3	45	31	21	80	2	0	29	31	18	113	6	0	40	35	32	115	17	3	32	35	31
19	cas-qpu	6	2	1	3	0	1	40	12	1	21	0	7	111	14	1	40	32	25	92	14	1	25	33	20
20	cas-qil75aru25	113	17	3	45	31	20	93	6	2	40	31	16	111	4	0	40	35	32	100	0	0	34	35	31
21	cas-qil50aru50	114	17	3	45	31	21	89	4	1	36	31	18	116	7	3	42	35	32	98	1	0	31	35	31
22	cas-qil25aru75	114	17	3	45	31	21	97	10	0	35	31	21	119	10	3	42	35	32	104	6	2	32	35	31
23	cas-qil25cas75	79	3	2	45	26	5	40	3	0	13	15	9	94	0	0	35	35	24	90	0	0	27	34	29
24	cas-qil50era50	109	15	3	45	31	18	89	14	2	34	23	18	114	17	3	42	30	25	114	15	2	36	34	29
25	cas-qil25era75	109	14	3	45	31	19	87	17	3	37	20	13	84	17	3	42	4	21	114	17	3	36	35	26
26	cas-qil75phl25	113	17	3	45	31	20	85	7	2	32	30	16	104	0	0	37	35	32	102	0	0	36	35	31
27	cas-qil75qpu25	109	14	3	45	31	19	86	10	2	32	28	16	108	3	0	38	35	32	104	2	0	36	35	31
28	cas-qil50qpu50	100	7	2	45	31	17	86	12	2	32	25	17	112	6	0	39	35	32	108	7	0	35	35	31
29	cas-qil25qpu75	69	6	2	36	14	13	79	13	1	30	19	17	120	14	1	39	35	32	110	11	0	34	34	31
30	era-phl	69	14	1	27	7	21	90	17	3	24	29	20	126	17	3	42	35	32	98	16	3	29	25	28
31	era-qil	82	8	1	33	20	21	94	17	3	26	31	20	123	17	3	39	35	32	87	17	3	17	22	31
32	era-qpu	109	17	3	45	31	16	58	16	3	14	8	20	124	15	3	42	35	32	98	14	3	32	30	22
33	era-qil75aru25	96	13	2	31	31	21	99	17	3	32	31	19	126	17	3	42	35	32	104	17	3	34	22	31

		25° optic												Leaf clip											
		reflectance						continuum-removed						reflectance						continuum-removed					
		full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
34	era-qil50aru50	99	17	1	30	31	21	101	17	3	33	31	20	126	17	3	42	35	32	99	17	3	32	19	31
35	era-qil25aru75	100	17	2	31	31	21	103	17	3	34	31	21	126	17	3	42	35	32	97	17	2	29	20	31
36	era-qil25cas75	102	10	2	45	31	16	84	17	3	32	15	20	125	17	3	42	35	31	106	17	3	35	28	26
37	era-qil50era50	23	0	0	2	6	15	92	13	3	29	30	20	126	17	3	42	35	32	96	15	3	30	20	31
38	era-qil25era75	0	0	0	0	0	0	61	7	3	19	21	14	67	3	1	0	32	32	63	6	1	14	12	31
39	era-qil75phl25	80	13	2	30	16	21	99	17	3	31	31	20	126	17	3	42	35	32	103	16	3	37	19	31
40	era-qil75qpu25	55	2	1	44	5	4	97	16	3	31	31	19	126	17	3	42	35	32	104	16	3	35	22	31
41	era-qil50qpu50	78	7	1	44	24	3	86	16	3	26	24	20	126	17	3	42	35	32	111	16	3	35	29	31
42	era-qil25qpu75	106	15	3	45	31	15	78	16	3	23	19	20	126	17	3	42	35	32	108	16	3	35	28	29
43	phl-qil	8	6	0	0	0	2	81	10	0	26	30	15	104	13	2	40	19	32	111	17	3	28	35	31
44	phl-qpu	114	17	3	45	31	21	94	17	2	30	31	16	46	15	1	0	4	27	89	11	1	34	18	26
45	phl-qil75aru25	3	3	0	0	0	0	76	11	1	31	25	9	92	0	0	36	24	32	94	2	2	26	35	31
46	phl-qil50aru50	47	0	0	0	31	16	79	1	0	31	31	16	85	0	2	42	14	29	88	0	0	22	35	31
47	phl-qil25aru75	55	4	0	0	31	20	91	7	0	32	31	21	71	9	3	42	11	9	86	0	0	22	35	29
48	phl-qil25cas75	114	17	3	45	31	21	78	17	2	19	31	11	36	0	2	2	6	28	57	0	2	25	14	18
49	phl-qil50era50	35	14	1	13	0	8	59	17	3	19	10	13	68	15	2	40	3	10	95	11	2	28	35	21
50	phl-qil25era75	62	16	1	21	6	19	79	17	3	25	21	16	98	17	3	42	7	32	91	15	2	28	29	19
51	phl-qil75phl25	7	7	0	0	0	0	30	13	1	13	0	4	75	3	2	23	17	32	99	3	2	30	35	31
52	phl-qil75qpu25	40	17	1	0	6	17	41	17	1	17	1	6	67	1	2	24	15	27	94	2	2	26	35	31
53	phl-qil50qpu50	113	17	3	44	31	21	59	17	2	23	7	12	5	1	0	4	0	0	85	0	2	21	34	30
54	phl-qil25qpu75	114	17	3	45	31	21	96	17	2	32	30	17	14	8	0	0	0	6	42	6	1	20	4	12
55	qil-qpu	114	17	3	45	31	21	81	7	2	22	31	21	124	17	2	40	35	32	113	17	3	30	35	31
56	qil-qil75aru25	0	0	0	0	0	0	11	0	0	11	0	0	34	12	2	19	0	3	60	17	3	31	9	3
57	qil-qil50aru50	5	4	0	0	1	0	46	4	0	21	15	6	59	17	3	3	7	32	95	17	3	29	23	26
58	qil-qil25aru75	56	7	0	0	30	19	88	13	0	25	30	20	65	17	3	5	11	32	113	17	3	30	35	31
59	qil-qil25cas75	114	17	3	45	31	21	83	1	0	31	31	20	116	9	1	40	35	32	116	17	3	33	35	31
60	qil-qil50era50	77	3	1	39	14	21	93	17	3	26	31	19	83	17	3	2	32	32	88	17	3	24	17	30
61	qil-qil25era75	87	7	1	40	19	21	98	17	3	31	31	19	112	17	3	28	35	32	97	17	3	27	22	31
62	qil-qil75phl25	1	0	0	0	0	1	58	0	0	13	28	17	55	4	1	38	5	8	83	17	3	35	25	6
63	qil-qil75qpu25	62	10	1	0	31	21	73	2	0	19	31	21	91	13	1	40	9	29	97	17	3	33	33	14
64	qil-qil50qpu50	114	17	3	45	31	21	80	9	1	19	31	21	112	17	1	39	24	32	117	17	3	34	35	31
65	qil-qil25qpu75	114	17	3	45	31	21	97	15	2	30	31	21	124	17	2	40	35	32	119	17	3	36	35	31
66	qpu-qil75aru25	114	17	3	45	31	21	89	13	3	27	31	18	104	16	1	21	35	32	117	17	3	34	35	31
67	qpu-qil50aru50	114	17	3	45	31	21	96	17	2	27	31	21	117	9	1	41	35	32	115	17	1	32	35	31
68	qpu-qil25aru75	114	17	3	45	31	21	97	17	2	28	31	21	117	8	2	42	35	32	116	13	3	37	35	31
69	qpu-qil25cas75	7	5	1	1	0	1	36	13	1	18	3	2	54	17	3	34	3	0	67	17	3	28	1	21
70	qpu-qil50era50	113	17	3	45	31	20	64	11	3	23	20	10	93	9	3	42	21	21	110	10	2	37	34	29
71	qpu-qil25era75	111	17	3	45	31	18	61	16	3	20	12	13	85	15	3	42	15	13	111	14	3	37	35	25
72	qpu-qil75phl25	114	17	3	45	31	21	79	6	3	28	30	15	110	17	3	27	34	32	121	17	3	38	35	31

		25° optic												Leaf clip											
		reflectance						continuum-removed						reflectance						continuum-removed					
		full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
73	qpu-qil75qpu25	113	17	3	45	31	20	68	0	2	24	30	14	111	17	3	28	34	32	118	15	3	37	35	31
74	qpu-qil50qpu50	100	7	3	45	31	17	54	0	0	19	25	10	101	11	3	24	34	32	114	11	3	37	35	31
75	qpu-qil25qpu75	0	0	0	0	0	0	12	0	0	7	0	5	75	0	0	13	32	30	99	2	0	31	35	31
76	qil75aru25-qil50aru50	1	1	0	0	0	0	46	7	0	19	14	6	43	2	2	21	0	20	93	3	2	29	34	27
77	qil75aru25-qil25aru75	62	8	1	2	31	21	90	13	0	25	31	21	83	7	3	21	25	30	108	7	3	35	35	31
78	qil75aru25-qil25cas75	114	17	3	45	31	21	89	3	1	39	31	16	109	1	1	41	35	32	100	2	0	32	35	31
79	qil75aru25-qil50era50	91	10	1	29	31	21	98	17	3	31	31	19	126	17	3	42	35	32	91	16	3	26	18	31
80	qil75aru25-qil25era75	97	14	2	31	31	21	98	17	3	31	31	19	126	17	3	42	35	32	96	16	3	28	21	31
81	qil75aru25-qil75phl25	0	0	0	0	0	0	68	5	0	23	29	11	20	0	1	20	0	0	38	0	0	10	23	5
82	qil75aru25-qil75qpu25	107	17	2	38	31	21	91	14	1	29	31	17	63	0	1	38	0	25	78	3	2	15	32	28
83	qil75aru25-qil50qpu50	114	17	3	45	31	21	96	17	1	31	31	17	95	0	0	29	34	32	96	8	0	22	35	31
84	qil75aru25-qil25qpu75	114	17	3	45	31	21	99	17	3	33	31	18	100	12	1	21	35	32	110	17	2	27	35	31
85	qil50aru50-qil25aru75	5	4	0	0	1	0	79	8	0	22	31	18	30	4	2	12	0	14	80	4	2	31	19	26
86	qil50aru50-qil25cas75	114	17	3	45	31	21	99	10	2	39	31	19	118	9	3	42	35	32	106	5	2	35	35	31
87	qil50aru50-qil50era50	97	15	2	30	31	21	102	17	3	33	31	21	119	16	2	36	35	32	94	17	2	29	22	26
88	qil50aru50-qil25era75	100	17	2	31	31	21	103	17	3	34	31	21	126	17	3	42	35	32	103	17	3	30	25	31
89	qil50aru50-qil75phl25	60	4	2	5	31	20	93	9	0	32	31	21	37	7	3	22	0	8	62	4	2	34	0	24
90	qil50aru50-qil75qpu25	85	17	1	16	31	21	101	17	1	32	31	21	36	5	3	31	0	0	40	5	2	19	2	14
91	qil50aru50-qil50qpu50	114	17	3	45	31	21	101	17	2	32	31	21	107	0	2	42	33	32	97	8	2	23	35	31
92	qil50aru50-qil25qpu75	114	17	3	45	31	21	101	17	2	32	31	21	117	8	2	42	35	32	106	15	2	25	35	31
93	qil25aru75-qil25cas75	114	17	3	45	31	21	106	17	2	37	31	21	120	12	3	42	34	32	106	11	3	29	35	31
94	qil25aru75-qil50era50	97	15	2	30	31	21	106	17	3	37	31	21	113	14	1	32	35	32	102	16	2	37	24	25
95	qil25aru75-qil25era75	100	17	2	31	31	21	105	17	3	36	31	21	123	17	2	39	35	32	100	17	2	33	23	27
96	qil25aru75-qil75phl25	79	8	2	19	31	21	102	16	1	34	31	21	61	10	3	22	5	24	107	12	3	39	28	28
97	qil25aru75-qil75qpu25	83	17	1	14	31	21	102	17	1	33	31	21	43	8	3	22	3	10	77	11	2	35	5	26
98	qil25aru75-qil50qpu50	106	17	2	37	31	21	103	17	1	34	31	21	79	7	3	42	3	27	88	11	2	30	31	16
99	qil25aru75-qil25qpu75	114	17	3	45	31	21	103	17	2	34	31	21	119	11	3	42	34	32	106	11	2	29	35	31
100	qil25cas75-qil50era50	108	14	3	45	31	18	86	17	3	39	18	12	104	17	3	42	20	25	114	17	3	37	35	25
101	qil25cas75-qil25era75	105	11	2	45	31	18	81	17	3	36	14	14	96	17	3	42	22	15	113	17	3	37	35	24
102	qil25cas75-qil75phl25	112	16	3	45	31	20	78	3	2	31	30	14	89	0	0	22	35	32	104	3	0	35	35	31
103	qil25cas75-qil75qpu25	103	9	3	45	31	18	81	8	1	35	25	13	103	1	0	35	35	32	105	5	1	34	35	31
104	qil25cas75-qil50qpu50	91	6	2	44	26	15	81	13	1	34	19	15	119	14	1	38	35	32	106	11	0	30	34	31
105	qil25cas75-qil25qpu75	7	3	1	3	0	1	47	13	1	26	0	8	120	17	3	42	33	28	95	17	2	31	32	15
106	qil50era50-qil25era75	0	0	0	0	0	0	56	9	3	18	20	9	80	14	3	2	32	32	76	14	3	16	15	31
107	qil50era50-qil75phl25	60	9	2	23	7	21	92	17	3	31	26	18	126	17	3	42	35	32	100	15	3	29	25	31
108	qil50era50-qil75qpu25	69	2	1	44	23	0	59	15	3	13	14	17	126	17	3	42	35	32	102	14	3	29	28	31
109	qil50era50-qil50qpu50	110	17	3	45	31	17	43	14	3	16	2	11	124	15	3	42	35	32	93	14	3	30	29	20
110	qil50era50-qil25qpu75	113	17	3	45	31	20	64	16	3	26	12	10	74	14	3	42	8	10	104	11	3	32	35	26
111	qil25era75-qil75phl25	76	13	2	29	13	21	99	17	3	32	31	19	126	17	3	42	35	32	110	16	3	38	25	31

	25° optic												Leaf clip											
	reflectance						continuum-removed						reflectance						continuum-removed					
	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2	full	VIS	red	NIR	SWIR1	SWIR2
112 qil25era75-qil75qpu25	63	2	1	44	17	0	85	16	3	27	24	18	126	17	3	42	35	32	110	16	3	35	28	31
113 qil25era75-qil50qpu50	98	10	2	45	31	12	71	17	3	19	16	19	126	17	3	42	35	32	104	15	3	30	29	30
114 qil25era75-qil25qpu75	110	17	3	45	31	17	70	16	3	22	13	19	126	17	3	42	35	32	104	15	3	34	31	24
115 qil75phl25-qil75qpu25	114	17	3	45	31	21	27	10	1	13	0	4	10	0	0	0	0	10	41	0	0	19	4	18
116 qil75phl25-qil50qpu50	114	17	3	45	31	21	72	7	2	30	30	5	74	5	0	6	31	32	98	7	1	25	35	31
117 qil75phl25-qil25qpu75	114	17	3	45	31	21	87	11	3	34	30	12	105	17	2	22	34	32	115	15	3	34	35	31
118 qil75qpu25-qil50qpu50	98	7	3	45	30	16	10	0	0	9	0	1	71	0	0	8	31	32	95	3	2	26	35	31
119 qil75qpu25-qil25qpu75	114	17	3	45	31	21	57	0	0	24	29	4	98	8	2	24	34	32	109	11	3	32	35	31
120 qil50qpu50-qil25qpu75	71	2	2	45	24	0	12	0	0	7	3	2	76	3	0	9	32	32	97	5	3	26	35	31

b

1 aru	20	4	0	0	1	15	25	7	0	0	12	6	0	0	0	0	0	0	5	0	0	2	0	3
2 cas	3	0	1	3	0	0	8	0	0	4	0	4	9	0	0	9	0	0	62	0	0	16	33	13
3 era	0	0	0	0	0	0	18	7	3	6	0	5	24	3	1	0	12	9	22	6	1	6	1	9
4 phl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
5 qil	0	0	0	0	0	0	1	0	0	1	0	0	4	1	0	0	0	3	36	17	3	7	9	3
6 qpu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	4	0	8
7 qil75aru25	0	0	0	0	0	0	7	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 qil50aru50	0	0	0	0	0	0	30	0	0	13	14	3	0	0	0	0	0	0	7	0	0	0	0	7
9 qil25aru75	4	4	0	0	0	0	21	5	0	0	11	5	0	0	0	0	0	0	5	0	0	0	0	5
10 qil25cas75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	4	0	2
11 qil50era50	0	0	0	0	0	0	7	6	2	1	0	0	8	8	1	0	0	0	12	9	1	0	0	3
12 qil25era75	0	0	0	0	0	0	12	7	3	3	0	2	6	3	1	0	0	3	12	6	1	3	1	2
13 qil75phl25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 qil75qpu25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15 qil50qpu50	40	2	0	14	24	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	12	3
16 qil25qpu75	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	8	0	0	3	1	4

Appendix B – Spectral unmixing

The table displays summary statistics of species estimations (Θ) as they are derived from the image spectral mixture analysis.
The statistics are calculated on a subset of the full image. Only the 169 estimations at the locations of the validation plots are considered.

Values are expressed in percentage.

Table B.1 Summary statistics of species estimations at the sampling sites

		data type	**reflectance**						**continuum-removed**					
		library	**6**		**10**		**16**		**6**		**10**		**16**	
		spectrum range	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**
a	*st.dev. Θ*	**aru**	6	7	0	2	3	2	12	16	3	1	12	4
		cas	3	5	1	1	2	1	9	8	0	0	6	1
		era	3	2	1	1	1	0	7	14	2	2	4	4
		phl	0	0	0	0	1	1	9	8	0	0	4	0
		qil	9	10	0	4	3	2	14	20	1	1	11	2
		qpu	5	11	0	5	2	0	7	15	2	3	5	3
b	*average Θ*	**aru**	21	30	14	15	44	13	28	31	14	4	24	26
		cas	33	5	6	0	11	2	27	7	0	0	12	1
		era	0	0	4	1	3	9	2	7	21	6	5	5
		phl	0	0	0	0	5	14	2	3	0	0	1	0
		qil	33	10	63	61	32	50	28	31	50	50	48	48
		qpu	12	54	13	23	6	13	13	22	15	40	10	19
c	*min. Θ*	**aru**	0	9	11	11	21	8	0	0	10	3	0	11
		cas	27	0	4	0	8	0	0	0	0	0	0	0
		era	0	0	3	0	0	9	0	0	11	3	0	2
		phl	0	0	0	0	0	9	0	0	0	0	0	0
		qil	7	0	62	49	29	43	0	0	48	48	16	41
		qpu	1	26	11	13	0	12	0	0	4	6	0	15
d	*max. Θ*	**aru**	37	44	14	19	46	23	53	62	28	16	53	30
		cas	46	21	8	6	25	6	37	37	2	0	26	2
		era	25	25	7	7	4	9	40	59	23	24	19	21
		phl	0	0	0	0	6	16	53	48	4	0	27	2
		qil	53	35	64	67	54	53	75	70	54	54	70	51
		qpu	26	76	14	36	15	13	43	55	18	44	28	32

APPENDIX C – ACCURACY ASSESSMENT

Table 1 contains the field observations on canopy composition and height. X and Y columns indicate UTM coordinates of the site.
Table 2 contains the medians per-species errors ("error" is meant as the raw difference between estimation and observation).
Tables 3 and 4 display per-species and per-site RMSE, respectively.

In tables 1 and 4 the 169 sites are identified by a stratum letter (A, B, C, E), a cluster number (1-5; 9 is used for extra sites identified outside the random scheme) and a site number (1-8). So that the site code A11 identifies the first site of the first cluster of stratum A.

Values are expressed in percentage.

TABLE C.1 – FIELD OBSERVATIONS

stratum	cluster	site	X	Y	aru	cas	era	phl	qil	qpu	height (cm)
A	1	1	521511	4826453					100		600
		2	521554	4826426					100		700
		3	521602	4826441					90	10	800
		4	521601	4826394					50	50	1200
		5	521578	4826440					100		800
		6	521549	4826400					100		600
		7	521524	4826354					65	35	400
		8	521532	4826399					100		600
	2	1	521616	4826064	15			15	70		600
		2	521645	4826020	30				70		500
		3	521695	4826025	20				80		900
		4	521749	4826023					20	80	1200
		5	521774	4825981					20	80	1700
		6	521723	4825971				10	50	40	600
		7	521677	4825982	35				65		400
		8	521653	4826033				10	90		600
	3	1	521873	4826382					95	5	450
		2	521831	4826410					90	10	600
		3	521800	4826371	25				75		600
		4	521812	4826419	10				90		600
		5	521786	4826461	30				70		600
		6	521808	4826430	5				95		600
		7	521821	4826383					100		500
		8	521774	4826372						100	1200
	4	1	521056	4826009					30	70	1200
		2	521101	4826029					50	50	900
		3	521095	4825980						100	1200
		4	521121	4826022					10	90	1100

stratum	cluster	site	X	Y	aru	cas	era	phl	qil	qpu	height (cm)
		5	521129	4826069					20	80	1200
		6	521084	4826050						100	1500
		7	521068	4826092					10	90	1400
		8	521108	4826063					20	80	1200
	5	1	521152	4826254					20	80	1400
		2	521129	4826209	10				20	70	1200
		3	521116	4826162						100	1000
		4	521086	4826199						100	1100
		5	521059	4826156						100	1200
		6	521017	4826189						100	1300
	5	7	520965	4826186					90	10	1300
		8	520962	4826241						100	1400
	9	1	521750	4826368						100	1300
		2	521739	4826381					10	90	1500
B	1	1	522823	4826233					100		750
		2	522844	4826195				20	80		400
		3	522802	4826131				25	75		500
		4	522769	4826095	50			10	15	25	600
		5	522772	4826146	15			10	60	15	600
		6	522799	4826195	10			10	40	40	600
		7	522752	4826187				20	60	20	600
		8	522771	4826233					90	10	700
	2	1	524243	4825795					100		850
		2	524201	4825770					100		500
		3	524246	4825791					100		1000
		4	524288	4825753					100		1000
		5	524238	4825742					100		900
		6	524243	4825695					100		400
		7	524194	4825682					100		600
		8	524213	4825638					100		600
	3	1	523203	4825977					100		900
		2	523246	4825953		90			10		1200
		3	523211	4825993	40				60		700
		4	523177	4826031	15	15			70		800
		5	523131	4826054	25				75		700
		6	523088	4826024		100					1700
		7	523130	4825994	30				20	50	700
		8	523081	4826002	70			5	25		500
	4	1	523720	4825820					100		700
		2	523770	4825830					100		500
		3	523738	4825861					100		600
		4	523784	4825836					100		500
		5	523784	4825787					100		500
		6	523740	4825762				20	80		500
		7	523788	4825739				20	80		500
		8	523826	4825709					100		500
	5	1	522966	4826294					20	80	1200
		2	523016	4826296	15				85		700

stratum	cluster	site	X	Y	aru	cas	era	phl	qil	qpu	height (cm)
		3	523047	4826260					80	20	700
		4	522990	4826268					100		800
		5	522943	4826291					100		800
		6	522978	4826327					40	60	1100
		7	523024	4826339					75	25	700
		8	522992	4826373	20				80		1200
C	1	1	524954	4823251					100		750
		2	524985	4823292	50				50		400
		3	524946	4823327	30				70		400
		4	524994	4823345	60				40		600
		5	525011	4823394	65		5		30		600
		6	525005	4823445	60			10	30		600
		7	524964	4823472	30		5	5	60		550
		8	525017	4823458	55		10		35		600
	2	1	525340	4823423	70				30		350
		2	525290	4823419	60		10		30		350
		3	525270	4823371	30		10		60		500
		4	525278	4823421	60		5		35		600
		5	525228	4823411	50				50		400
		6	525202	4823369	40				60		500
		7	525226	4823324	60		5		35		600
		8	525262	4823283	90				10		700
	3	1	524803	4823203	50				50		600
		2	524797	4823153	70				30		500
		3	524834	4823182	30		40		30		250
		4	524830	4823228	30		10		60		500
		5	524808	4823275	30				70		550
		6	524764	4823254	25				75		400
		7	524814	4823263					100		700
		8	524855	4823289					100		600
	4	1	525421	4823996	35				65		500
		2	525376	4824016	75				25		500
		3	525326	4824025	40				60		600
		4	525300	4823982	30				50	20	700
		5	525276	4823935	20				80		600
		6	525281	4823885	20				80		600
		7	525234	4823872	40				60		800
		8	525276	4823902	30				70		500
	5	1	524880	4824025	10				90		600
		2	524834	4824004	45				55		600
		3	524837	4823955	10		5		85		600
		4	524832	4823904	35			10	55		700
		5	524883	4823887	50				50		600
		6	524871	4823937	70				30		700
		7	524822	4823925	15			15	70		900
		8	524781	4823892	40			5	55		600
E	1	1	523466	4823095			60		40		250
		2	523506	4823070	10		15		75		400

stratum	cluster	site	X	Y	aru	cas	era	phl	qil	qpu	height (cm)
		3	523569	4823052	30		5		65		400
		4	523553	4823103			5		95		500
		5	523586	4823059	40		10		50		500
		6	523584	4823102	40		20		40		400
		7	523630	4823110	20		70		10		300
		8	523591	4823135	5			20	75		650
	2	1	523819	4823186	40		30		30		400
		2	523782	4823154			10		90		600
		3	523736	4823133			10		90		600
		4	523761	4823169			65		35		250
		5	523720	4823200	40		20		40		500
		6	523706	4823154			45		55		400
		7	523663	4823139	25				75		1000
		8	523691	4823097	15		50		35		350
	3	1	523427	4823410	40				60		800
		2	523429	4823460	80		10		10		550
		3	523390	4823500	70		10		20		500
		4	523436	4823488	50				50		550
		5	523482	4823476	35				65		600
		6	523433	4823497	30				70		600
		7	523432	4823548	60				40		600
		8	523397	4823583	70				30		700
	4	1	523532	4823793	25				75		600
		2	523572	4823822	35				65		500
		3	523521	4823826	50			10	40		650
		4	523472	4823836	20				80		600
		5	523496	4823792	50				50		600
		6	523448	4823793	30				70		600
		7	523432	4823841				20	80		600
		8	523381	4823844	30			5	65		700
	5	1	523451	4823650	80		10		10		500
		2	523410	4823685	70		15		15		400
		3	523386	4823726	70		10		20		500
		4	523344	4823753	40		10		50		500
		5	523386	4823781	30		10		60		600
		6	523431	4823805	40				60		500
		7	523449	4823759	35			5	60		900
		8	523406	4823738	90				10		900
	9	1	523837	4823072			80		20		350
		2	523850	4823050	10		60		30		400
		3	523824	4823117			70		30		400
		4	523866	4823139	20		60		20		300
		5	523856	4823170	50		40		10		400
		6	523549	4823289				35	65		900
		7	523431	4823041	30		60		10		300

TABLE C.2 – ERROR MEDIANS PER-SPECIES

	data type	reflectance						continuum-removed					
	library	6		10		16		6		10		16	
	spectrum range	7	126	7	126	7	126	7	126	7	126	7	126
species	**aru**	7	15	1	1	30	3	9	14	1	-9	8	12
	cas	32	3	6	0	10	1	31	3	0	0	12	1
	era	0	0	4	0	3	9	0	0	21	5	2	3
	phl	0	0	0	0	5	14	0	0	0	0	0	0
	qil	-28	-51	4	1	-29	-11	-33	-29	-11	-10	-6	-11
	qpu	10	51	13	21	6	13	11	14	15	40	7	18

TABLE C.3 – RMSE PER-SPECIES

	data type	reflectance						continuum-removed								
	library	6		10		16		6		10		16				
	spectrum range	7	126	7	126	7	126	7	126	7	126	7	126	**avg.**	**min**	**max**
species	**aru**	23	22	25	25	33	26	24	24	26	30	26	24	26	22	33
	cas	33	12	11	10	14	10	29	14	10	10	15	10	15	10	33
	era	16	16	15	16	16	16	12	10	23	15	15	14	15	10	23
	phl	6	6	6	6	6	13	11	10	6	6	7	6	7	6	13
	qil	40	56	32	32	41	33	44	37	33	32	31	33	37	31	56
	qpu	31	51	30	30	30	30	28	32	29	40	27	30	32	27	51
	avg.	25	27	20	20	23	21	25	21	21	22	20	19			

TABLE C.4 – RMSE PER-SITE

data type		**reflectance**						**continuum-removed**								
library		**6**		**10**		**16**		**6**		**10**		**16**				
spectrum range		**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**avg.**	**min.**	**max.**
sites	**A11**	28	36	17	17	34	24	31	15	24	26	21	26	25	15	36
	A12	30	39	17	19	34	23	31	19	24	26	16	25	25	16	39
	A13	26	34	12	14	31	19	28	15	20	21	12	21	21	12	34
	A14	22	25	17	12	27	17	22	19	17	5	19	16	18	5	27
	A15	30	39	17	19	34	23	31	19	24	26	16	25	25	16	39
	A16	31	40	17	20	34	23	34	22	24	26	16	25	26	16	40
	A17	21	22	11	8	26	14	24	14	15	7	12	15	16	7	26
	A18	31	40	17	20	34	23	34	22	24	26	16	25	26	16	40
	A21	23	26	9	13	21	11	11	11	15	19	10	14	15	9	26
	A22	17	28	9	9	18	14	23	9	14	22	13	14	16	9	28
	A23	19	34	9	10	22	15	26	17	17	21	14	16	18	9	34
	A24	34	17	33	32	35	31	33	27	31	21	32	30	30	17	35
	A25	33	13	33	27	34	31	26	17	30	22	30	28	27	13	34
	A26	20	23	14	13	22	13	27	18	15	5	18	15	17	5	27
	A27	17	34	11	11	16	14	20	16	14	21	14	12	17	11	34
	A28	25	38	14	14	31	19	34	21	21	24	26	24	24	14	38
	A31	33	44	15	19	32	21	30	29	22	24	16	22	25	15	44
	A32	31	37	13	18	31	19	24	16	20	21	16	20	22	13	37
	A33	28	34	9	17	21	14	20	15	16	22	12	14	18	9	34
	A34	34	48	13	17	28	19	21	32	20	24	19	20	25	13	48
	A35	26	41	10	13	18	14	16	26	16	22	11	13	19	10	41
	A36	33	40	15	21	31	21	26	19	22	25	18	22	24	15	40
	A37	38	52	17	21	34	23	25	35	24	27	22	25	29	17	52
	A38	39	24	44	35	45	42	48	49	41	32	43	38	40	24	49
	A41	31	11	28	20	31	27	20	25	25	18	25	23	23	11	31
	A42	23	25	17	12	26	17	24	27	17	5	15	17	19	5	27
	A43	46	26	44	37	43	43	32	35	40	35	37	39	38	26	46
	A44	39	16	39	31	39	37	30	30	36	28	34	33	33	16	39
	A45	36	12	33	27	34	32	30	27	31	22	30	30	29	12	36
	A46	47	20	44	39	41	43	37	31	41	34	39	40	38	20	47
	A47	42	15	39	34	35	37	34	28	36	28	34	34	33	15	42
	A48	36	12	33	27	34	32	30	27	31	22	30	30	29	12	36
	A51	35	12	33	28	32	32	27	21	30	22	30	29	28	12	35
	A52	25	12	29	28	30	27	29	0	27	0	29	0	26	12	30
	A53	45	26	44	37	42	43	35	35	41	34	38	39	38	26	45
	A54	42	28	44	34	44	43	36	42	42	35	38	37	39	28	44
	A55	43	20	44	36	44	42	37	33	41	33	39	39	38	20	44
	A56	43	20	44	36	44	42	37	33	41	33	39	39	38	20	44
	A57	23	36	12	15	30	17	32	31	20	20	21	20	23	12	36
	A58	42	18	44	36	44	42	33	31	41	33	37	38	37	18	44
	A91	39	24	44	35	45	42	48	49	41	32	43	38	40	24	49
	A92	35	15	39	33	40	36	37	26	36	26	37	35	33	15	40

data type	reflectance						continuum-removed								
library	6		10		16		6		10		16				
spectrum range	7	126	7	126	7	126	7	126	7	126	7	126	avg.	min.	max.
B11	34	44	17	19	34	24	34	25	24	27	19	26	27	17	44
B12	33	37	14	18	29	16	21	15	19	23	15	20	22	14	37
B13	26	32	14	17	27	14	26	18	19	22	16	21	21	14	32
B14	19	16	26	24	12	22	15	6	23	25	18	16	19	6	26
B15	23	28	5	8	18	6	16	10	11	13	8	8	13	5	28
B16	20	15	15	9	21	13	20	18	15	7	21	11	15	7	21
B17	26	29	11	12	24	10	19	14	14	13	12	14	16	10	29
B18	32	37	13	19	31	19	29	23	20	21	18	21	23	13	37
B21	29	36	17	18	34	23	36	25	23	26	30	25	27	17	36
B22	27	40	17	17	34	22	41	33	23	26	33	26	28	17	41
B23	29	36	17	18	34	23	36	25	23	26	30	25	27	17	36
B24	27	35	17	17	34	23	37	23	23	26	31	26	27	17	37
B25	27	34	17	17	34	23	32	20	23	27	29	26	26	17	34
B26	30	36	17	18	34	23	31	24	23	26	30	26	26	17	36
B27	28	36	17	20	34	22	44	37	24	26	33	27	29	17	44
B28	33	37	17	22	34	23	32	23	24	26	23	25	27	17	37
B31	29	42	17	17	34	23	36	26	24	27	28	27	27	17	42
B32	27	43	41	42	35	40	33	40	41	43	33	42	38	27	43
B33	17	26	12	12	13	14	20	10	15	23	10	12	16	10	26
B34	17	31	7	11	21	13	18	11	15	20	11	15	16	7	31
B35	20	37	9	9	20	14	27	24	15	21	22	15	19	9	37
B36	34	40	47	49	43	45	33	42	47	49	42	46	43	33	49
B37	20	7	24	21	20	21	19	17	22	17	21	16	19	7	24
B38	23	25	28	28	11	26	22	18	28	33	24	21	24	11	33
B41	32	34	17	17	34	25	18	20	24	27	23	26	25	17	34
B42	28	37	17	17	34	23	34	23	24	27	30	27	27	17	37
B43	29	36	17	18	34	24	31	20	24	26	27	27	26	17	36
B44	28	37	17	17	34	23	34	23	24	27	30	27	27	17	37
B45	26	38	17	17	34	23	35	25	23	26	32	27	27	17	38
B46	24	32	13	14	29	16	28	17	19	23	23	22	21	13	32
B47	23	34	13	14	29	15	31	21	19	22	26	22	22	13	34
B48	27	32	17	18	34	23	31	17	24	26	27	27	25	17	34
B51	31	21	33	29	36	31	33	38	31	20	34	29	30	20	38
B52	28	36	11	14	26	17	26	17	18	23	13	18	21	11	36
B53	28	26	10	13	29	16	26	18	16	15	12	18	19	10	29
B54	32	37	17	20	34	24	34	22	24	27	18	26	26	17	37
B55	33	37	17	20	34	24	32	21	24	27	20	26	26	17	37
B56	24	17	22	18	29	22	26	25	21	9	23	20	21	9	29
B57	25	30	9	9	27	15	25	18	15	13	12	16	18	9	30
B58	22	34	9	11	23	16	26	16	17	22	15	16	19	9	34
C11	36	42	17	22	34	23	35	26	24	26	20	25	27	17	42
C12	20	33	17	18	10	17	19	14	19	25	14	11	18	10	33
C13	22	29	9	15	17	14	22	10	16	22	11	12	17	9	29
C14	19	31	22	23	9	21	20	19	23	29	19	16	21	9	31

data type	reflectance						continuum-removed								
library	**6**		**10**		**16**		**6**		**10**		**16**				
spectrum range	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**avg.**	**min.**	**max.**
C15	21	29	26	26	23	24	20	25	24	31	18	18	24	18	31
C16	21	28	24	25	11	22	19	17	24	30	20	17	22	11	30
C17	19	29	9	13	13	10	20	9	13	21	9	9	15	9	29
C18	19	29	21	22	9	20	19	18	20	27	18	14	20	9	29
C21	23	31	27	27	11	24	24	17	27	33	29	21	25	11	33
C22	19	25	24	23	9	22	19	12	23	30	22	17	20	9	30
C23	24	32	9	13	15	11	19	13	12	20	8	9	15	8	32
C24	19	26	23	22	8	21	19	11	23	29	21	16	20	8	29
C25	20	28	17	17	10	17	19	11	19	25	16	12	18	10	28
C26	23	31	12	15	13	14	19	13	17	23	12	10	17	10	31
C27	23	32	23	24	8	20	24	22	23	30	23	17	22	8	32
C28	31	32	38	38	22	37	26	25	37	42	34	30	33	22	42
C31	18	33	17	17	9	17	19	25	19	26	13	13	19	9	33
C32	22	29	27	27	12	26	20	16	27	33	21	20	23	12	33
C33	20	27	21	21	17	15	25	24	15	23	13	19	20	13	27
C34	19	34	9	12	14	11	22	15	12	20	9	9	16	9	34
C35	21	37	9	12	17	14	24	17	16	22	11	11	18	9	37
C36	24	38	9	13	20	14	25	18	16	22	11	13	19	9	38
C37	33	49	17	20	33	23	37	31	24	27	19	25	28	17	49
C38	38	51	17	22	34	24	28	22	24	27	14	25	27	14	51
C41	19	33	10	11	16	14	27	25	14	22	17	10	18	10	33
C42	26	27	30	30	14	29	21	15	28	35	18	22	25	14	35
C43	18	30	12	13	13	15	25	21	15	22	15	10	17	10	30
C44	15	20	10	7	12	11	21	13	11	13	12	1	12	1	21
C45	22	42	9	11	22	15	34	35	16	22	21	15	22	9	42
C46	22	42	9	10	23	15	32	33	17	22	16	15	21	9	42
C47	19	27	12	15	13	15	20	9	16	22	13	10	16	9	27
C48	19	38	9	11	17	14	30	29	15	21	15	11	19	9	38
C51	28	46	12	13	28	19	35	34	20	23	19	20	25	12	46
C52	19	29	14	16	11	15	19	10	18	24	17	10	17	10	29
C53	25	43	11	13	26	17	30	23	18	22	11	18	21	11	43
C54	18	35	12	14	12	11	19	18	16	22	10	9	16	9	35
C55	19	30	17	18	10	16	19	13	19	26	19	12	18	10	30
C56	24	28	27	28	12	26	21	19	27	33	26	20	24	12	33
C57	19	35	9	11	20	10	29	23	16	20	14	14	18	9	35
C58	18	34	13	14	11	13	18	17	16	23	14	10	17	10	34
E11	32	32	26	29	30	23	21	9	19	25	25	25	25	9	32
E12	28	36	8	8	23	14	39	37	12	10	22	18	21	8	39
E13	21	38	9	13	16	12	18	23	13	20	10	10	17	9	38
E14	31	30	15	24	32	20	35	26	22	24	23	23	25	15	35
E15	19	34	14	16	10	14	15	16	13	22	8	9	16	8	34
E16	19	32	17	19	10	15	15	12	13	23	10	11	16	10	32
E17	33	40	35	36	31	31	33	25	26	35	33	31	32	25	40
E18	24	37	12	15	26	12	31	28	18	21	18	18	21	12	37

data type	**reflectance**						**continuum-removed**								
library	**6**		**10**		**16**		**6**		**10**		**16**				
spectrum range	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**	**7**	**126**			
													avg.	**min.**	**max.**
E21	22	31	21	23	12	18	24	17	15	25	17	16	20	12	31
E22	31	43	14	20	31	19	23	20	19	24	16	21	23	14	43
E23	31	43	14	20	31	19	23	20	19	24	16	21	23	14	43
E24	29	42	27	28	31	25	20	21	26	30	25	21	27	20	42
E25	22	34	17	17	9	13	25	30	9	23	10	14	19	9	34
E26	30	43	18	20	27	18	9	28	13	23	18	15	22	9	43
E27	22	29	9	17	20	14	20	9	16	21	6	13	16	6	29
E28	26	39	23	25	23	19	24	8	14	26	21	18	22	8	39
E31	20	35	12	14	13	15	18	18	16	22	7	10	17	7	35
E32	28	30	35	35	18	34	21	18	32	39	20	27	28	18	39
E33	25	30	30	30	13	28	18	19	27	33	21	22	25	13	33
E34	19	33	17	18	10	18	19	27	19	25	13	12	19	10	33
E35	18	36	10	11	16	14	30	33	13	21	19	11	20	10	36
E36	20	38	9	11	18	14	32	35	14	21	20	12	20	9	38
E37	20	30	22	22	10	23	22	24	20	28	10	15	20	10	30
E38	24	29	27	28	12	27	20	22	25	32	11	20	23	11	32
E41	23	39	9	13	20	14	26	26	16	21	12	13	19	9	39
E42	20	38	11	13	15	14	22	25	16	22	11	10	18	10	38
E43	19	31	19	20	10	18	18	21	20	26	15	12	19	10	31
E44	22	42	9	12	22	15	29	32	17	22	13	15	21	9	42
E45	19	34	17	18	10	17	18	19	19	25	15	12	19	10	34
E46	21	39	9	12	18	14	23	26	16	21	10	12	18	9	39
E47	24	45	13	15	27	15	33	36	20	22	15	20	24	13	45
E48	17	37	9	10	15	12	25	25	15	20	10	9	17	9	37
E51	29	31	35	36	18	33	25	21	33	39	33	27	30	18	39
E52	24	28	31	31	14	29	18	18	28	34	21	23	25	14	34
E53	25	28	30	30	15	28	19	22	25	34	10	22	24	10	34
E54	17	33	14	15	10	14	20	25	14	22	10	8	17	8	33
E55	18	36	9	12	15	11	23	29	12	20	9	8	17	8	36
E56	18	36	12	14	13	15	22	24	17	22	13	9	18	9	36
E57	17	35	11	12	13	13	25	28	15	21	11	9	17	9	35
E58	36	30	39	39	22	38	21	26	34	42	18	30	31	18	42
E91	37	43	36	38	37	33	36	24	28	37	36	32	35	24	43
E92	30	35	27	28	27	23	23	11	19	28	24	24	25	11	35
E93	32	42	30	31	33	28	21	15	24	31	26	26	28	15	42
E94	31	38	29	31	26	25	25	23	23	28	23	26	27	23	38
E95	27	34	31	32	18	26	35	22	24	32	25	24	27	18	35
E96	27	43	17	19	27	14	27	34	20	23	18	20	24	14	43
E97	31	36	32	33	26	28	29	18	25	29	26	26	28	18	36
avg.	26	32	20	20	24	21	26	22	22	25	20	20			

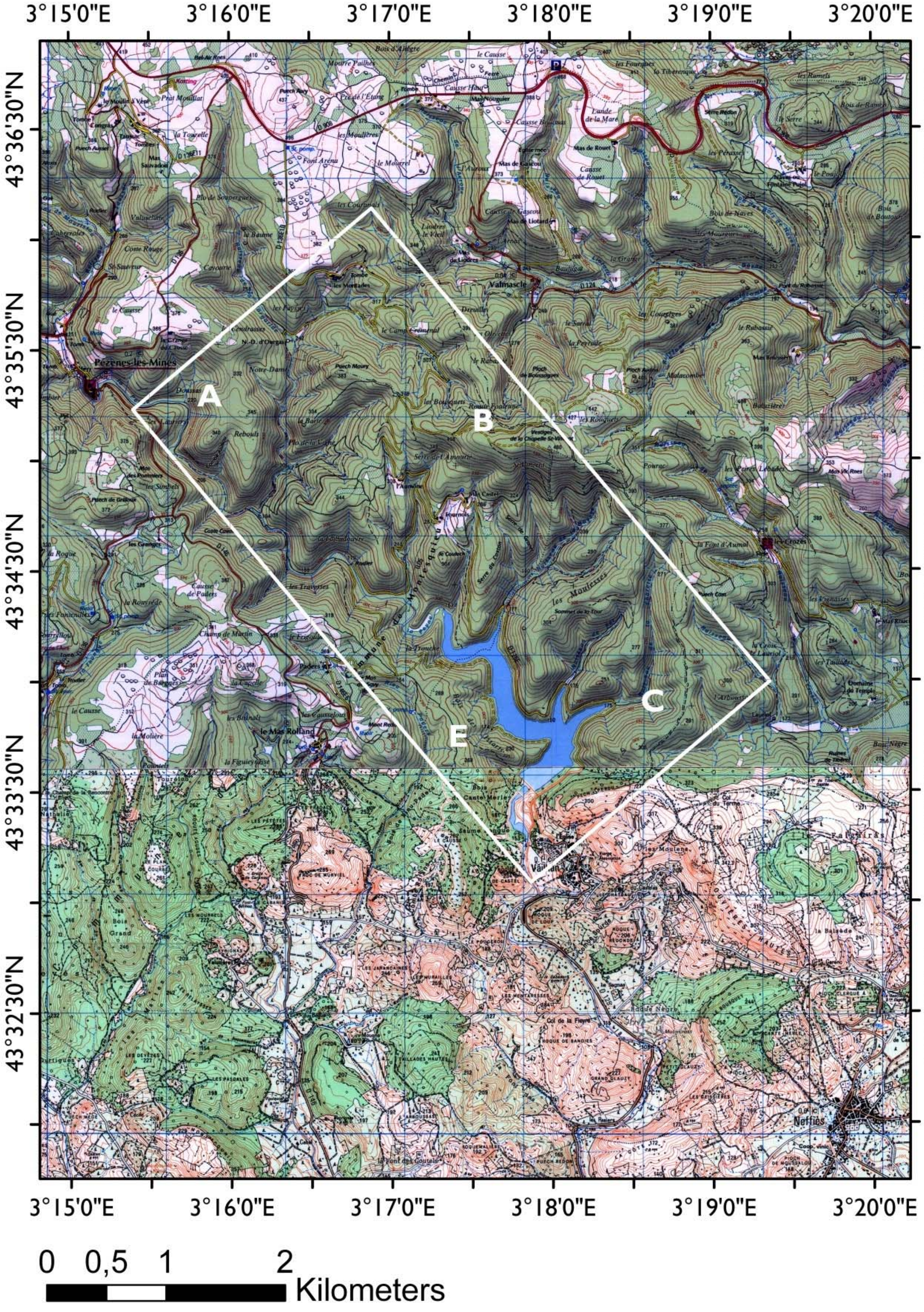
APPENDIX D – CARTOGRAPHY

Figure D.1 – Study area.
Mosaic of the two IGN sheets 2643 OT and 2644 O. Scale 1:25k.
The rectangle indicates the subarea selected for the research.
Letters indicate the strata of the sampling scheme.

Figure D.2 – HyMap image.
A subset of the HyMap dataset is overlaid on top of topographic cartography.
White dots indicate the position of the 169 sampling sites.

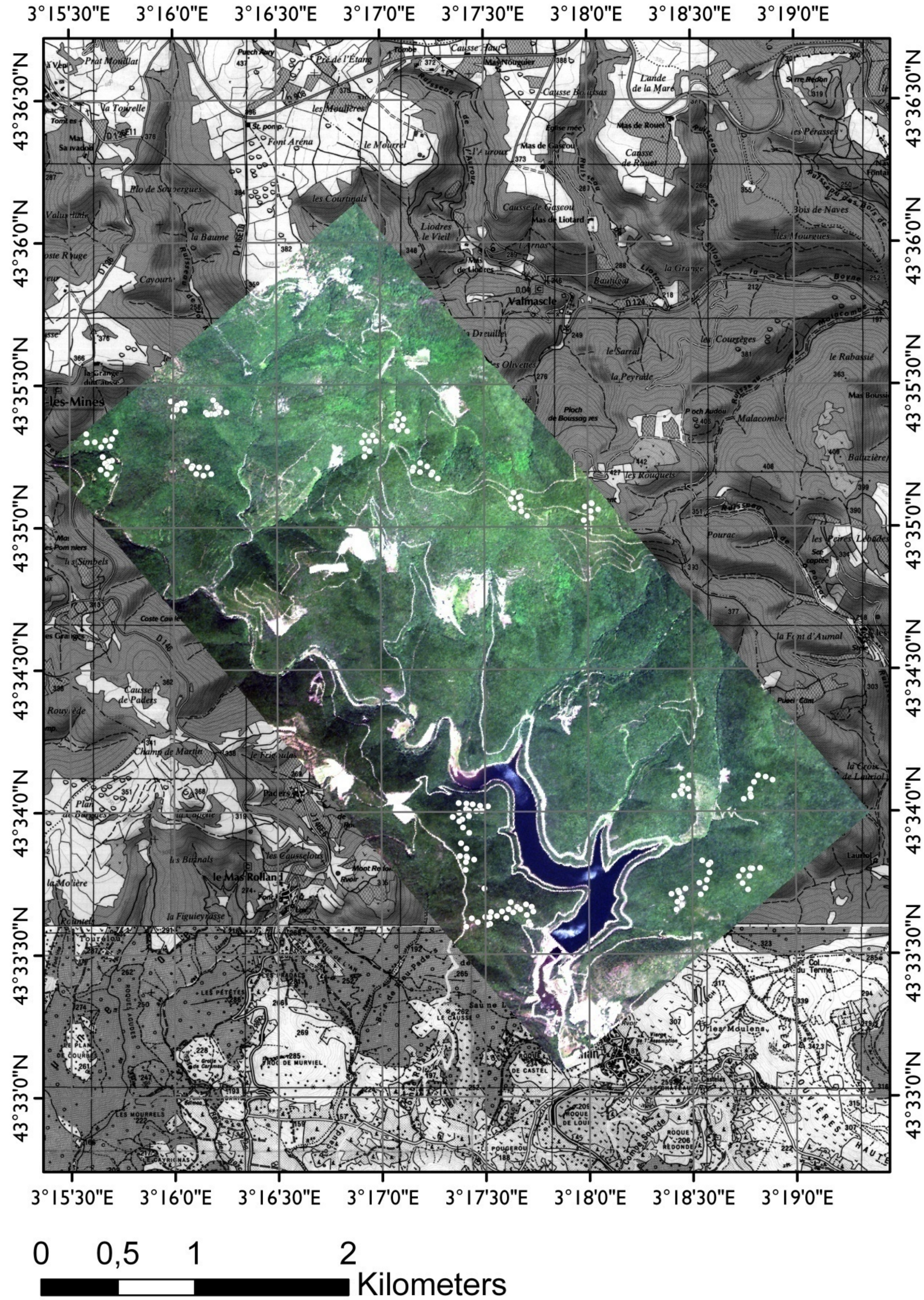
Figure D.3 – Stratified random cluster sampling scheme.
Clusters are indicated with white circles and roman numbers. Circles group sampling sites of the same cluster; sites are identified by numbers centered on their GPS coordinates.

FIGURE D.1 – STUDY AREA



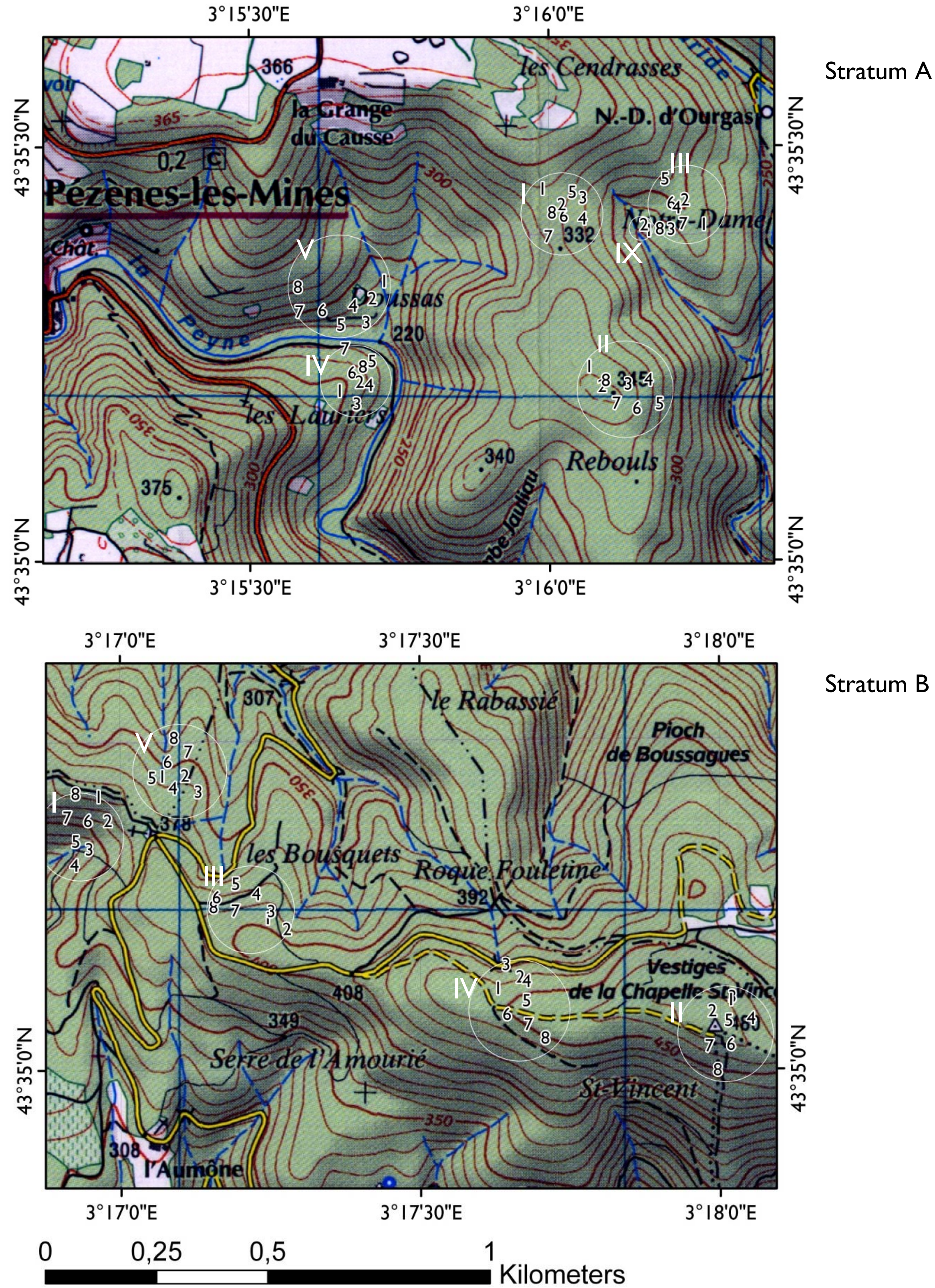
Study area and sampling scheme strata (letters) positioned on 1:25k topographic map.

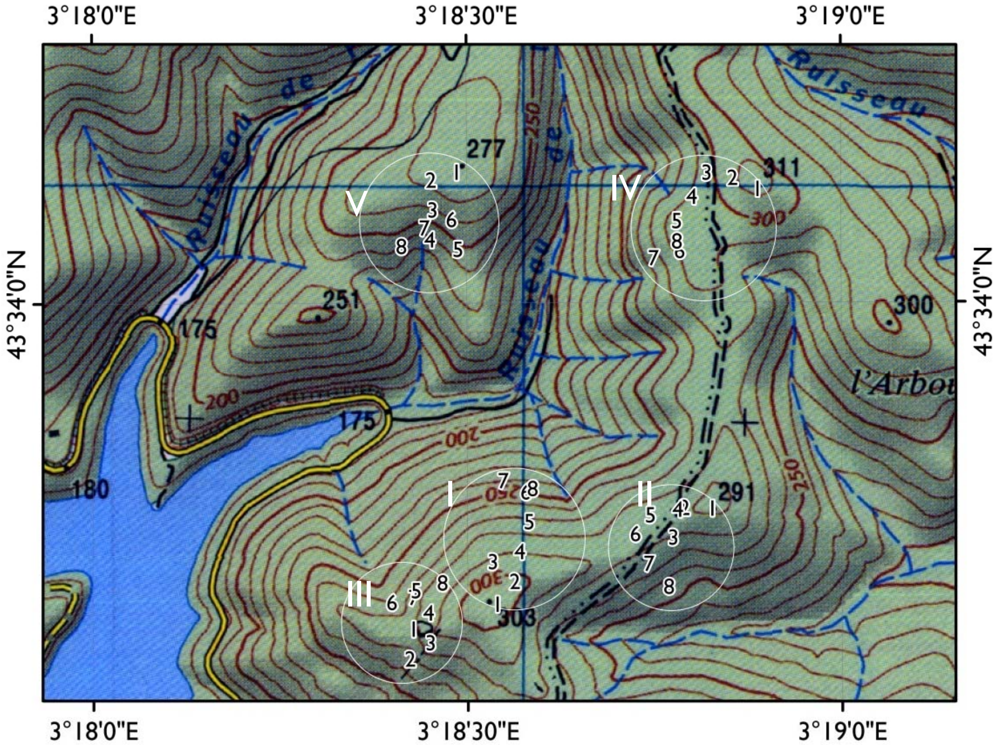
FIGURE D.2 – HYMAP IMAGE



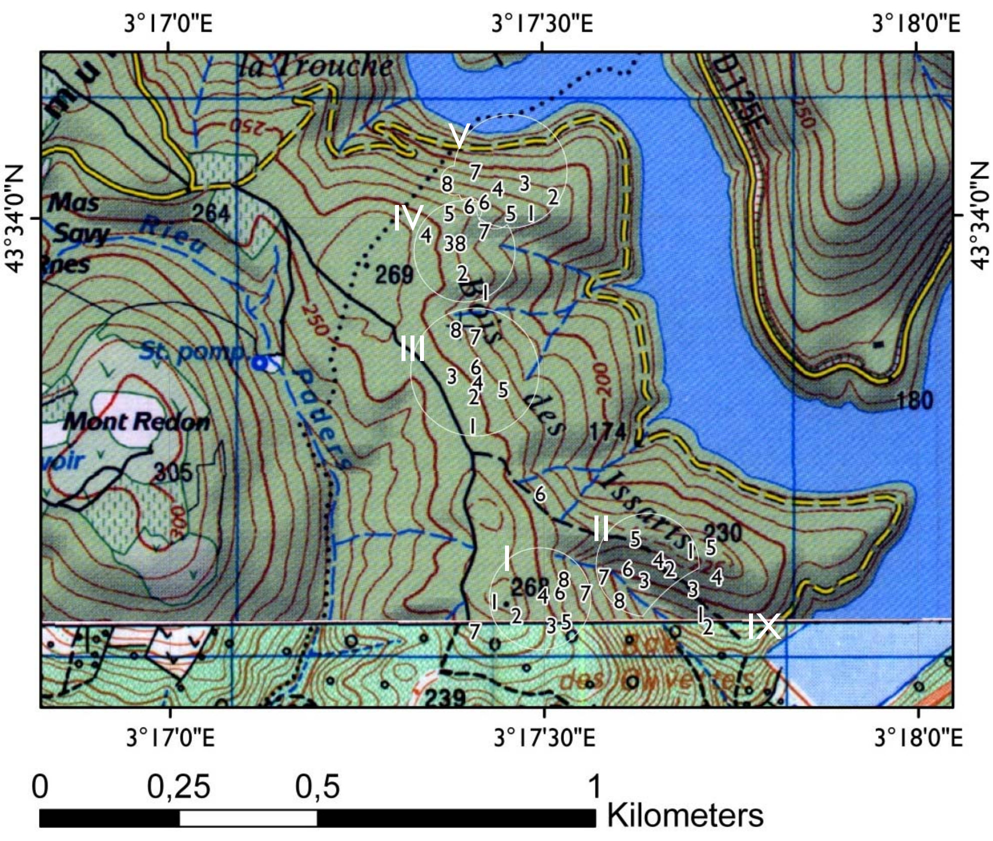
HyMap true-color overlaid on 1:25k topograhic map. White dots represent field plots.

FIGURE D.3 – STRATIFIED RANDOM CLUSTER SAMPLING SCHEME.





Stratum C



Stratum E

(sites left outside the circles belong to cluster 9)