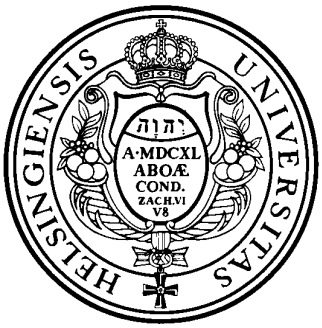


Master's thesis
Geography
Geoinformatics

TREE SPECIES DIVERSITY ESTIMATION USING AIRBORNE IMAGING SPECTROSCOPY

Elisa Schäfer

2014

Supervisors:
Petri Pellikka
Janne Heiskanen

UNIVERSITY OF HELSINKI
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
DIVISION OF GEOGRAPHY

P.O. Box 64 (Gustaf Hällströmin katu 2)
FIN-00014 University of Helsinki

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section	Laitos – Institution – Department	
Faculty of Science	Department of Geosciences and Geography	
Tekijä – Författare – Author		
Elisa Schäfer		
Työn nimi – Arbetets titel – Title		
Tree species diversity estimation using airborne imaging spectroscopy		
Oppiaine – Läroämne – Subject		
Geoinformatics		
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	December 2014	71

Tiivistelmä – Referat – Abstract

With the ongoing global biodiversity loss, approaches to measuring and monitoring biodiversity are necessary for effective conservation planning, especially in tropical forests. Remote sensing is a very potential tool for biodiversity mapping, and high spatial resolution imaging spectroscopy allows for direct estimation of tree species diversity based on spectral reflectance.

The objective of this study is to test an approach for estimating tree species alpha diversity in a tropical montane forest in the Taita Hills, Kenya. Tree species diversity is estimated based on spectral variation of high spatial resolution imaging spectroscopy data. The approach is an unsupervised classification, or clustering, applied to objects that represent tree crowns.

Airborne imaging spectroscopy data and species data from 31 field plots were collected from the study area. After preprocessing of the spectroscopic imagery, a minimum noise fraction (MNF) transformation with a subsequent selection of 13 bands was applied to the data to reduce its noise and dimensionality. The imagery was then segmented to obtain objects that represent tree crowns. A clustering algorithm was applied to the segments, with the aim of grouping spectrally similar tree crowns. Experiments were made to find the optimal range for the number of clusters.

Tree species richness and two diversity indices were calculated from the field data and from the clustering results. The clusters were assumed to represent species in the calculations. Correlation analysis and linear regression analysis were used to study the relationship between diversity measures from the field data and from the clustering results.

It was found that the approach succeeded well in revealing tree species diversity patterns with all three diversity measures. Despite some factors that added error to the relationship between field-derived and clustering-derived diversity measures, high correlations were observed. Especially tree species richness could be modelled well using the approach (standard error: 3 species). The size of the considered trees was found to be an important determinant of the relationships. Finally, a tree species richness map was created for the study area.

With further development, the presented approach has potential for other interesting applications, such as estimation of beta diversity, and tree species identification by linking the reflectance properties of individual crowns to their corresponding species.

Avainsanat – Nyckelord – Keywords

remote sensing, imaging spectroscopy, biodiversity, clustering, Taita Hills

Säilytyspaikka – Förvaringställe – Where deposited

Kumpula Science library, University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section	Laitos – Institution – Department
Matemaattis-luonnontieteellinen tiedekunta	Geotieteiden ja maantieteen laitos

Tekijä – Författare – Author
Elisa Schäfer

Työn nimi – Arbetets titel – Title
Tree species diversity estimation using airborne imaging spectroscopy

Oppiaine – Läroämne – Subject
Geoinformatiikka

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Pro gradu -tutkielma	Joulukuu 2014	71

Tiivistelmä – Referat – Abstract

Luonnon monimuotoisuuden maailmanlaajuisen vähenemisen vuoksi biodiversiteetin mittaus- ja tarkkailumenetelmiä tarvitaan tehokkaaseen suojelualueiden suunnitteluun, erityisesti trooppisissa metsissä. Kaukokartoitus on erittäin lupaava väline biodiversiteetin kartoitukseen, ja spatiaalisesti tarkka hyperspektraalinen aineisto (kuvantava spektroskopia) mahdollistaa puiden lajidiversiteetin arvioinnin suoraan niiden spektraalisen heijastuksen perusteella.

Tämän tutkimuksen tarkoitus on kokeilla lähestymistapaa puulajien alfadiversiteetin mittaamiseen trooppisessa vuoristometsässä Kenian Taitavuorilla. Puulajien monimuotoisuutta arvioidaan spatiaalisesti tarkan hyperspektraalisen aineiston heijastuksen vaihtelun avulla. Lähestymistapa on puunlatvuksia edustaville kohteille tehty ohjaamaton luokittelu, tarkemmin ilmaistuna klusterointi.

Tutkimusalueelta kerättiin hyperspektraalista ilmakuva-aineistoa sekä puulajitiedot 31 maastokoealalta. Hyperspektraalisen aineiston esikäsittelyn jälkeen sen hälyä ja ulottuvuuksia vähennettiin tekemällä sille MNF (minimum noise fraction) –muunnos ja valitsemalla 13 parasta kanavaa. Tämän jälkeen ilmakuva segmentoitiin puunlatvuksia kuvaaviksi kohteiksi. Kohteet klusteroitiin klusterointialgoritmia käyttäen, tarkoituksena ryhmitellä spektraalisesti samankaltaiset puunlatvukset. Ihanteellisen klusterimäärän löytämiseksi tehtiin kokeiluja.

Puulajirunsaus ja kaksi diversiteetti-indeksiä laskettiin maastoaineistolle ja klusteroinnin tuloksille. Klustereiden oletettiin edustavan puulajeja laskelmissa. Maastoaineistosta ja klusterointituloksista laskettujen diversiteettimittareiden suhdetta tutkittiin korrelaatioanalyysin ja lineaarisen regressioanalyysin avulla.

Lähestymistapaa soveltaen onnistuttiin hyvin paljastamaan puulajien monimuotoisuuden piirteitä kaikkien kolmen diversiteettimittarin avulla. Huolimatta tekijöistä, jotka aiheuttivat virhettä maastoaineistoon ja klusterointituloksiin perustuvien diversiteettimittareiden suhteeseen, korrelaatioasteet olivat korkeita. Varsinkin puiden lajirunsautta pystyttiin mallintamaan hyvin lähestymistavan avulla (keskivirhe: kolme lajia). Mukaanluettujen puiden koko oli tärkeä tekijä muuttujien suhteissa. Lopuksi tehtiin kartta puulajirunsaudesta tutkimusalueelle.

Jatkokehittämisen avulla esitellyllä lähestymistavalla on mahdollisuuksia muihinkin mielenkiintoisiin sovelluksiin, kuten betadiversiteetin arvioimiseen, sekä puulajien tunnistukseen, kun yksittäisten latvusten heijastusominaisuudet liitetään niitä vastaaviin lajeihin.

Avainsanat – Nyckelord – Keywords
Kaukokartoitus, kuvantava spektroskopia, hyperspektraalinen, biodiversiteetti, klusterointi, Taitavuoret

Säilytyspaikka – Förvaringställe – Where deposited
Kumpulan kampuskirjasto

Muita tietoja – Övriga uppgifter – Additional information

CONTENTS

LIST OF FIGURES III
LIST OF TABLES V
LIST OF ABBREVIATIONS VI
1 INTRODUCTION 1
1.1 Biodiversity research in the tropics 1
1.2 Context of the study 2
1.3 Objectives 2
2 BACKGROUND 3
2.1 Measures of biodiversity 3
2.2 Physical background of optical remote sensing of vegetation 4
2.2.1 Atmospheric effects on remote sensing 4
2.2.2 Reflectance characteristics of vegetation and the separability of species 6
2.3 Background of image analysis techniques 7
2.3.1 Imaging spectroscopy and feature extraction methods 7
2.3.2 Supervised vs. unsupervised classification 8
2.4 Remote sensing of biodiversity 9
3 STUDY AREA 11
4 DATA 13
4.1 Imaging spectroscopy data 13
4.2 Field data 15
4.2.1 Data collection 15
4.2.2 Diversity measures from field data 17
4.3 Additional data 20
5 METHODS 21
5.1 Preprocessing 21
5.1.1 Radiometric calibration 21
5.1.2 Atmospheric correction 23
5.1.3 Georectification 25
5.1.4 Study area mosaicking and delineation 27
5.1.5 Shadow removal 28

5.2 Feature extraction using Minimum Noise Fraction transformation 29
5.3 Tree crown segmentation 31
5.4 Spectral clustering of segments 32
5.4.1 Algorithm *clara* 32
5.4.2 Implementation 33
5.5 Calculation of biodiversity indices from clustering results 33
5.6 Analyses 34
5.7 Tree species richness map 35
6 RESULTS 35
6.1 Segmentation and clustering 35
6.1.1 Segmentation 35
6.1.2 Spectral differences between segments 36
6.1.3 Number of segments on plots 38
6.1.4 Clustering 38
6.2 Diversity measures from clustering 40
6.3 The effect of number of clusters on the relationship between biodiversity measures 43
6.4 Modelling of species diversity measures 48
6.5 Tree species richness map of Ngangao 53
7 DISCUSSION 55
7.1 Performance of the approach in estimating tree species diversity 55
7.2 The approach in context of previous research in the field 56
7.3 Evaluation of factors that affected tree diversity measures 57
7.3.1 Optimal number of clusters 57
7.3.2 Optimal tree size 58
7.3.3 Sources of error in the relationships between field data and clustering results . 59
7.3.4 Factors that affected species discrimination 60
7.4 Diversity measures as indicators of conservation value 63
8 CONCLUSIONS 65
9 ACKNOWLEDGEMENTS 66
10 REFERENCES 67

LIST OF FIGURES

Figure 2-1. Radiance reflected by the target of interest (A), scattered from neighbouring targets (B) and scattered from the atmosphere (C). 5
Figure 2-2. Reflectance spectrum of a vegetation target in the study area. 6
Figure 2-3. Forest canopy viewed from a cliff in in the study area. 7
Figure 3-1. Maps of Africa, Kenya and the Taita Hills. 12
Figure 3-2. Aerial image of the study area, the Ngangao forest. 13
Figure 4-1. Field plot locations in the study area. 16
Figure 4-2. Species abundance distribution, with the five most common species named. 17
Figure 4-3. Simpson's index values of the field plots. Plot names are shown only for the plots that got low index values. 19
Figure 4-4. Shannon–Wiener index values of the field plots. Plot names are shown only for the plots with highest and lowest values. 19
Figure 4-5 a – c. Species accumulation curves for a) all measured trees, b) 50 % of largest trees and c) 25 % of largest trees. 20
Figure 5-1. The spectral profile of a vegetation pixel before radiometric calibration. The values are unitless digital numbers. Note the wavelength axis running from larger to smaller values. 22
Figure 5-2 The spectral profile of a vegetation pixel after radiometric calibration. The spectrum exhibits absorption features due to gases and water vapour in the atmosphere. The units are radiance, (mW/cm^2*sr* µm) * 1000. 22
Figure 5-3. The full width at half maximum values for each band in the 4x spectral binning mode, based on laboratory measurements by Specim Ltd.. 23
Figure 5-4. The response curves of bands 22 to 29. As can be seen, there are slight differences between the curves. The wavelength is unit µm. 23
Figure 5-5. The spectral profile of a vegetation pixel after atmospheric correction. The units are reflectance values, percentage multiplied by 100. 25
Figure 5-6 a – b. A part of the image a) before and b) after georectification. The images are false-colour infrared compositions of bands 90, 65 and 39 (center wavelengths at around 812, 693, and 572 nm, respectively). 26
Figure 5-7 a – b. a) Edge between two flightlines in the image mosaic (bands 90, 65, 39). b) The same area in the canopy height model, where relatively higher areas have a lighter shade. 27

Figure 5-8 a – b. a) The study area delineated with the land cover classification; RGB bands: 90, 65, 39 of the spectroscopic image. b) The study area after removing shadowed pixels and areas with height less than 7 m, and manual cleaning of edges. 28
Figure 5-9. The MNF transformed image (MNF bands 1, 2 and 3). 29
Figure 5-10. Eigenvalues for the 129 MNF bands. 30
Figure 5-11 a – f. A part of the image depicted with MNF bands 11–16. 30
Figure 5-12. Some of the segments selected for comparison (white outlines). The pine and green segments appear visually similar, but the two cypress plots have segments of somewhat different colours. 32
Figure 6-1. a) A part of the MNF image and b) the segments created from it. 35
Figure 6-2 a – d. The mean MNF values for each crown show clustering patterns. The different groups of trees are distinguishable even at the higher MNF bands (figure d). 36
Figure 6-3 a – f. MNF values plotted for each group of trees. The value is the mean MNF value of the segment pixels. 37
Figure 6-4. Number of segments and trees on the plots. The boxes represent half of the 31 plots and the line in the box corresponds to the mean. The whiskers show the range of the values and the points are outliers. 38
Figure 6-5. Examples of clustering results with a) k = 50 and b) k =10. 39
Figure 6-6 a) Clustering results on some plots shown on the MNF transformed image. b) The MNF transformed image of the same area. 40
Figure 6-7. Maximum, mean, and minimum species richness values for clustering results with different k values. The dark blue dots show the same for the field data. 41
Figure 6-8. Simpson's index values for clustering results with different k values. 42
Figure 6-9. Shannon–Wiener index values for clustering results with different k values. 42
Figure 6-10 a – c. The effect of k on the correlation between species richness values from field data and from clustering. 44
Figure 6-11 a – c. The effect of k on the correlation between the Simpson's index values from field data and from clustering. 45
Figure 6-12 a – c. The effect of k on the correlation between the Shannon–Wiener index values from field data and from clustering. 46
Figure 6-13 a – c. Effect of k on correlations between species richness from field data and from mean of 100 clustering results. 47
Figure 6-14 a – c. Effect of k on correlations between Simpson's index from field data and from mean of 100 clustering results. 47

Figure 6-15 a – c. Effect of k on correlations between the Shannon–Wiener index from field data and from mean of 100 clustering results. 48
Figure 6-16. The relationship of species richness obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered. 51
Figure 6-17. The relationship of Simpson's index obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered. 52
Figure 6-18. The relationship of the Shannon–Wiener index obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered. 52
Figure 6-19. Tree species richness in the study area as the mean prediction of 5 clustering results. 53
Figure 6-20. Canopy height model of the study area. 54
Figure 8-1. From left to right, the figure shows a fraction of the study area a) as a true-colour composition of the spectroscopic image, b) after shadow removal and MNF transformation, c) after spectral segmentation, d) with clustering results, e) with predicted species richness. 65

LIST OF TABLES

Table 4-1. Sensor characteristics and configurations for the flight campaign. 14
Tables 4-2 a – c. Species richness of the plots with different tree sizes. 18
Table 5-1. Naming of diversity variables from clustering results (3 x 79). 34
Table 5-2. Naming of diversity variables from second round of clustering (3 x 16). 34
Table 6-1. Coefficients of linear regression models for species richness with all measured trees. 49
Table 6-2. Coefficients of linear regression models for species richness with 50 % of largest trees. 49
Table 6-3. Coefficients of linear regression models for species richness with 25 % of largest trees. 50

LIST OF ABBREVIATIONS

AISA	Airborne Imaging Spectrometer for Applications
BIL	Band Interleaved by Line
BIODEV	Building Biocarbon and Rural Development in West Africa
CBD	Convention for Biological Diversity
CCD	Charge-Coupled Device
CHM	Canopy Height Model
DBH	Diameter at Breast Height
DDV	Dense dark vegetation
DN	Digital Number
FODIS	Fibre Optic Downwelling Irradiance Sensor
FOV	Field of View
FWHM	Full Width at Half Maximum
GPS	Global Positioning System
IMU	Inertial Measurement Unit
LIDAR	Light Detection And Ranging
LUT	Look-up table
MNF	Minimum Noise Fraction (transformation)
NIR	Near-Infrared
PCA	Principal Component Analysis
REDD+	Reducing Emissions from Deforestation and Forest Degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries
RMSE	Root Mean Square Error
SWIR	Short-Wave Infrared
UNFCCC	United Nations Framework Convention on Climate Change

1 INTRODUCTION

1.1 Biodiversity research in the tropics

Biodiversity is defined as "the variability among living organisms from all sources (…) and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems" (CBD 1992). It is essential for the functioning of ecosystems, and thus ecosystem services such as hydrological and climatic regulation (Duffy 2009; Hector & Bagchi 2007). Despite initiatives such as the Convention on Biological Diversity (CBD), global biodiversity is rapidly declining (Butchart *et al.* 2010).

Most biodiversity on Earth is found in the tropics (Gaston 2000). This is reflected by the locations of biodiversity hotspots (Myers *et al.* 2000), areas that have a high number of endemic species and are also highly endangered. However, tropical ecosystems, and especially tropical forests, are poorly known to science in terms of their vegetation diversity and its significance to the ecosystem (Milliken *et al.* 2010). The lack of research is understandable, because tropical forests pose a very challenging environment for research with their climate, inaccessibility and vast amount of diversity. Adequate information on biodiversity, however, is essential for effective conservation planning (Nagendra *et al.* 2013).

Remote sensing has been recognized as a very potential tool for assessing and monitoring biodiversity (e.g. Nagendra *et al.* 2013; Kuenzer *et al.* 2014; Pettorelli *et al.* 2014; Gillespie *et al.* 2008; Turner *et al.* 2003). It enables data collection from large areas that may otherwise be inaccessible, and the development of technologies such as imaging spectroscopy allows new approaches for biodiversity studies.

REDD+ is an initiative by UNFCCC (United Nations Framework Convention on Climate Change) to financially compensate developing countries for the protection of their forests (REDD = Reducing Emissions from Deforestation and Forest Degradation). To date, biodiversity is recognized as an important co-benefit in the protection of tropical forest carbon stocks (Gardner *et al.* 2012), but the tools for carbon monitoring in REDD+ projects often do not include ways to estimate biodiversity (Imai *et al.* 2014). Monitoring based on remote sensing, however, could serve the needs of both carbon and biodiversity assessment (Imai *et al.* 2014).

1.2 Context of the study

In this thesis, the use of airborne imaging spectroscopy is studied for assessment of tree species diversity in a tropical montane forest. The study is part of the BIODEV research project (Building Biocarbon and Rural Development in West Africa) funded by the Ministry of Foreign Affairs in Finland. The project aims to link climate change and mitigation strategies to enhance the livelihoods in rural communities, and research in the Taita Hills serves for testing the methodology applied in the target countries in West Africa.

One of the project aims is to develop methods for carbon measurement that simultaneously provide information on biodiversity and other ecosystem services. The field data of this study was also used for remote sensing based estimates of carbon stocks in the BIODEV research project, and thus it serves as an example of combining carbon and biodiversity assessment.

1.3 Objectives

Motivated by previous research, the objective of this study is to test an approach for mapping tree species alpha diversity in a tropical montane forest in the Taita Hills, Kenya. Tree species diversity is predicted based on spectral variation of high spatial resolution imaging spectroscopy data. The approach is an unsupervised classification, or clustering, applied to objects that represent tree crowns.

The research questions are the following.

1. How accurately can we estimate tree species diversity measures based on spectral differences between tree crowns?

2. How does tree species richness vary spatially in the Ngangao forest in the Taita Hills?

2 BACKGROUND

2.1 Measures of biodiversity

Much of the research in biodiversity is focused on species diversity, because it is the taxonomic level that is best defined for many organisms (Krebs 2014). Species diversity studies, in turn, usually consider only one taxon, such as trees, birds, or butterflies. However, it is important to keep in mind that species diversity in one taxon forms only part of the biodiversity of the area of interest.

Ecologists often differentiate between diversity on a local scale (alpha diversity) and on a regional scale (gamma diversity). The former contains the diversity of a community; the latter comprises many communities in a larger area. Beta diversity is a measure of how different communities are within an area or along an environmental gradient, and thus links local and regional diversity (Krebs 2014).

Species diversity itself can be broken down to two components: species richness and evenness. Species richness is the simplest measure of diversity, and is simply the number of species on the area of interest. Evenness is the measure of how equally abundant species are in a community (Krebs 2014). If the community is dominated by a few abundant species, its diversity is lower than that of a community where abundances are more equal.

While species richness is probably the most common measure of biodiversity, indices have been developed to measure other aspects of it. Two of them are used in this study: Simpson's index and the Shannon–Wiener index. While Simpson's index is more sensitive to changes in the abundant species of the community, the Shannon–Wiener index better measures the changes in the rare species.

Simpson's index of diversity expresses the probability that two randomly selected organisms are different species. For a finite population, sampling without replacement should be assumed. Then the index is calculated as

$$D = \sum\left[\frac{n_i(n_i-1)}{N(N-1)}\right] \qquad \text{Equation 1.}$$

where n_i is the number of individuals of species *i* in the community, N is the total number of individuals in the sample, and D is the index proposed by Simpson (1949). However, the complement (1 – D) of Simpson's original measure is most often used (Krebs 2014), also in

this study. The index will get values close to 0 when species in the community are not evenly abundant, and values close to 1 occur when the species are similar in abundance.

The Shannon–Wiener index expresses how difficult it is to predict the species of the next individual collected. It is based on information theory (Shannon 1948) and according to (Krebs 2014) derived independently by Shannon and Wiener. The index is calculated as

$$H' = -\sum_{i=1}^{s}(p_i)(\log_2 p_i) \qquad \text{Equation 2.}$$

where *s* is the number of species, and p_i is the proportion of the sample belonging to the *i* th species. The index is a measure of uncertainty in the prediction and its value increases with the number of species in the community. In theory it can reach very large values, but in practice does not exceed five in biological communities.

2.2 Physical background of optical remote sensing of vegetation

2.2.1 Atmospheric effects on remote sensing

Passive optical remote sensing provides information of ground targets by measuring the electromagnetic radiance they reflect. Remote sensors can be mounted on aircrafts (airborne) or satellites (spaceborne). Sensors record radiation as digital numbers (DN's) which can be converted back to radiance units, Watts per square meter per steradian (W m^{-2} sr^{-1}).

When the amount of incoming solar radiation (irradiance) is known, the radiance values of ground targets can be further converted to reflectance values. Reflectance is the ratio of radiant exitance with the irradiance (Schaepman-Strub *et al.* 2006), or in other words, the proportion of incoming radiation that a surface reflects back. It varies with wavelength because surfaces selectively absorb incoming radiation. It also depends on the illumination and view angles, as surfaces may have different reflectance in different directions, described by their bidirectional reflectance distribution function (BRDF). A hypothetical surface that has the same reflectance regardless of view angle is called a Lambertian surface. The reflectance of natural surfaces, such as vegetation canopies, provides information on their properties.

The signal that remote sensors aim to measure is the radiance that originates in the sun and is reflected by ground targets (path A in figure 2-1). However, also radiance from other sources reaches the sensors. The most important of these are the radiance scattered from the atmosphere (path radiance, path C in figure 2-1) and from neighbouring targets (adjacency effect, path B in figure 2-1 (Jones & Vaughan 2010)). Scattering in the atmosphere occurs due to gas molecules and larger particles such as aerosols and atmospheric water.

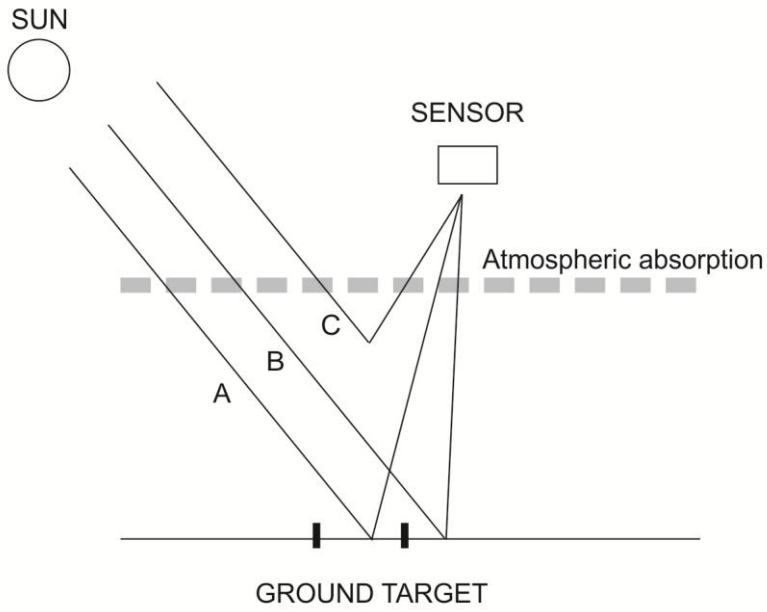


Figure 2-1. Radiance reflected by the target of interest (A), scattered from neighbouring targets (B) and scattered from the atmosphere (C).

Some of the radiation is absorbed by the atmosphere when it comes from the sun and again when it is reflected or scattered. The main absorbing gases in the atmosphere are water vapour (H_2O), carbon dioxide (CO_2), oxygen (O_2), ozone (O_3) and nitrous oxide (N_2O) (Jones & Vaughan 2010). All of these absorb radiation selectively depending on the wavelength, and therefore the amount of radiance that reaches the ground or the sensor at different wavelengths is not equal to the radiance that is emitted be the sun.

To obtain the reflectance of ground targets, the effects of path radiance, adjacency effects and atmospheric absorption have to be accounted for. This is the aim of atmospheric corrections to remotely sensed images.

2.2.2 Reflectance characteristics of vegetation and the separability of species

The spectral reflectance of vegetation canopies is determined by leaf biochemistry, leaf and canopy structure, and the properties of the background of the canopy (if visible). These vary with the species present, vegetation type, the age and health status of the plants, and seasonal changes in phenology. Leaf biochemical composition has been shown to be often unique for different species, and it causes fine-scale differences in the reflectance of species (Asner & Martin 2009).

In general, leaf reflectance in the visible portion of the spectrum is low due to absorption by photosynthetic pigments (figure 2-2). In the near-infrared (NIR) region leaf reflectance is relatively high, with absorption features by leaf water and leaf tissue components (cellulose and lignin). Reflectance in short-wave infrared (SWIR) is relatively low and mostly characterized by water absorption features. At the canopy scale, the NIR reflectance is further increased by multiple scattering of NIR radiation in the canopy. Also non-photosynthetic tissues such as bark and flowers affect the reflectance signal of the canopy. Shadows are an important component of canopy reflectance, and are affected by crown and canopy structure (Clark 2012).

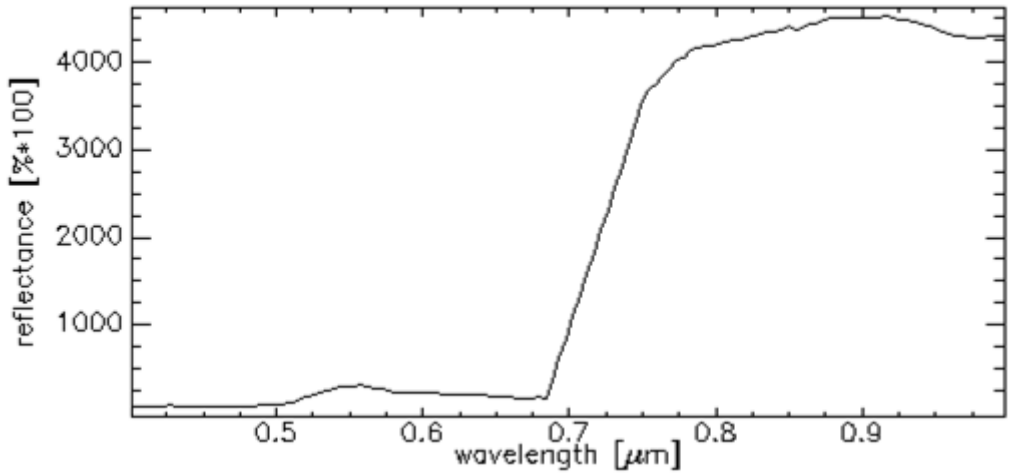


Figure 2-2. Reflectance spectrum of a vegetation target in the study area.

For remote sensing of species diversity an important question is, whether species are separable by their reflectance properties. This requires the differences between species being larger than the variation within species. Tropical forests have characteristics that make species discrimination very challenging: they have a very high number of tree species, tree crowns are

multilayered and intermingled, epiphytes such as lianas are common, and the seasonal phenological changes such as leaf drop and flowering may occur non-synchronized within the same species (Clark 2012). Figure 2-3, which is a photograph from the study area, illustrates the complexity of the canopy, but also the different colours and textures of tree crowns.

Figure 2-3. Forest canopy viewed from a cliff in in the study area.

2.3 Background of image analysis techniques

2.3.1 Imaging spectroscopy and feature extraction methods

Imaging spectroscopy, also known as hyperspectral remote sensing, is a technique for measuring radiance in many narrow bands that continuously cover a proportion of the spectrum (Schaepman *et al.* 2009). Because of the high spectral resolution, it allows much more detailed information to be acquired from targets, such as vegetation, than multispectral remote sensing (e.g. Kuenzer *et al.* 2014; Gillespie *et al.* 2008).

However, spectroscopic imagery also poses new challenges to data analysis compared to multispectral imagery. In addition to the high storage and computing capacities that spectroscopic imagery requires, the main issues are data redundancy and the high dimensionality (Bajwa & Kulkarni 2012). Because there is a high correlation between most of the bands in hyperspectral datasets, their information content is not unique. In addition, as the number of bands in an image increases, it takes more observations to train a classifier and separate classes from each other ("the curse of dimensionality", Bajwa & Kulkarni 2012).

To extract relevant information from spectroscopic imagery, some form of feature selection methods need to be applied. These may be selection of individual bands, calculation of vegetation indices or other transformations to the original data (Bajwa & Kulkarni 2012). In this study, the Minimum Noise Fraction (MNF) transformation is applied for the purpose, motivated by its successful use in other biodiversity studies based on imaging spectroscopy (e.g. Leutner *et al.* 2012; Ghosh *et al.* 2014; Vaglio Laurin *et al.* 2014).

The MNF transformation is based on the Principal Component Analysis (PCA), a common technique for reducing data redundancy and dimensionality. The principal components are linear transformations of the original data, uncorrelated to each other and sorted based on their variance (Bajwa & Kulkarni 2012). It converts the data to as many PC bands as there were original bands, but most of the information content is in the first PCs.

The MNF transformation takes advantage of the PCA by first making a similar transformation to the data, but maximizing the noise content of each component rather than the variance (Green *et al.* 1988; Lee *et al.* 1990). This is done based on an estimated noise-covariance matrix and results in noise-decorrelated components. Taken in reverse order, the components are used in a subsequent PCA. This results in MNF bands that are inversely sorted by their noise content and decorrelated with each other. Examining the eigenvalues of the bands reveals which bands have high information content and are thus useful for further processing.

2.3.2 Supervised vs. unsupervised classification

Classification of image content is commonly used to extract information from remotely sensed data. Supervised classification techniques require training a classification algorithm with reflectance data from known targets. The algorithm is then applied to the whole image to

assign the pixels to classes. The accuracy of the classification has to be validated with samples that are independent from the training data, and it depends on the spectral separability of classes as well as the amount of training samples (Bajwa & Kulkarni 2012).

Supervised classification methods can be useful and accurate especially in the cases when certain, well-defined spectral classes have to be found from remotely sensed data. However, the availability of training and validation data is often a constraint. Unsupervised classification methods are therefore an appealing alternative for image analysis, as they do not require previous knowledge of the image contents.

Unsupervised classification methods are usually based on cluster analysis (or clustering, as referred to hereafter; (Bajwa & Kulkarni 2012)). Clustering groups the data based on the similarity of observations. Clustering methods can be roughly divided to partitional and hierarchical methods (Tan *et al.* 2006). Partitional clustering algorithms, such as k-means, iteratively optimize the division of the observations to a user-defined number of clusters. Hierarchical clustering algorithms produce a nested clustering, where clusters have subclusters.

2.4 Remote sensing of biodiversity

With coarse spatial and spectral resolution data, remote sensing of biodiversity is mostly limited to indirect assessment, meaning modelling based on environmental variables. High spectral and spatial resolution data allow for more direct approaches, such that species or species diversity can be mapped based on spectral reflectance (Gillespie *et al.* 2008; Turner *et al.* 2003).

Mapping e.g. individual tree species using imaging spectroscopy is a typical situation where supervised classification is used. Discrimination of tree species is an active and progressing field of research, and several authors have studied the spectral separability of tropical rain forest tree species on leaf level in laboratory conditions (e.g. Cochrane 2000; Clark *et al.* 2005; Asner *et al.* 2009) and on the canopy level using remote sensing (e.g. Vaglio Laurin *et al.* 2014; Asner *et al.* 2008; Clark & Roberts 2012). Efforts have been made to link also the leaf biochemical properties to their spectral and taxonomic diversity (Asner *et al.* 2009).

Research in this field was reviewed by Clark (2012). Important conclusions were that species discrimination is most successful with hyperspectral data that covers the full range from visible light to SWIR, and that has a high spatial resolution. He also recognized that the temporal domain of differences between species has not been explored yet. Also, as research is still in an experimental phase, a wide range of data processing and analytical techniques have been applied, and a systematic comparison of the approaches is lacking. Some later studies have found that combining LIDAR data to imaging spectroscopy may be of advantage in species discrimination (e.g. Higgins *et al.* 2014), but others have questioned the usefulness of it (Leutner *et al.* 2012; Ghosh *et al.* 2014).

For biodiversity assessment, however, mapping species one by one is not practical. Especially in tropical forests with their vast amount of tree species, many of them rare, it is practically impossible to obtain the training and validation data for every species that is needed for supervised classification. The need for unsupervised classification approaches has been recognized recently, but so far only they have been applied in only a few experiments. Of these, Baldeck & Asner (2013) focused only on beta diversity, whereas Medina *et al.* (2013) studied alpha diversity, and Féret & Asner (2014) both.

An approach that overcomes the challenge of species identification is to avoid the species level completely, and use spectral variation as a proxy of biodiversity. It is most often justified with the spectral variation hypothesis by (Palmer *et al.* 2000; Palmer *et al.* 2002). According to the hypothesis, environmental heterogeneity is linked to species richness, and at the same time causes variation in the spectral signature. Therefore the amount of spectral variation in the remotely sensed signal could serve as an estimate of biodiversity. In cases like this, however, spectral variation comes from the canopy itself. As the subject of interest is species diversity of canopy trees, the assumption of a varying environment is not needed.

As described previously, spectral variation in vegetation canopies has many sources. Still, it has still been successfully linked with species diversity in a range of environments (e.g. Vaglio Laurin *et al.* 2014; Carlson *et al.* 2007; Rocchini *et al.* 2007; Lucas & Carter 2008; Oldeland *et al.* 2010; Maeda *et al.* 2014). There is no single measure for spectral variation, but different approaches have been reviewed by Rocchini *et al.* in 2010. The authors also recognize the need for diversity estimation using object-oriented methods, as most of the research has focused on the pixel scale. This is particularly relevant for high spatial resolution data such as aerial spectroscopic imagery.

Previous research gives motivation to study the use of imaging spectroscopy for biodiversity mapping in a novel way. The few studies that applied unsupervised classification for biodiversity estimation performed clustering on the pixel scale. However, an object-based approach has the advantage of averaging reflectance variation within a tree crown (Lucas *et al.* 2008; Féret & Asner 2012). This diminishes for instance the problem that different parts of a tree crown have different illumination conditions.

3 STUDY AREA

The study area is the Ngangao forest fragment in the Taita Hills, in the Taita-Taveta district of Southern Kenya (figure 3-1). It consists of a hilltop covered with moist montane forest at an altitude of 1700–1952 m. The climate in the study area is tropical, characterized by a shorter rainy season in November–December, and a longer rainy season in March–May. The forested hilltops of the Taita Hills trap moisture-laden clouds coming from coastal areas, and therefore the forest remains relatively humid throughout the year (also called cloud or mist forest (Pellikka *et al.* 2009)).

The forests in the Taita Hills are remnants of a larger forest cover, now covering only the highest hilltops (Pellikka *et al.* 2013). The Taita Hills belong to a globally important biodiversity hotspot together with other mountains of the Eastern Arc, which extend to Tanzania (Myers *et al.* 2000). The hotspots are characterized by a high degree of endemic species and a high threat of extinction, and together contain a large portion of the world's biodiversity while covering only a small area.

In the forests of the Taita Hills, 100 tree species have been recorded during the years 1877–1985 (Beentje & Ndiang'ui 1988, cited by Aerts *et al.* 2011). In a recent survey, 73 woody species were recorded for Ngangao and another forest fragment together (Mbuthia 2003, cited by Aerts *et al.* 2011). The total number of tree species in Ngangao is probably somewhat smaller, as the survey included also palms.

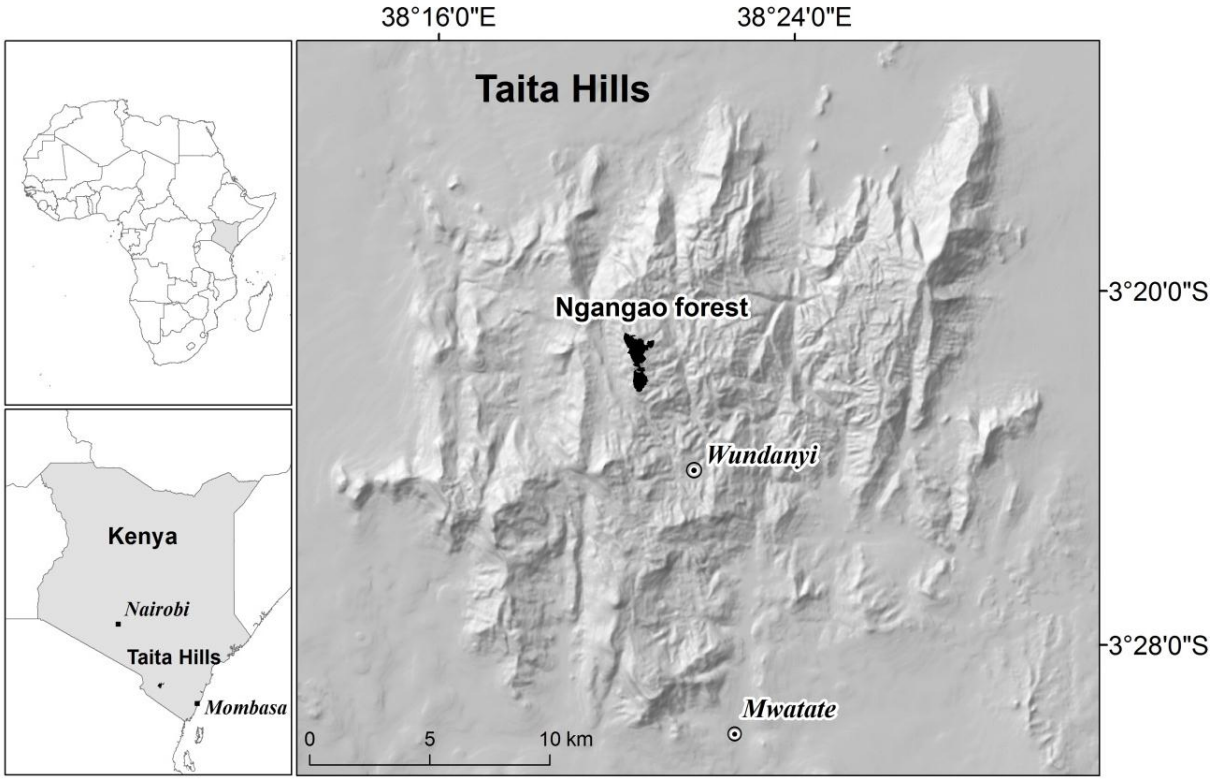


Figure 3-1. Maps of Africa, Kenya and the Taita Hills.

The Ngangao forest fragment covers 120 ha, including 18 ha of plantations of exotic pine (*Pinus patula*) and cypress (*Cupressus lusitanica*) (figure 3-2). The plantations were established in the 1970s mostly on cleared land, so that despite forest loss to agricultural expansion, the total forest area has remained about the same since 1955 (Pellikka *et al.* 2009). Long before this, the forests in the Taita Hills have been under human influence because of the long history of settlement (Pellikka *et al.* 2013).

The most abundant indigenous tree species in the Ngangao forest as recorded by (Mbuthia 2003) are *Tabernaemontana stapfiana, Macaranga conglomerata* and *Albizia gummifera*, which are early-successional species typical of forest edges and gaps (Aerts *et al.* 2011). Other abundant species include *Craibia Zimmermannii, Syzygium sclerophyllum, Pouteria adolfi-friederici, Strombosia scheffleri, Milletia oblata, Cussonia spicata* and *Newtonia buchananii*. Of these, only *P. adolfi-friederici* is associated with old-growth cloud forest. According to other sources, also *Syzygium guineense, Maesa lanceolata* and *Cola greenwayi* are among the most common trees in Ngangao (Omoro *et al.* 2010). The high proportion of secondary successional species in the forest indicates that the community composition has undergone disturbance (Aerts *et al.* 2011; Omoro *et al.* 2010).

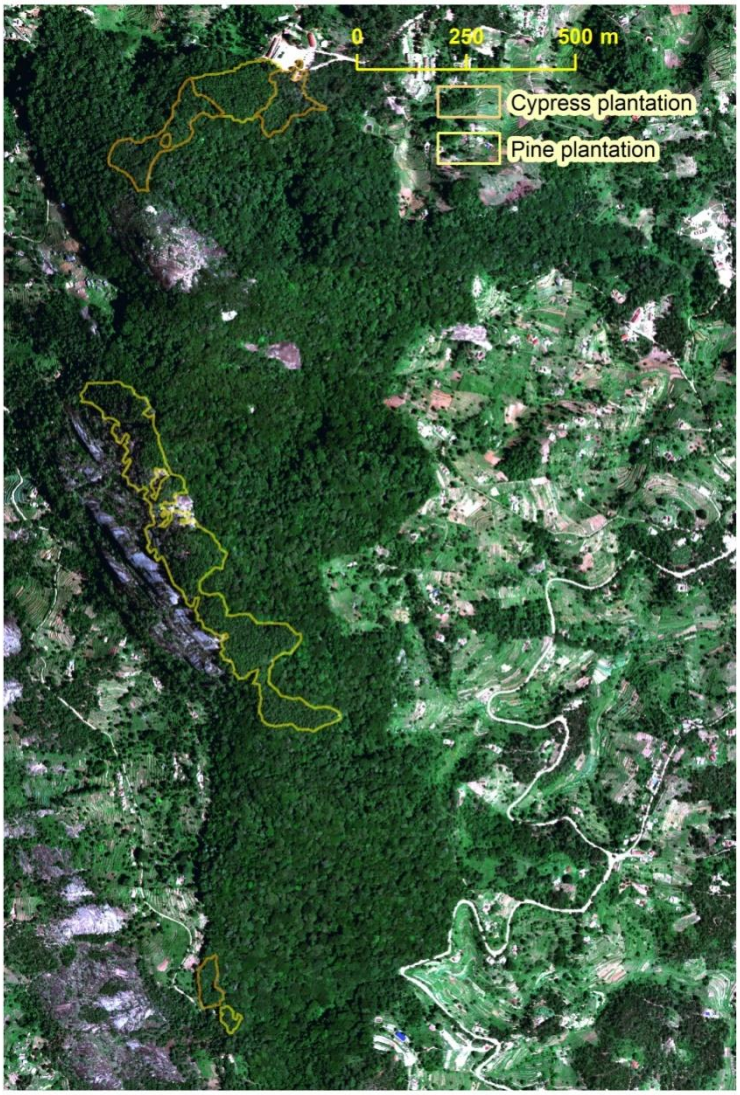


Figure 3-2. Aerial image of the study area, the Ngangao forest.

4 DATA

4.1 Imaging spectroscopy data

The remote sensing data were aerial spectroscopic imagery recorded in February 2013. The raw imaging spectroscopy data consisted of 14 East-West oriented images, or flightlines, that covered the study area.

The sensor was an AisaEAGLE imaging spectrometer which is used in research, commercial use and public services. It is manufactured by a Finnish company, and applications include

forestry management, environmental investigations, precision farming, water assessment and land use planning (Specim Ltd. 2013). The specific sensor that was used was purchased by the University of Helsinki in 2011 for research purposes.

The AisaEAGLE is a pushbroom type spectroscopic sensor for airborne remote sensing. It uses a charge-coupled device (CCD) to record radiance as 12 bit digital numbers (DN's). It records at the spectral range of 400 – 1000 nm, which covers the visible and NIR regions of the spectrum. The sensor has a field of view (FOV) of 37.7 degrees, and a swath width that gives a spatial resolution of 0.68 m at 1000 m flight altitude. It has the flexibility of acquiring data at various spatial and spectral resolutions, according to the binning configurations. The system includes a fibre optic downwelling irradiance sensor (FODIS). The measurements of downwelling diffuse irradiance could be used in preprocessing of the imagery, but were not used in this study.

The detector array has 1024 pixels, 55 of which are used by the FODIS. For this flight campaign, a 2x spatial binning was applied, so that two detector pixels record the radiance for one image pixel. With the specified flight altitude this gives a spatial resolution of approximately 1 m, and the resulting number of pixels per image line is 485.

Also a 4x spectral binning was applied, so that four detecting elements record the radiance for one spectral band. This ensures a stronger signal, but reduces the spectral resolution and limits the number of spectral bands to 129. The details of the sensor and its configurations during the flight campaign are summarized in table 4-1.

Table 4-1. Sensor characteristics and configurations for the flight campaign.

Sensor characteristics		**Configurations for flight campaign**	
Spectral range	400–1000 nm	Spatial binning	2x
FOV (with FODIS)	37.7°	Spectral binning	4x
Swath width	0.68x altitude	Pixels per line	485
Numerical aperture	F/2.4	Spatial resolution	1 m
Radiometric resolution	12 bits	Number of bands	129
		FWHM (spectral resolution)	4.9 nm (average)
		Spectral sampling (band interval)	4.6 nm (average)

Pushbroom type sensors have a linear array of pixels that record simultaneously, one line at a time. Because of sensitivity differences between the detecting elements, the exact center wavelengths of the bands depend on the position of the pixel in the array. This small wavelength shift, called the spectral smile effect, should be accounted for during data preprocessing. The smile effect for this sensor is ±0.35 nm (Specim Ltd. 2013).

The sensor measures dark current at the end of each flightline. The dark current is the electromagnetic noise that is produced by the sensor itself as it warms up, and is measured when by closing the shutter during recording. Dark current measurements are necessary for radiometric calibration of the imagery.

For recording position and attitude during the flight, the AISA system contains a GPS/ inertial measurement unit (Oxford RT3100 Inertial and GPS Navigation System). It records the X, Y and Z position, heading, roll, pitch and speed of the aircraft, as well as the line number and the exact time. The information is needed for georeferencing the imagery.

4.2 Field data

4.2.1 Data collection

Field data was collected from the study area in January and February of 2013 and 2014. 31 field plots were established by manually placing them on the map, aiming for a spatially representative sampling but ensuring that the pine and cypress plantation sites were also sampled (figure 4-1). In the field the plots were located using a GPS device. The exact positions of the plot centres were recorded during data collection and later differentially corrected, using simultaneous measurements of a GPS base station.

The 0,1 ha-sized field plots were circular in shape with a radius of 17,84 m. Each tree with diameter at breast height (DBH) ≥ 10 cm was measured for DBH and their species was identified by a local expert.

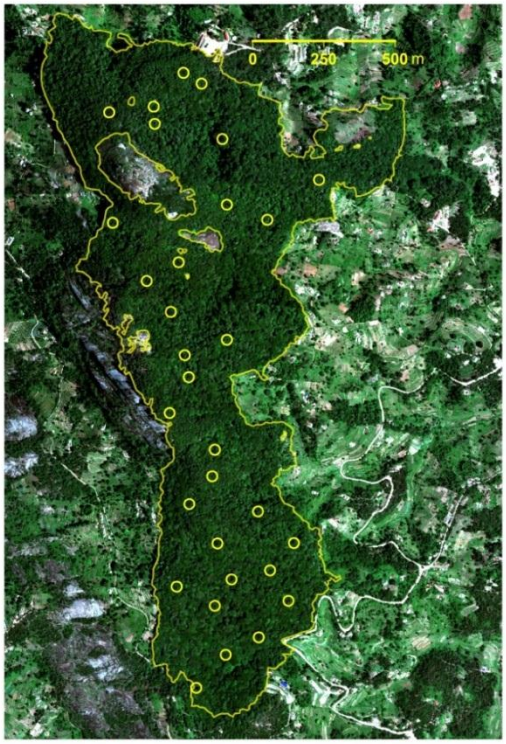


Figure 4-1. Field plot locations in the study area.

It was observed during field work that in average about 30 % of the measured trees seemed to reach the highest canopy level. The estimate was rough because in practice it was not always visible if the tree crown reached the canopy or not, and the proportion of canopy trees varied on the plots. But because it was known that not all trees were visible to the airborne sensor, three different samples were taken from the field data:

1. All measured trees on the plots (DBH ≥ 10 cm)
2. 50 % of the largest trees on each plot
3. 25 % of the largest trees on each plot

Three measures of tree species diversity were calculated for each plot. Species richness was simply the number of different tree species. Simpson's index was calculated assuming sampling without replacement, using the complement of the original index in equation 1. The Shannon–Wiener index was calculated as in equation 2. The calculation of both indices required the count of trees of every occurring species and the total count of trees on the plots.

Species accumulation curves were calculated for the trees of different sizes. They show the accumulative number of species found as a function of sampled plots. To achieve smooth curves, the mean curves of 100 permutations of the data in a random order were calculated with the function *specaccum* in R (package *vegan*, R version 3.0.2).

4.2.2 Diversity measures from field data

In total 53 different species were recorded in the field plots. Figure 4-2 shows the relative abundance of species and the names of the five most common species. The figure does not include the pine and cypress trees. They were recorded with similar abundances as *Oxyanthus speciosus*, but were found only on the plantation sites. Approximately 3 % of the trees were unidentified.

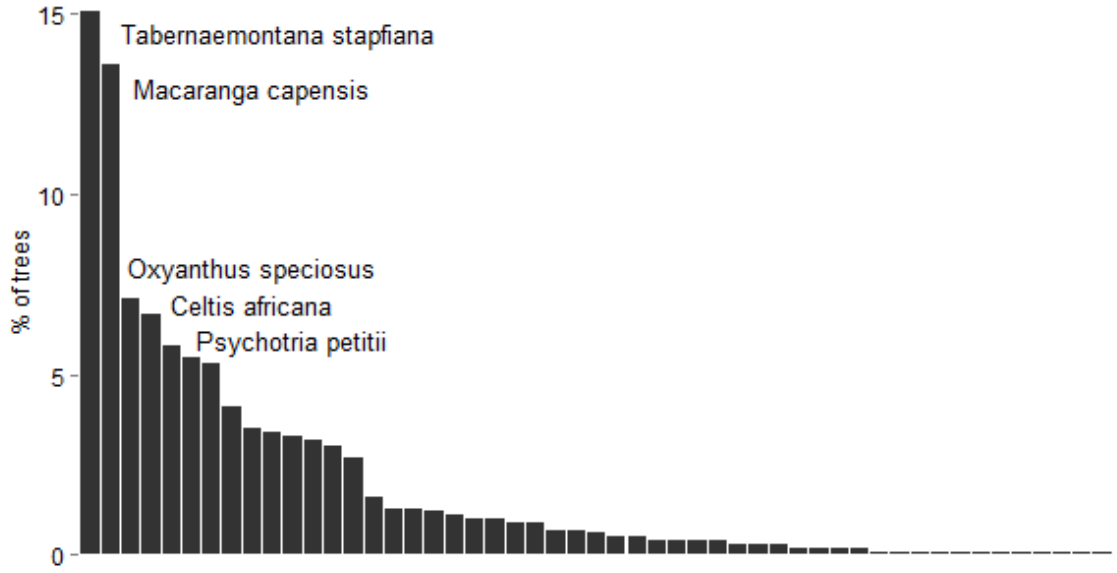


Figure 4-2. Species abundance distribution, with the five most common species named.

The tree species richness values calculated for different tree sizes are shown in tables 4-2 a – c. The lowest species richness was found on the plantation plots (Cypress_1, Cypress_2, Pine_1, Pine_2 and 11_12). The highest species richness was found on plot Indi_22 with 22 tree species. As expected, species richness was lower when only 50 % or 25 % of the largest trees were considered.

Tables 4-2 a – c. Species richness of the plots with different tree sizes.

a)

All measured trees	
Plot	Sp. richness
Indi_22	22
Indi_11	19
12_7 Indi_1 Indi_9 Indi_12 Indi_13 Indi_17 Indi_28	18
Indi_2 Indi_15 Indi_19 Indi_20	17
Indi_8	16
Indi_14 Indi_18	15
Indi_5	14
12_5 Indi_10	13
Indi_4	12
Indi_3 Indi_6 Indi_21	11
12_3	10
12_4 Indi_7	9
11_12 Pine_2	8
Pine_1	6
Cypress_1 Cypress_2	3

b)

50 % of largest trees	
Plot	Sp. richness
Indi_22	18
Indi_28	16
Indi_11	15
Indi_1 Indi_9 Indi_12 Indi_15	14
12_7 Indi_2	13
Indi_8 Indi_20	11
Indi_10 Indi_13 Indi_19	10
12_5 Indi_17 Indi_18	9
Indi_5 Indi_7 Indi_14 Indi_21	8
11_12 12_3 Indi_3 Indi_4	7
Indi_6	6
Pine_2	4
12_4	3
Cypress_2	2
Cypress_1 Pine_1	1

c)

25 % of largest trees	
Plot	Sp. richness
Indi_2 Indi_9 Indi_12 Indi_15	10
Indi_19 Indi_28	9
Indi_8 Indi_10 Indi_20 Indi_22	8
12_7 Indi_1 Indi_11 Indi_13 Indi_17	7
12_3 12_5 Indi_18	6
11_12 Indi_3 Indi_4 Indi_5	5
Indi_7 Indi_21	4
Indi_6 Indi_14	3
Cypress_2 Pine_2	2
12_4 Cypress_1 Pine_1	1

The Simpson's index values calculated for the field data are summarized in figure 4-3. Most plots have values of 0.8–0.9 with all tree sizes, but the plantation plots get always lower values. When only the larger trees are considered, the values of the plantation plots are considerably lower than those of the indigenous plots, which in addition get slightly more varying values. When only one species was observed on the plot, the Simpson's index value was 0.

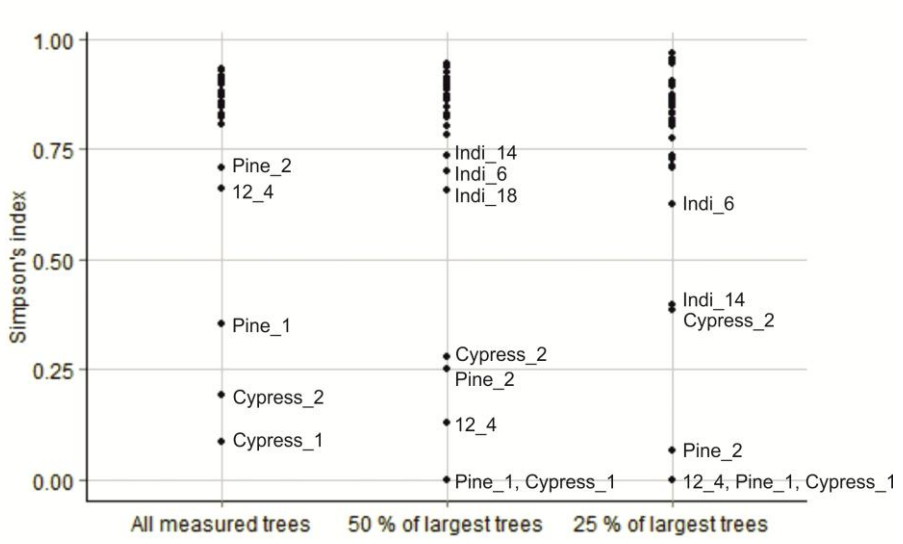


Figure 4-3. Simpson's index values of the field plots. Plot names are shown only for the plots that got low index values.

The Shannon–Wiener index values calculated for the field data are summarized in figure 4-4. Again, the plantation plots get lower values than the indigenous plots, and considerably so when only the larger trees are included. The values of the indigenous plots show more variation than with the Simpson's index, and decrease when fewer trees are under consideration.

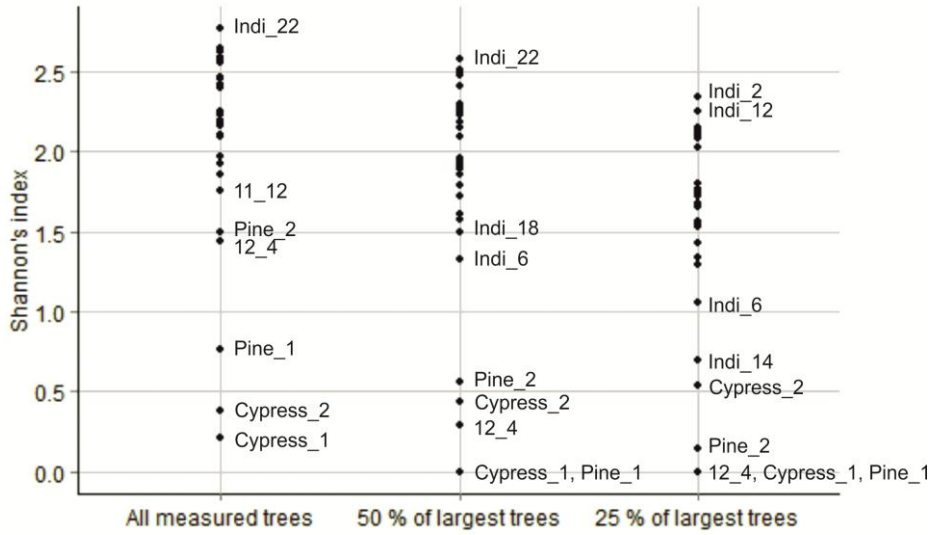
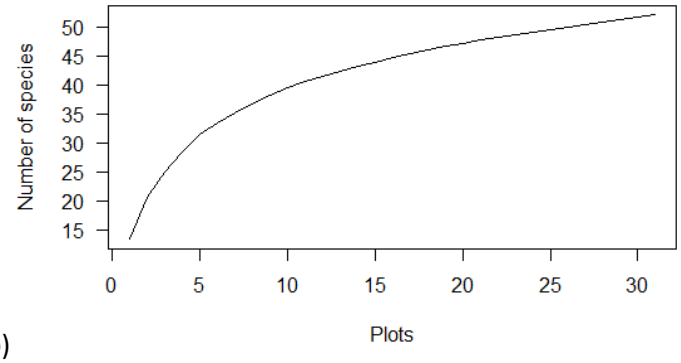


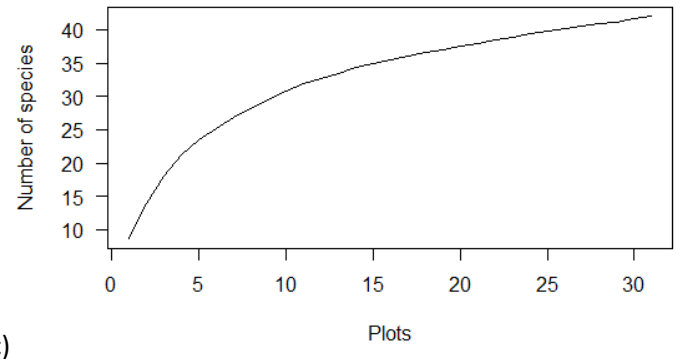
Figure 4-4. Shannon–Wiener index values of the field plots. Plot names are shown only for the plots with highest and lowest values.

The species accumulation curves with different tree sizes are shown in figure 4-5. The most species are found when all measured trees are considered, and the corresponding curve (figure 4-5 a) rises least steeply.

a)



b)



c)

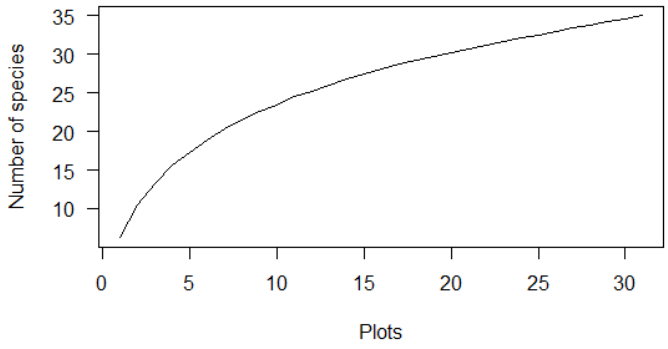


Figure 4-5 a – c. Species accumulation curves for a) all measured trees, b) 50 % of largest trees and c) 25 % of largest trees.

4.3 Additional data

Other data that were used in the process were provided by the BIODEV research group. They included a digital terrain model (DTM) and a canopy height model (CHM), both of 1

m spatial resolution. These were derived from LIDAR measurements that were acquired simultaneously with the collection of the imaging spectroscopy data.

Additionally, a land cover vector file was used for delineating the study area and the plantation sites in the forest. It was based on the work presented by Pellikka *et al.* (2009). Also sensor model files used in preprocessing of the data were provided by research group, and partly by the manufacturer of the sensor.

5 METHODS

5.1 Preprocessing

5.1.1 Radiometric calibration

A radiometric calibration was done to the raw imagery to convert the digital number (DN) values to at-sensor radiance values. In the process also noise caused by the sensor itself (dark current) was removed. Additionally, the image data was synchronized with the GPS/IMU data to allow georectification at a later stage.

The conversion relates the DN values to the at-sensor radiance (L_{sensor}) as in the following equation:

$$L_{sensor} = c_0 + c_1 * DN \qquad \text{Equation 3.}$$

where the offset (c_0) and slope (c_1) coefficients are sensor and band specific. The dark current is removed from the data by subtracting the average dark current values for each band from the recorded values. The calibration also reverses the order of the bands, so that small band numbers have small wavelengths.

The radiometric calibration was done using software provided by the manufacturer of the sensor (CaliGeo 4.9.15, Specim Ltd.). The required inputs were a sensor calibration file also provided by Specim Ltd., the raw imagery which contained also dark current measurements, and the raw navigation data recorded by the GPS/IMU.

The output from the calibration were radiometrically corrected images consisting of radiance values, unit mW/cm^2*sr* µm multiplied by a scaling factor of 1000. An additional output was a synchronized navigation data file that was used later for

georectification. Figures 5-1 and 5-2 show the spectral profiles of vegetation targets before and after radiometric calibration.

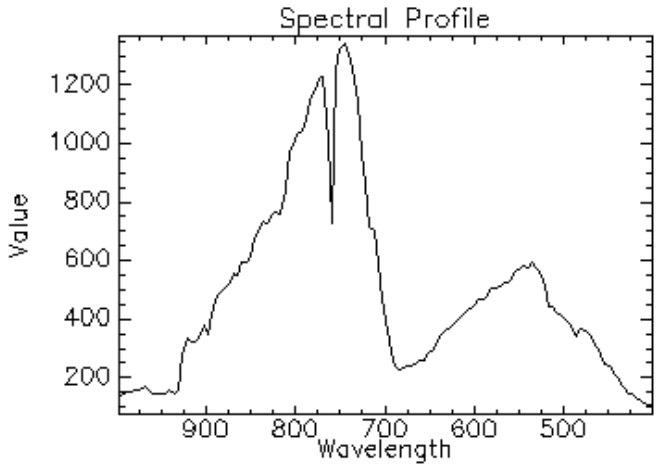


Figure 5-1. The spectral profile of a vegetation pixel before radiometric calibration. The values are unitless digital numbers. Note the wavelength axis running from larger to smaller values.

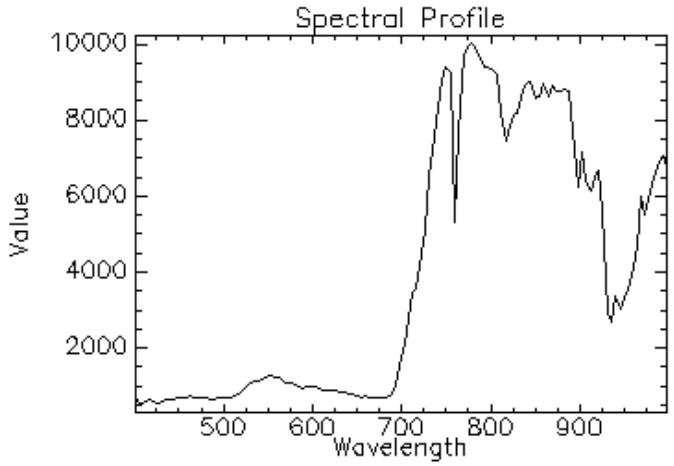


Figure 5-2 The spectral profile of a vegetation pixel after radiometric calibration. The spectrum exhibits absorption features due to gases and water vapour in the atmosphere. The units are radiance, (mW/cm^2*sr* µm) * 1000.

5.1.2 Atmospheric correction

An atmospheric correction was applied to each flightline to convert the radiance values to reflectance values, with the effects of the atmosphere removed. The atmospheric correction was made using ATCOR-4 software for airborne sensors (version 6.2.0, ReSe Applications Schläpfer).

A sensor model was created in ATCOR that corresponds to the sensor and its configurations during the campaign. A response file (channel filter file) was created for each band, which defines the way in which the band is sensitive to radiation. For this, measurements of band center wavelengths and full width at half maximum (FWHM) values were used (figure 5-3). These were provided by the sensor manufacturer from laboratory measurements. The response type was defined as a Butterworth 2 type function, which is close to a Gaussian curve and is the best approximation for a 4x spectral binning (figure 5-4). The sensor model was a spectral smile sensor type, which means that the exact band center positions for each sensor pixel depends on its location in the detector pixel row.

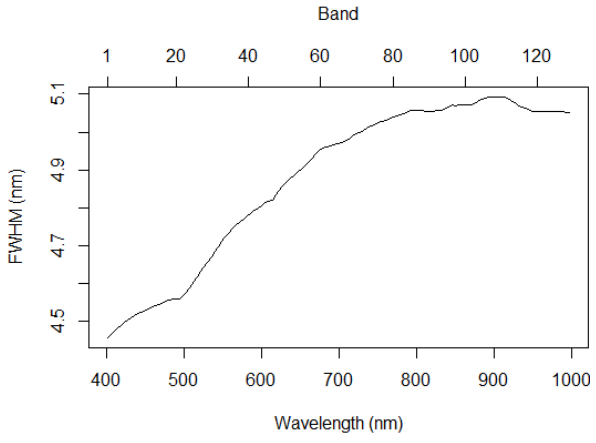


Figure 5-3. The full width at half maximum values for each band in the 4x spectral binning mode, based on laboratory measurements by Specim Ltd..

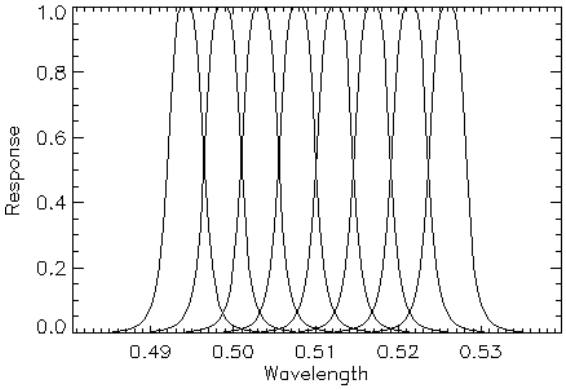


Figure 5-4. The response curves of bands 22 to 29. As can be seen, there are slight differences between the curves. The wavelength is unit µm.

ATCOR was run in the flat terrain mode, although the study area is on a hilltop. In the rugged terrain mode, the program would calculate the solar incidence angle for each pixel based on a DEM. However, with high spatial resolution imagery, and a scene that has forest cover, the true incidence angles for each pixel cannot be derived from a digital terrain model because tree crowns have different illumination on different sides. The flat terrain mode was therefore used.

The main atmospheric parameters that have to be adjusted in ATCOR are aerosol type, visibility and water vapor. Visibility is related to the optical thickness of the atmosphere, which is affected by molecular scattering and absorption and aerosol scattering (Richter & Schläpfer 2011).

The estimation of visibility and aerosols in ATCOR is based on the dense dark vegetation (DDV) approach. For vegetation pixels, the radiance in the red wavelength region is assumed to be 0.1 times the radiance in the NIR region. Visibility is then automatically estimated using look-up tables (LUTs) that are included in the program. The visibility estimates are obtained by comparing the radiance values of reference vegetation pixels to a LUT that is modelled with corresponding solar geometry and aerosol type. A spatial interpolation was chosen for calculating the visibility of non-reference pixels. The same visibility is applied to the blue spectral regions. Path radiance is estimated based on its increasing effects towards shorter wavelengths.

ATCOR allows selection of an atmospheric database that best corresponds to the scene in terms of aerosols and altitude. Altitude affects Rayleigh scattering caused by nitrogen and oxygen gases, because their concentration is dependent on air pressure. The aerosol types for selection are rural, urban, maritime, or desert type, with varying water vapor contents. The selected atmospheric file had aerosols of rural type, an altitude of approximately 2 km and a water vapor column of 2.9 g / cm^2 from ground to space. The parameters are slightly adjusted in ATCOR during the process.

After visibility estimation, ATCOR calculates atmospheric water vapor for each pixel. This is based on a linear interpolation between window channels that surround a water vapor absorption feature in the spectrum. The depth of the absorption feature is a measure of the water vapor column content (Richter & Schläpfer 2011). The 940 nm region was chosen for the water vapor algorithm, because it allows the non-linear effect of vegetation to be included. The reflectance of vegetation is affected by leaf water absorption and therefore cannot be

interpolated linearly. The correct choice of absorption and window regions was important for good results. The window region, which is not much affected by water vapor, was set to 870–887 nm, and the absorption feature to 899–970 nm. The resulting average water vapor content per scene varied around 1.3 g / cm^2.

After atmospheric correction, a spectral smile interpolation was applied in ATCOR. It corrects for the spectral smile effect, bringing the spectral bands to a common center wavelength in across-track direction.

The atmospheric correction seemed successful based on the inspected spectral profiles of sample pixels. The spectral profile of a vegetation pixel in figure 5-5 shows the typical pattern of low reflectance in the visible wavelength region, and high reflectance in NIR. Some noise was encountered at approximately 850–875 nm, but in further processing a noise-removing technique was applied.

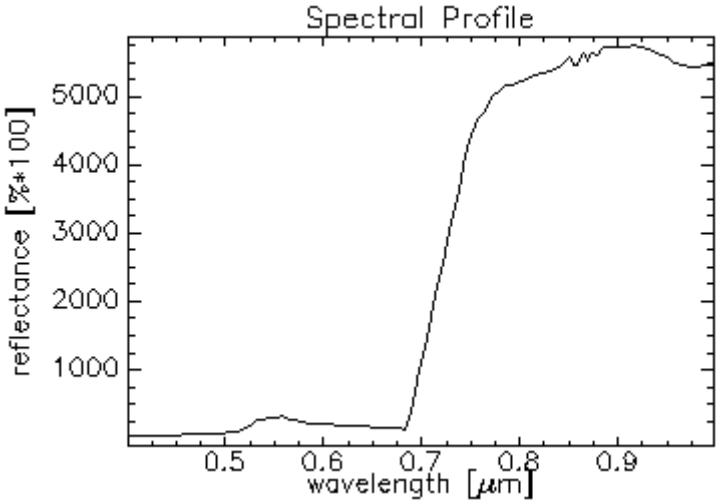


Figure 5-5. The spectral profile of a vegetation pixel after atmospheric correction. The units are reflectance values, percentage multiplied by 100.

5.1.3 Georectification

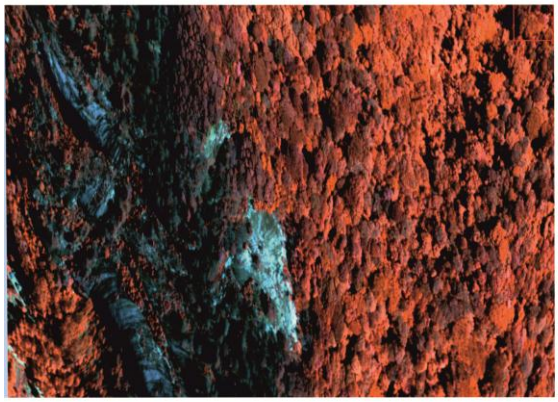
Georectification of the atmospherically corrected flightlines was performed with PARGE (PARametric GEocoding, v. 3.1, ReSe Applications Schläpfer). A LIDAR-derived digital

surface model (DSM) with 1 m resolution was used for rectification. Other input data were the synchronized navigation data file from the CaliGeo processing step and a sensor model file.

To correct for the slight difference in position of the GPS/IMU system and the AISA sensor, a boresight calibration had to be performed. The calibration values were calculated by the program based on ground control points that were collected from the DSM and from the image in corresponding locations. The resulting roll, pitch and heading angles were 1.78, -0.22, and 0.12 degrees, respectively. The resampling option for georectification was set to fast nearest neighbour, which does not alter the original pixel values.

Results of the georectification are shown in figure 5-6 b. The accuracy was visually assessed against a canopy height model (CHM). The accuracy was at best in the centres of the flightlines, where the difference between crowns in the two images was approximately 3 pixels at lowest. The positioning error increased towards the edges of the flightlines, and was particularly high at the left edge in flight direction. This was due to the FODIS pixels in the other edge of the sensor, and the sensor model file which did not represent the positions of the pixels in the most accurate way.

a)



b)

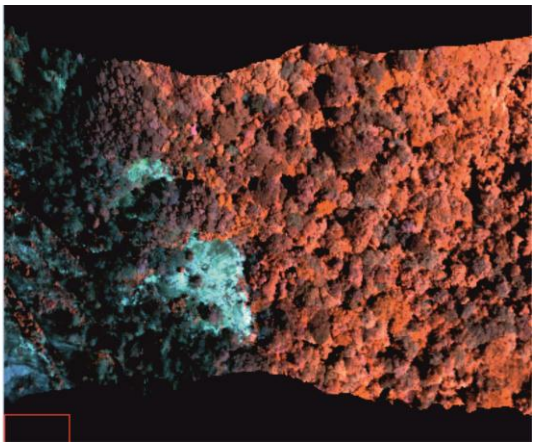


Figure 5-6 a – b. A part of the image a) before and b) after georectification. The images are false-colour infrared compositions of bands 90, 65 and 39 (center wavelengths at around 812, 693, and 572 nm, respectively).

A better accuracy of the navigation data, which the georectification was based on, may have been achieved if a differential correction had been made to the GPS measurements of the

sensor system. This would have required simultaneous GPS measurements on a base station in a known location. However, it would not have affected the distortion at the edges.

5.1.4 Study area mosaicking and delineation

To bring the imaging spectroscopy data from the study area together in one image, the flightlines were combined to an image mosaic and the study area was delineated from it. Because geometric accuracy was at highest in the centres of the flightlines, the overlapping parts of the lines were clipped from the edges before mosaicking. However, the edges of flightlines were still visible in the mosaic, and comparison with the CHM (figure 5-7) indicated that some data was missing at the edges.

a) b)

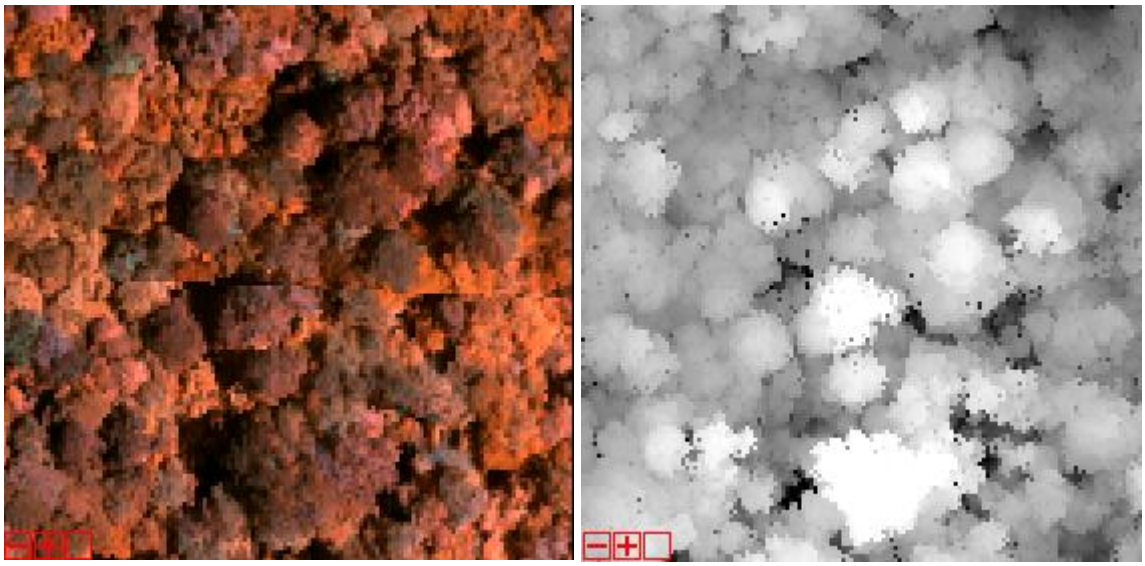


Figure 5-7 a – b. a) Edge between two flightlines in the image mosaic (bands 90, 65, 39). b) The same area in the canopy height model, where relatively higher areas have a lighter shade.

The study area was then delineated from the image mosaic using a land cover vector file and the CHM. The area classified as forest was first extracted from the image, then cleared areas and bushland were excluded by selecting only pixels with a height of 7 m or more.

5.1.5 Shadow removal

Previous studies (Clark *et al.* 2005; Lucas *et al.* 2008; Féret & Asner 2013) have indicated that excluding shaded canopy pixels might result in a better discrimination of species. In addition to canopy self-shading, the imagery had some shadows due to topography. The Northwestern corner of the study area lies on a west-facing slope and is more shaded, while most of the area faces east, approximately towards the sun azimuth angle (108–121 degrees from North during data acquisition).

To minimize the effects of shadows, the most shaded areas were removed using the eCognition software (eCognition Developer 8.9, Trimble Navigation Ltd.). The image was first divided to small segments, then the shaded segments were excluded applying a brightness treshold value which evaluated the intensity of all spectral bands. Figure 5-8 shows the forest area before and after removing low-height and shadowed pixels.

a) b)

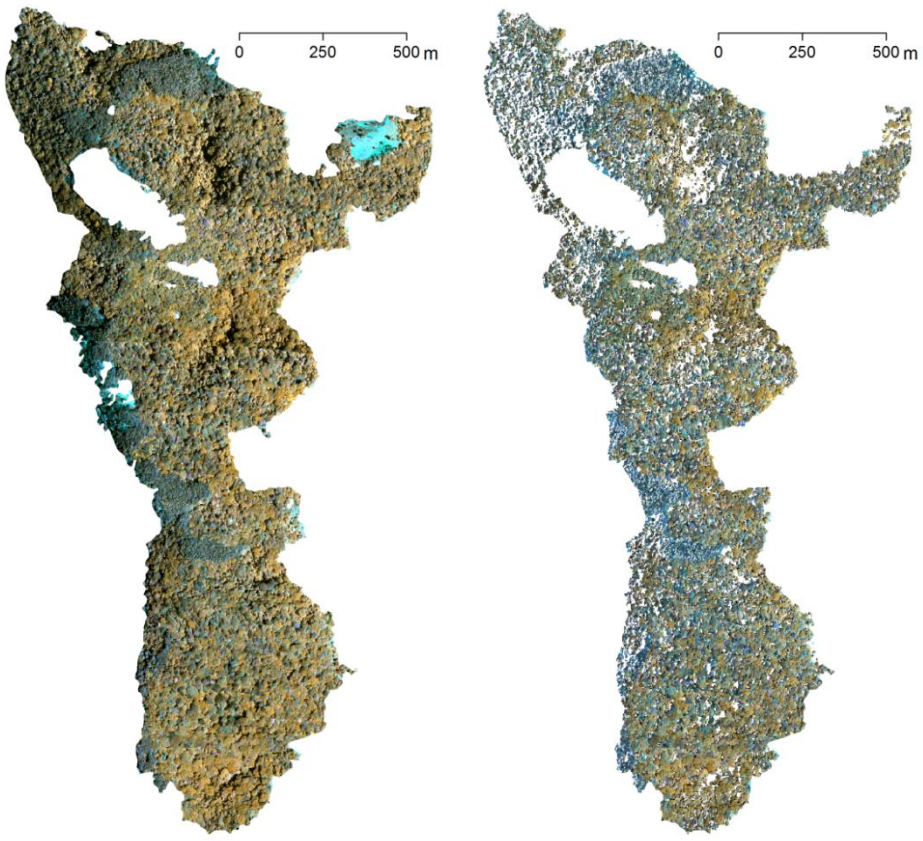


Figure 5-8 a – b. a) The study area delineated with the land cover classification; RGB bands: 90, 65, 39 of the spectroscopic image. b) The study area after removing shadowed pixels and areas with height less than 7 m, and manual cleaning of edges.

5.2 Feature extraction using Minimum Noise Fraction transformation

The Minimum Noise Fraction transformation (MNF, Lee *et al.* 1990) and a subsequent selection of MNF bands were applied to reduce dimensionality and noise of the imaging spectroscopy data. The transformation was also necessary, and more suitable than a PCA, because clustering is based on distance measures between observations. The PCA bands are ordered by their variance, and therefore the distances of observations within different PCA bands would not be comparable. The MNF transformation was perfomed with ENVI software (version 4.8, Exelis VIS Ltd.). The transformed image is shown in figure 5-9, where the first three MNF bands are visualized in red, green and blue, respectively.

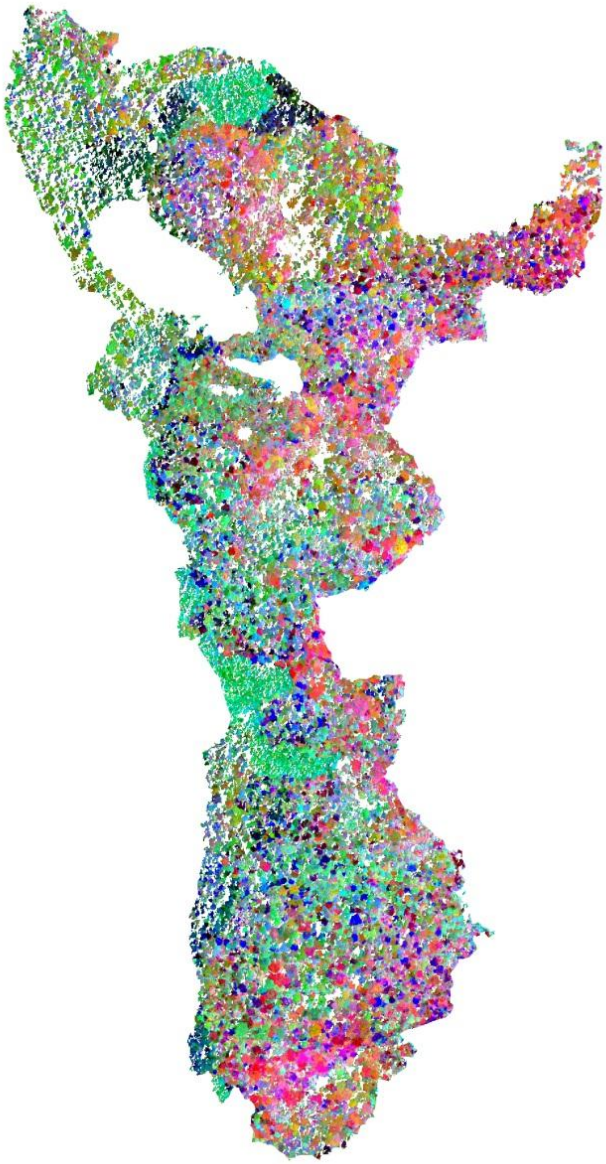


Figure 5-9. The MNF transformed image (MNF bands 1, 2 and 3).

The eigenvalue plot and the output bands were visually evaluated to select bands for further processing. The eigenvalue plot (figure 5-10) indicated that the first 15–20 MNF bands contained useful information and the rest mostly noise. Visual assessment of the MNF bands, however, showed that the brightness differences between flightlines became more pronounced from band 14 onwards, while tree crowns were less visible (figure 5-11). Because the differences between flightlines would be undesired noise in the study, only the first 13 MNF bands were selected for further analyses.

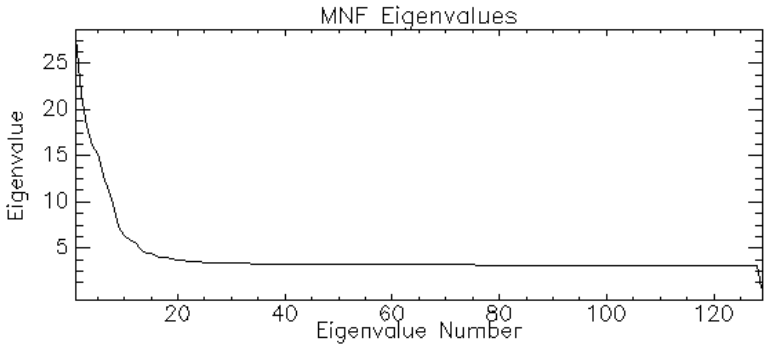
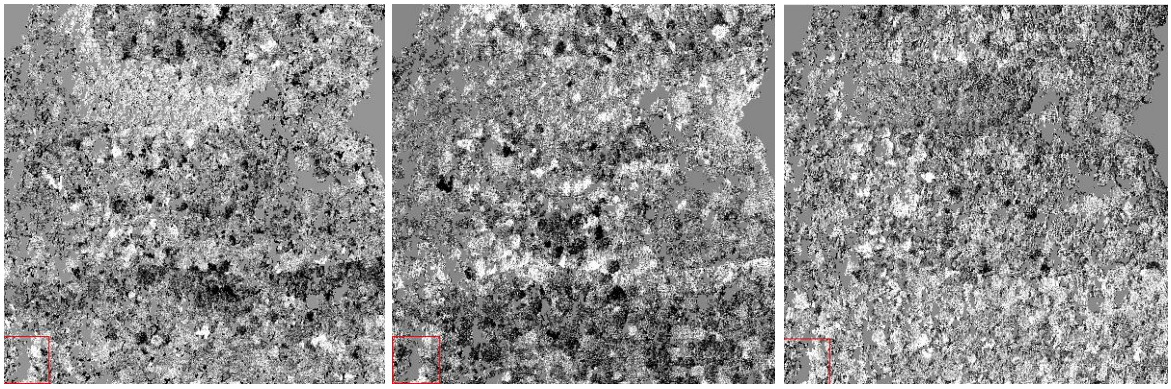


Figure 5-10. Eigenvalues for the 129 MNF bands.

a) MNF band 11 b) MNF band 12 c) MNF band 13



d) MNF band 14 e) MNF band 15 f) MNF band 16

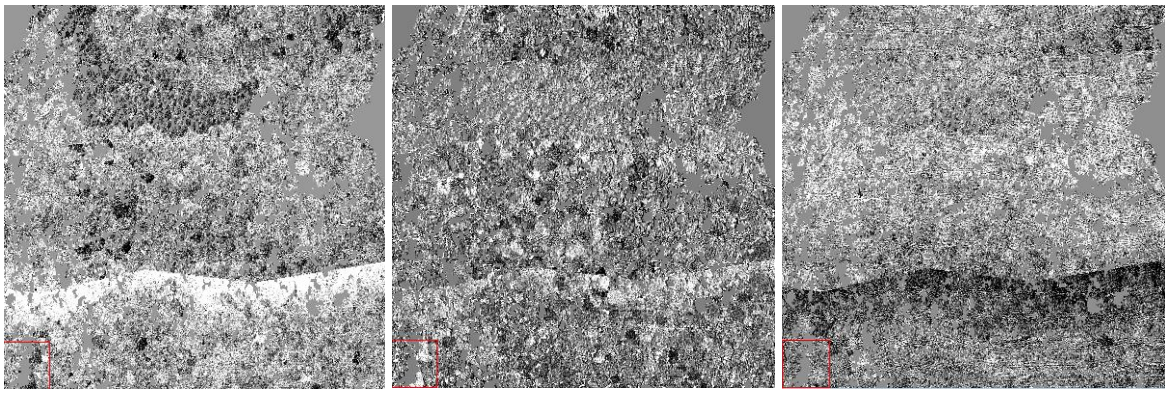


Figure 5-11 a – f. A part of the image depicted with MNF bands 11–16.

5.3 Tree crown segmentation

The MNF image was segmented with the aim to obtain objects that represent tree crowns. Segmentation was performed using themultiresolution segmentation algorithm in the eCognition software. A combination of spectral and spatial information has been proved useful for delineating individual tree crowns in mixed-species forests (Féret & Asner 2013; Bunting & Lucas 2006).

Here, spatial information was given less importance (shape factor 0.1, compactness value 0.1), since it was assumed that tree crowns may not have a compact shape but are separable by their spectral properties. The 13 MNF bands were weighted inversely according to the MNF band number, because of increasing noise with increasing MNF band number.

The scale factor parameter for the segmentation algorithm affects the size of the output segments. It was set to 5, which resulted in segments that seemed to resemble tree crowns. Although some crowns were clearly divided to several segments, this was preferred rather than having several crowns in one segment. The reason was that segments from one crown were expected to be spectrally similar, possibly enough to later be assigned to the same cluster. Splitting a crown was therefore expected to give more realistic estimations of species richness than having several crowns in one segment. There was no way to quantitatively assess the success of the segmentation, so I had to rely on visual assessment.

The segmentation resulted in approximately 20 000 segments for the whole forest. Segments smaller than 2 m^2 were filtered out. The mean values and standard deviations of each MNF band were calculated for the segments. The number of segments on each plot was compared to the number of trees on the plots with different size limits.

Six groups of segments were chosen for visualization of their spectral properties (figure 5-12). The aim was to assess whether the segments can be clustered based on the similarity of their MNF values (mean or standard deviations). Segments were selected from pine and cypress plots, because they were known to represent mainly one species. For the other segments, a coarse assumption was made that visually similar segments might represent spectrally similar species. Segments that appeared blue, pink or green on the image formed by the first three MNF bands were assumed to represent clusters of spectrally similar tree crowns.

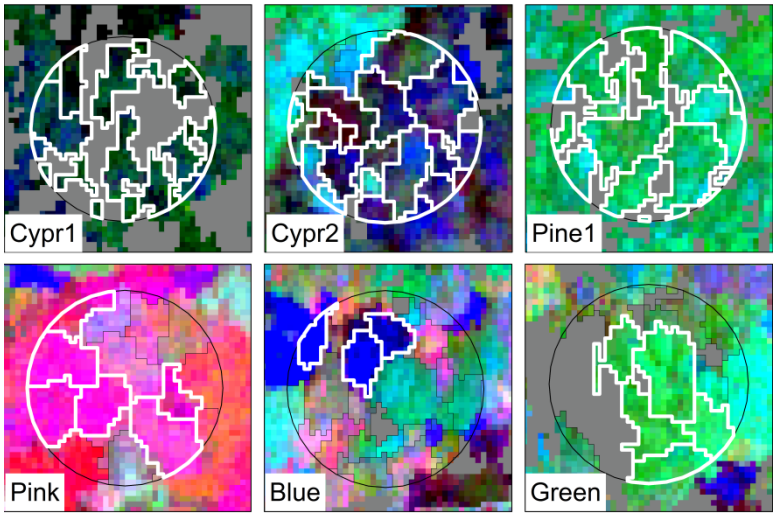


Figure 5-12. Some of the segments selected for comparison (white outlines). The pine and green segments appear visually similar, but the two cypress plots have segments of somewhat different colours.

The selected segments were plotted by their mean MNF values on different bands and by the standard deviations of the MNF values on different bands. Based on this, the standard deviations were left out from further analyses and only the mean MNF values were considered in the clustering. Also the MNF spectra of the selected segments were plotted.

5.4 Spectral clustering of segments

5.4.1 Algorithm *clara*

The clustering of the segments was performed using function *clara* in package *cluster* in R (R version 3.0.2, R Foundation for Statistical Computing). *Clara* (Clustering Large Applications) performs a partitional clustering around medoids (Kaufman & Rousseeuw 1990). It can handle large datasets by dividing them to sub-datasets of fixed size.

The algorithm requires as an input the parameter *k*, which is the desired number of clusters. Each sub-dataset is partitioned to *k* clusters by searching for *k* medoids (representative objects) and assigning each observation to the nearest medoid. The objective is to find *k* medoids that minimize the mean of the dissimilarities of the observations to their closest medoid. Compared to the k-means algorithm, *clara* is more robust because the parameter to

be minimized is the mean of Euclidean (root sum-of-squares) distances instead of a mean of squared Euclidean distances.

The sub-dataset for which the mean of the dissimilarities is minimal, is retained, and each sub-dataset is forced to contain the medoids obtained from the best sub-dataset until then. Randomly drawn observations are added to this set until the determined size of the sub-dataset has been reached.

The output of the algorithm is a cluster label assigned to each observation (in total *k* different labels). Here, the cluster labels were taken to represent different species. If the initial medoids are not given by the user (which was the case here), *clara* creates the seeds with a random number generator. This random component in the algorithm may cause the clustering result to be somewhat different on every repetition.

5.4.2 Implementation

The *clara* algorithm was applied to the dataset that contained the mean MNF values of each segment. In the first phase, the effect of the parameter *k* was tested and the algorithm was run with *k* values ranging from 2 to 80. The number of samples drawn (or subdatasets) from the data was kept at 100, with 240 observations in each sample.

After preliminary inspection of the results, the clustering was performed again, this time for a more limited range of *k* values from 45 to 60. This is later referred to as the second round of clustering. The clusterings with each *k* value were iterated 100 times to enable calculation of a mean result.

5.5 Calculation of biodiversity indices from clustering results

The clusters produced by *clara* were taken to represent species. Thus, the species richness values of the clustering results were the number of clusters found on each plot. With the second round of clustering, species richness was calculated as the mean of 100 clustering results for each plot.This was to average the variation from the clustering results which were slightly different on each repetition.

The Simpson's index and the Shannon–Wiener index were calculated in a similar manner as for the field data, after equation 2 and and the complement of equation 1. It required countinf the number of segments belonging to each cluster, and the total number of segments on the plot. Tables 5-1 and 5-2 summarize the calculated biodiversity variables which were calculated for each of the 31 plots.

Table 5-1. Naming of diversity variables from clustering results (3 x 79).

RESULT OF ONE CLUSTERING	***k* = 2**	**…**	***k* = 80**
Species richness (number of clusters)	Spr_k2	…	Spr_k80
Simpson's index	Simp_k2	…	Simp_k80
Shannon–Wiener index	Shan_k2	…	Shan_k80

Table 5-2. Naming of diversity variables from second round of clustering (3 x 16).

MEAN OF 100 CLUSTERINGS	***k* = 45**	**…**	***k* = 60**
Species richness (number of clusters)	Spr_mean_k45	…	Spr_mean_k60
Simpson's index	Simp_mean_k45		Simp_mean_k60
Shannon–Wiener index	Shan_mean_k45		Shan_mean_k60

5.6 Analyses

First, the field-derived biodiversity measures were compared to the diversity measures from clustering results by correlation analyses. The effect of the increasing number of clusters (parameter *k*) was illustrated by plotting the Pearson's correlation coefficients.

Next, linear regression analysis was used to study using the relationship between diversity measures from field data and from clustering. A threefold cross-validation was used to assess the accuracy (root mean square error, RMSE) of the model predictions. Some of the relationships were plotted for illustration, and the coefficients of the models were compared. A model for predicting species richness was selected for making a tree diversity map for the study area. The model with an intercept closest to 0 and slope closest to 1 was determined best, because it most directly related the predicted species richness to the field-measured species richness.

5.7 Tree species richness map

A tree species richness map was created for the Ngangao forest fragment based on the clustering. The best model determined above was chosen for predicting tree species richness for the whole study area. The species richness of each pixel was calculated in R with a moving window. A circular neighbourhood of 17.84 m radius (same as the field plot size) was used as the window. The number of different clusters in the neighbourhood represented species richness from clustering, and the output pixel size was set to 5 m. Because the clustering results were known to vary, the whole process was repeated five times and the final tree species richness map was calculated as the mean value of the five maps. The final map was visually compared to the canopy height model of the study area.

6 RESULTS

6.1 Segmentation and clustering

6.1.1 Segmentation

By visual assessment, segmentation seemed to delineate tree crowns well from the MNF image, with some larger trees clearly divided to several segments (figure 6-1).

a) b)

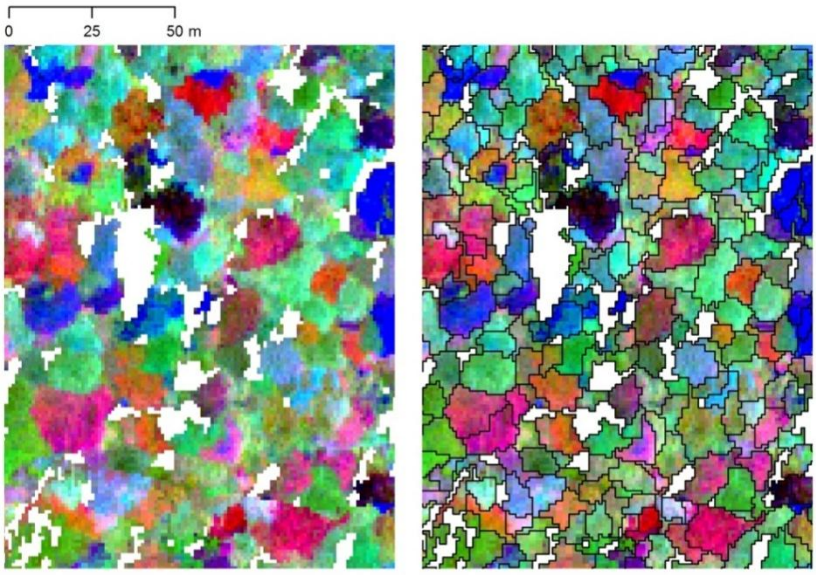


Figure 6-1. a) A part of the MNF image and b) the segments created from it.

6.1.2 Spectral differences between segments

The results of the comparison between six groups of segments is shown in figures (X_X). When plotted by the mean MNF values, the groups showed different degrees of clustering with different band combinations (figure 6-2). However, even with higher MNF bands the groups could be distinguished to some extent.

The plots (not shown) of the standard deviations of the MNF values did not reveal any clustering patterns for the groups. Therefore the standard deviations were left out from further analyses and only the mean MNF values were considered in the clustering.

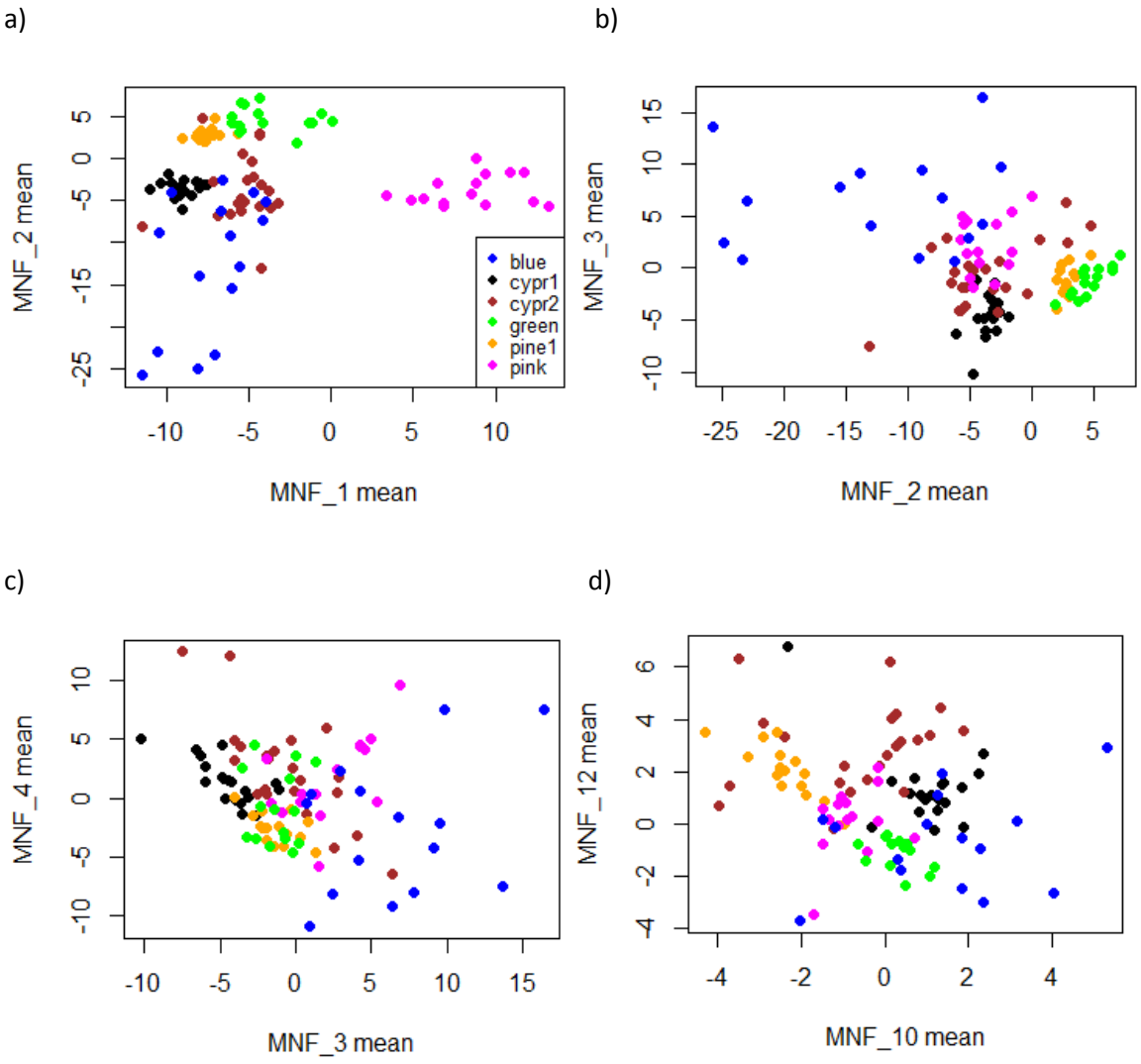


Figure 6-2 a – d. The mean MNF values for each crown show clustering patterns. The different groups of trees are distinguishable even at the higher MNF bands (figure d).

Also the MNF spectra of the selected segments were plotted (figure 6-3). The only very homogenous group were the pine segments, which had very similar values throughout the MNF bands (figure 6-3 c). The segments on the second cypress plot (figure 6-3 b) had surprisingly varying spectra, knowing that the canopy consists almost exclusively of cypress trees. For the segments in the pink, blue and green groups homogeneity was not much expected, because it was not known if they represent spectrally similar species. However, the groups seemed separable from each other at least on some bands. This gave good grounds to expect success with the clustering.

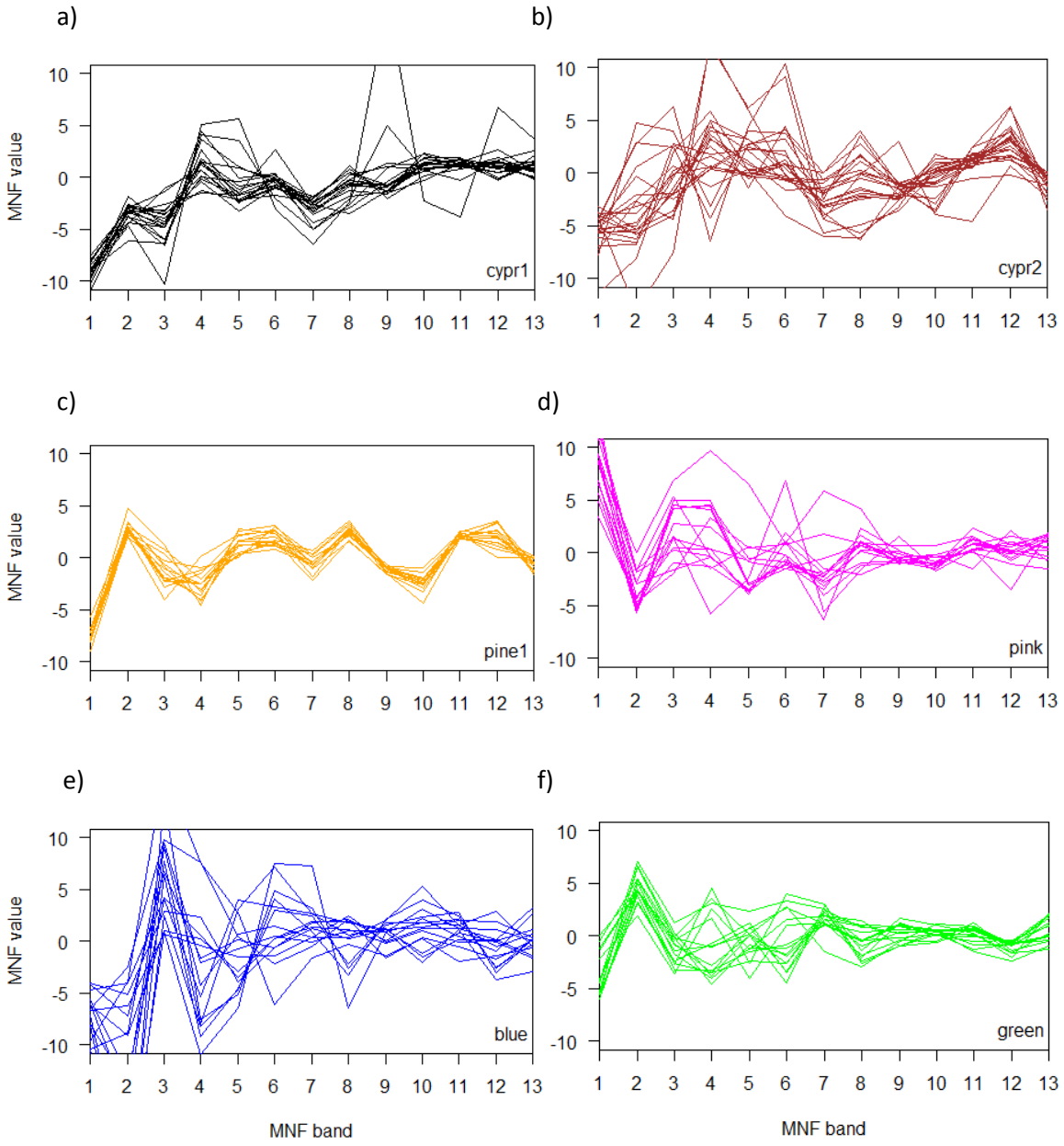


Figure 6-3 a – f. MNF values plotted for each group of trees. The value is the mean MNF value of the segment pixels.

6.1.3 Number of segments on plots

The number of segments in the AISA image created by the segmentation algorithm was compared to the number of trees on the plots (figure 6-4). When all measured trees (DBH ≥ 10 cm) were included, the number of trees on the plots was much higher than the number of segments. The average number of segments corresponded better to the field data when only 50 % or 25 % of the largest trees on the plots were included. The individual plots showed no relationship between the number segments and the number of trees.

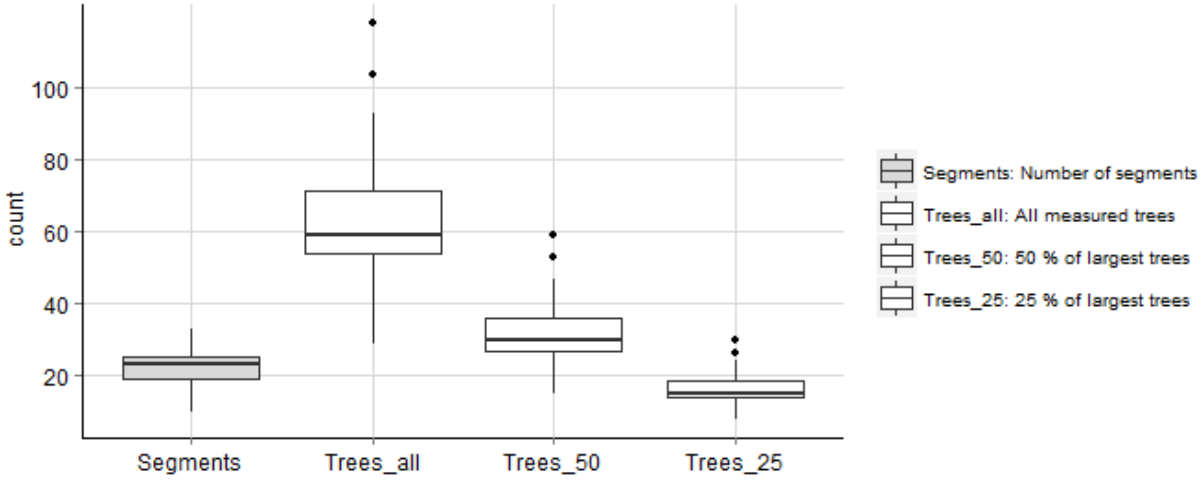


Figure 6-4. Number of segments and trees on the plots. The boxes represent half of the 31 plots and the line in the box corresponds to the mean. The whiskers show the range of the values and the points are outliers.

6.1.4 Clustering

Two examples of the clustering results are shown in figure 6-5. Similar patterns can be distinguished in both results, but the clustering with 50 clusters detected more small clusters. It also distinguished between pine and cypress plantations, unlike the clustering with 10 clusters.

a) b)

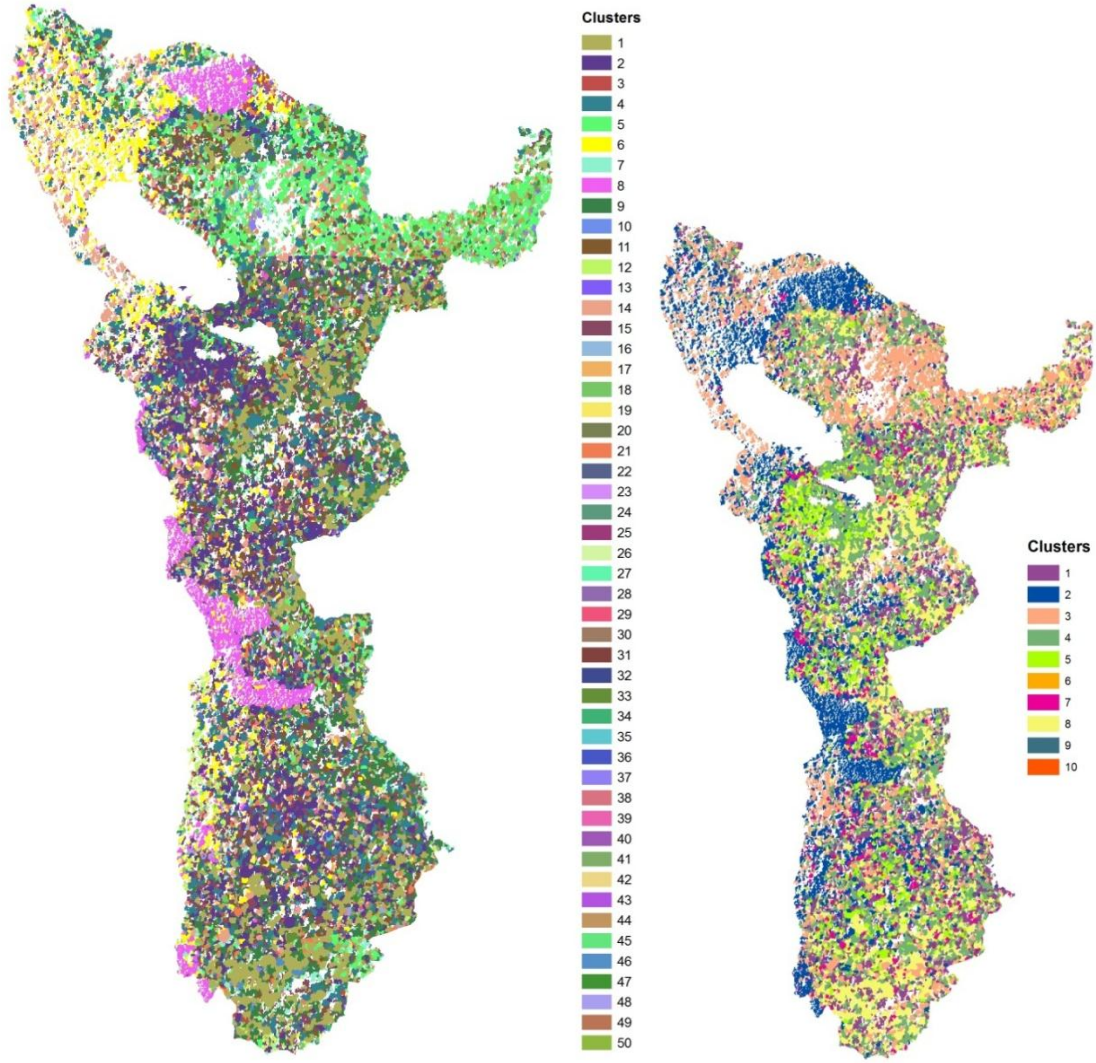


Figure 6-5. Examples of clustering results with a) *k* = 50 and b) *k* =10.

Figure 6-6 shows a comparison of the clustering on some field plots against the MNF image. As can be seen, all segments in the pine plot have been assigned to one cluster, but the other cypress plot has several clusters assigned to it (figure 6-6 a). The plot has somewhat varying colours in the MNF image as well figure 6-6 b).

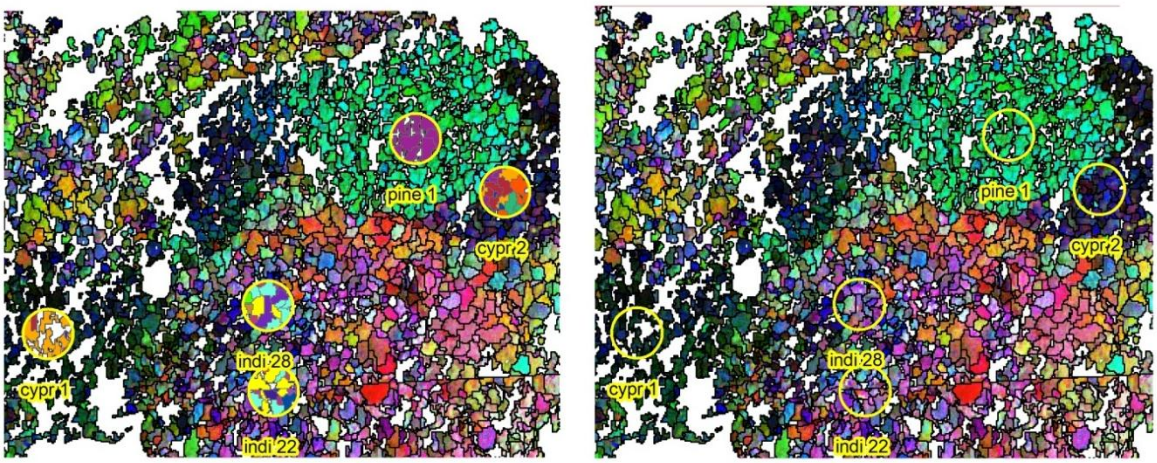


Figure 6-6 a) Clustering results on some plots shown on the MNF transformed image. b) The MNF transformed image of the same area.

6.2 Diversity measures from clustering

Tree species richness values derived from the clustering are summarized in figure 6-7. It shows the maximum, minimum and mean species richnesses of 31 plots calculated for all the clustering results. Also the maximum, minimum and mean species richnesses from field data are shown in dark blue for comparison. The highest (maximum) species richness assigned to the plots by clustering cannot exceed the total number of clusters (*k*), but it increased with *k* as expected (red points). The mean species richness similarly increased with the total number of clusters, but more slowly (green points). The lowest species richness assigned to the plots mostly remained at one (light blue points).

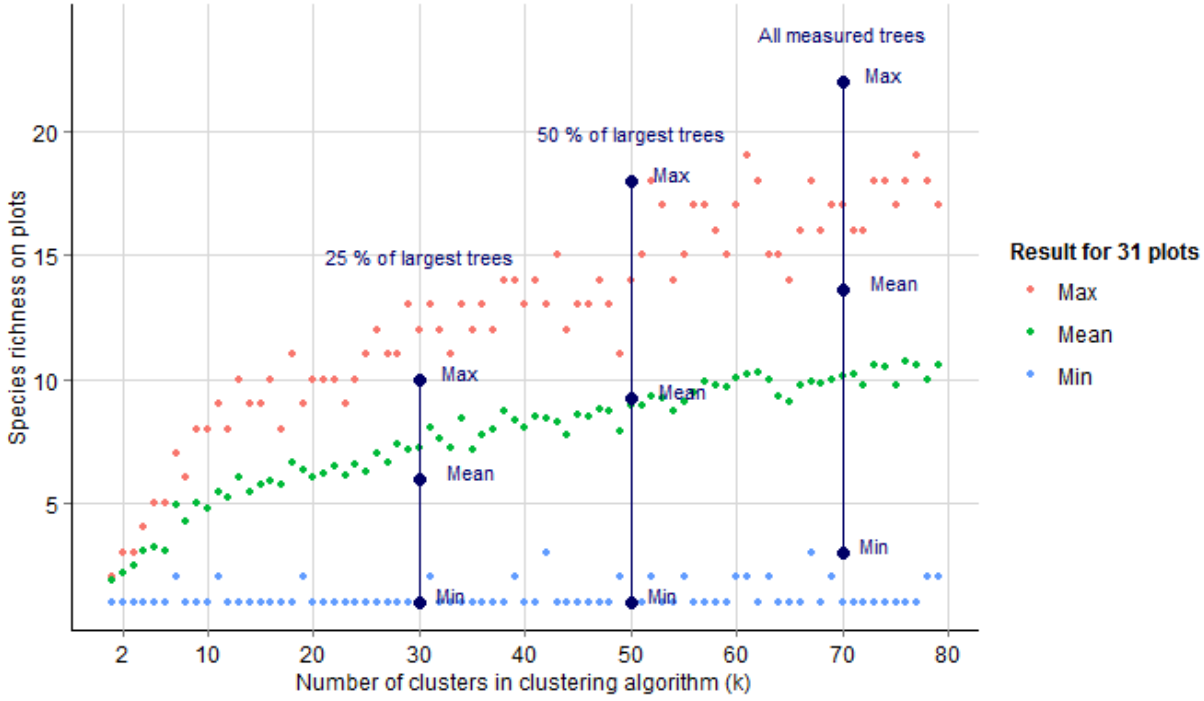


Figure 6-7. Maximum, mean, and minimum species richness values for clustering results with different *k* values. The dark blue dots show the same for the field data.

The plot shows that when 50 % of the largest trees are selected, the number of clusters that gives similar mean species richness values is approximately 45–60. Therefore this range was selected for the second round of clustering.

The Simpson's index values calculated from the clustering results showed good agreement with the field-derived values, with slightly more variation (figure 6-8). The plots that got low values were the pine and cypress plantation plots, similarly to the field data. The increase in the number of clusters (*k*) did not seem to affect much the results, with the exception of the smallest *k* values.

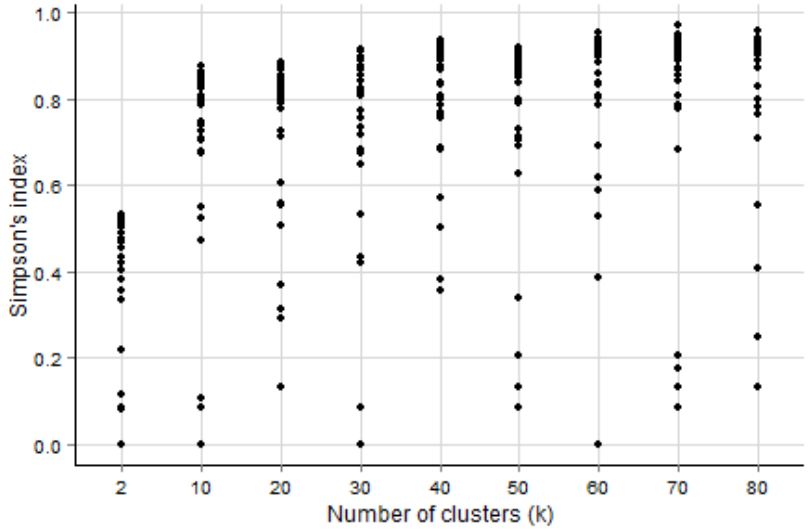


Figure 6-8. Simpson's index values for clustering results with different *k* values.

The Shannon–Wiener index from the clustering results increase with *k* (figure 6-9). The index reaches similar values with the field data when *k* reaches about 30. The lowest values are assigned to the plantation plots as with the field data, but the variation in values of the indigenous plots is larger.

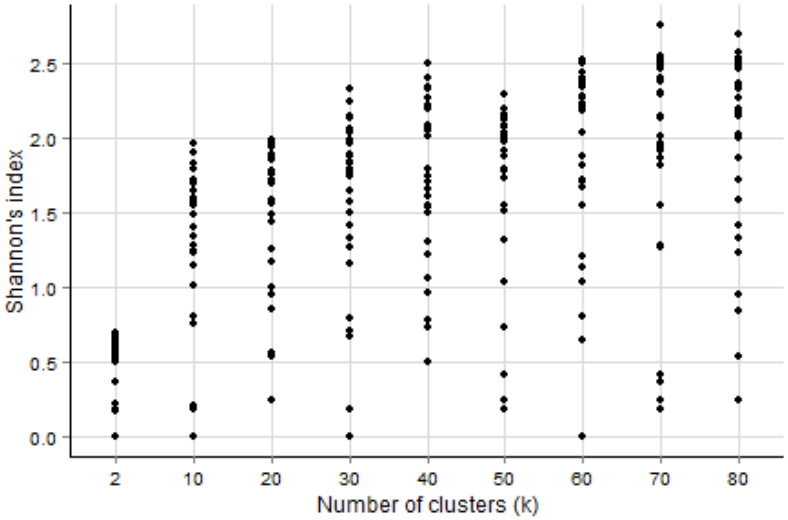
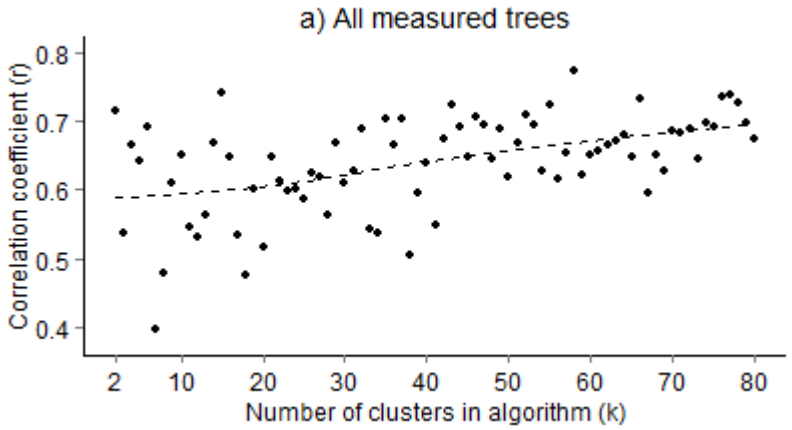


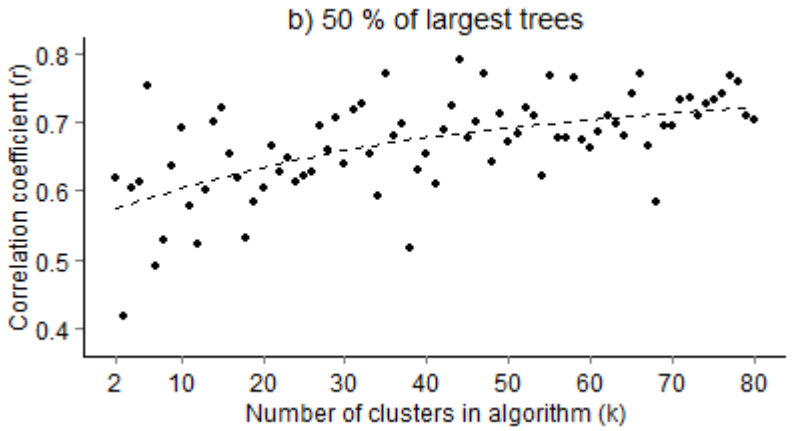
Figure 6-9. Shannon–Wiener index values for clustering results with different *k* values.

6.3 The effect of number of clusters on the relationship between biodiversity measures

Correlation analyses revealed positive correlations in all cases between the diversity measures from field data and from clustering, most of them significant. With the first round of clustering, the values of the Pearson's correlation coefficient had large variation due to the varying clustering results (figures 6-10 to 6-12). Nevertheless, similar trends could be observed with all tree sizes and diversity measures. The correlations improved with increasing number of clusters, although after around 40 clusters the improvement was small.

The correlation coefficients were generally slightly higher when not all the measured trees were included, and slightly higher for the diversity indices than for species richness. When all measured trees were considered, the two indices showed slightly better correlations with the smallest *k* values (figures 6-11 a and 6-12 a).





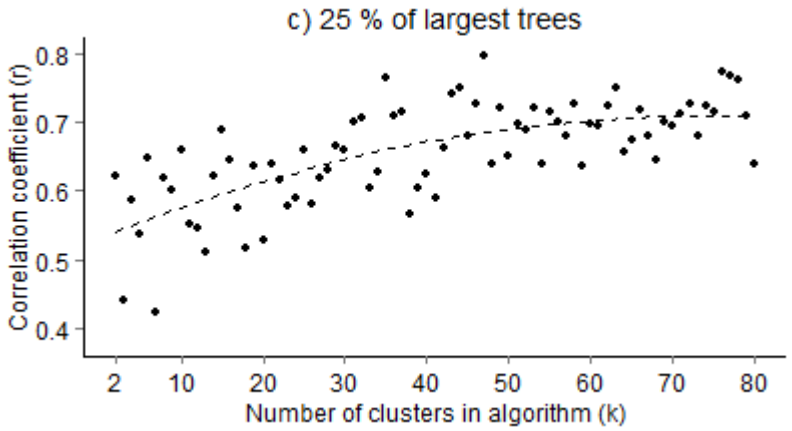


Figure 6-10 a – c. The effect of k on the correlation between species richness values from field data and from clustering.

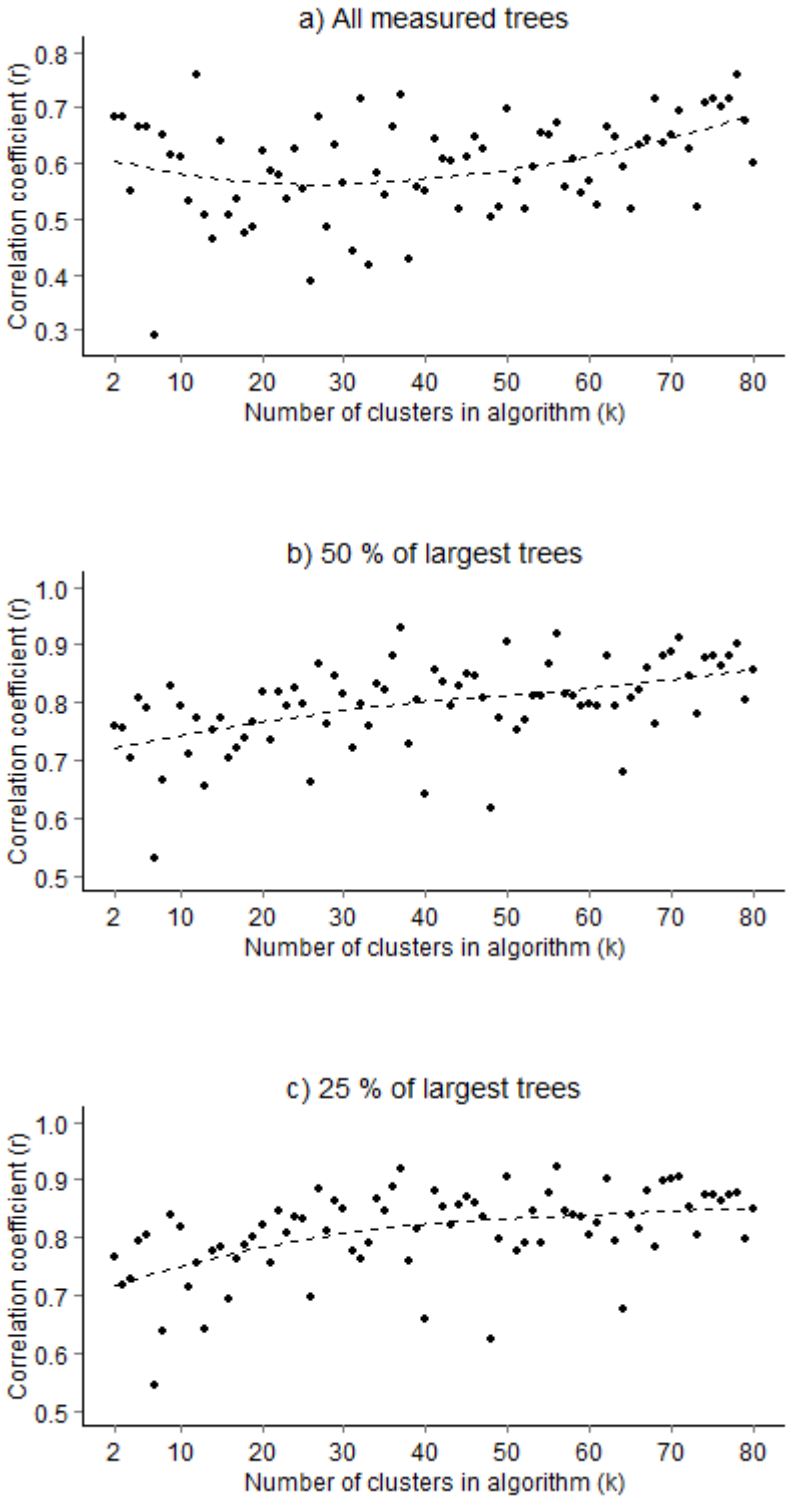


Figure 6-11 a – c. The effect of k *on the correlation between the Simpson's index values from field data and from clustering.*

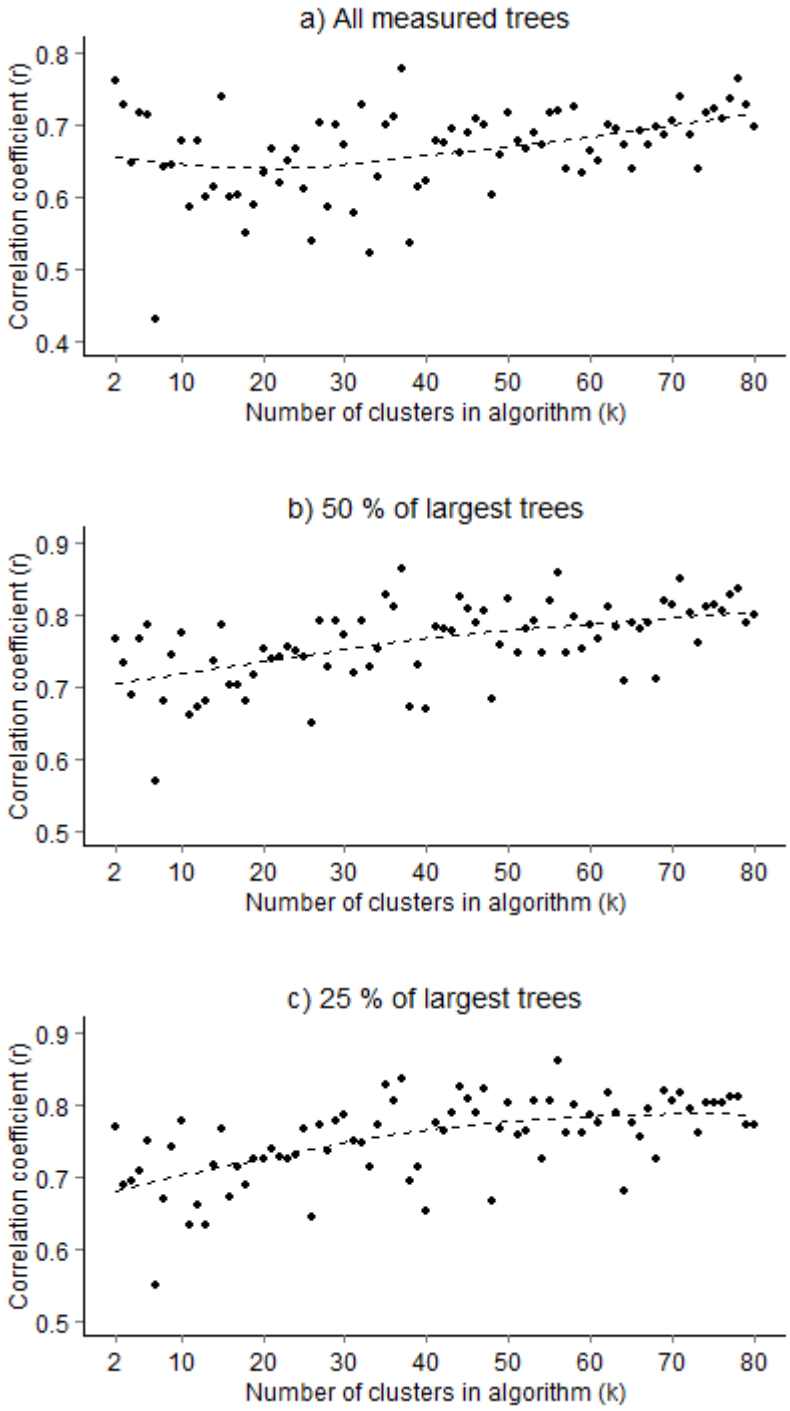


Figure 6-12 a – c. The effect of *k* on the correlation between the Shannon–Wiener index values from field data and from clustering.

Figures 6-13 to 6-15 show the Pearson's correlation coefficients of the analysis of the second round of clustering. The effect of *k* was now studied for the relationship between diversity measures obtained from field data and from the mean clustering results of 100 iterations. The

number of clusters did not seem to have a significant effect on the correlation coefficients, and the variation in the correlation coefficients was minimal. However, including only the larger trees improved the correlation coefficients in all cases, and considerably so with the Simpson's index.

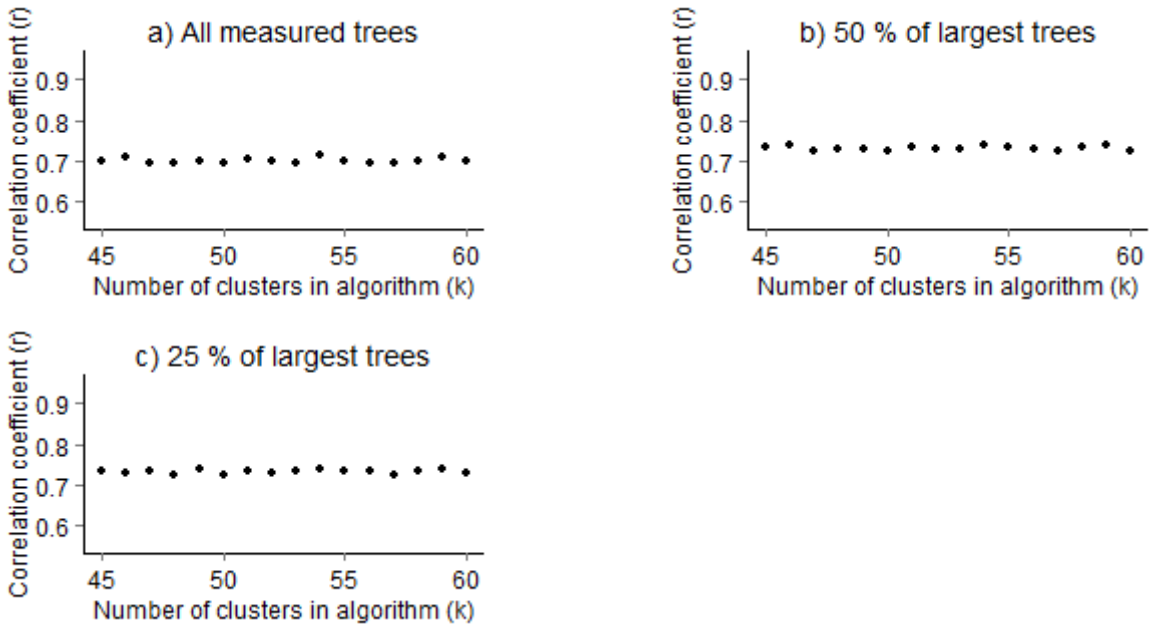


Figure 6-13 a – c. Effect of k on correlations between species richness from field data and from mean of 100 clustering results.

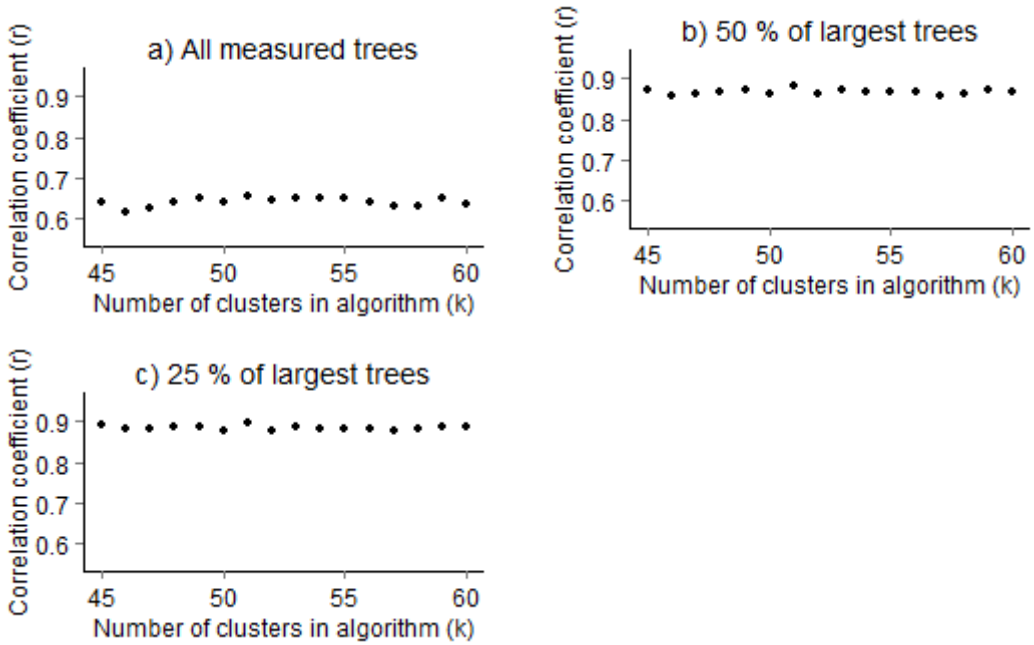


Figure 6-14 a – c. Effect of *k* on correlations between Simpson's index from field data and from mean of 100 clustering results.

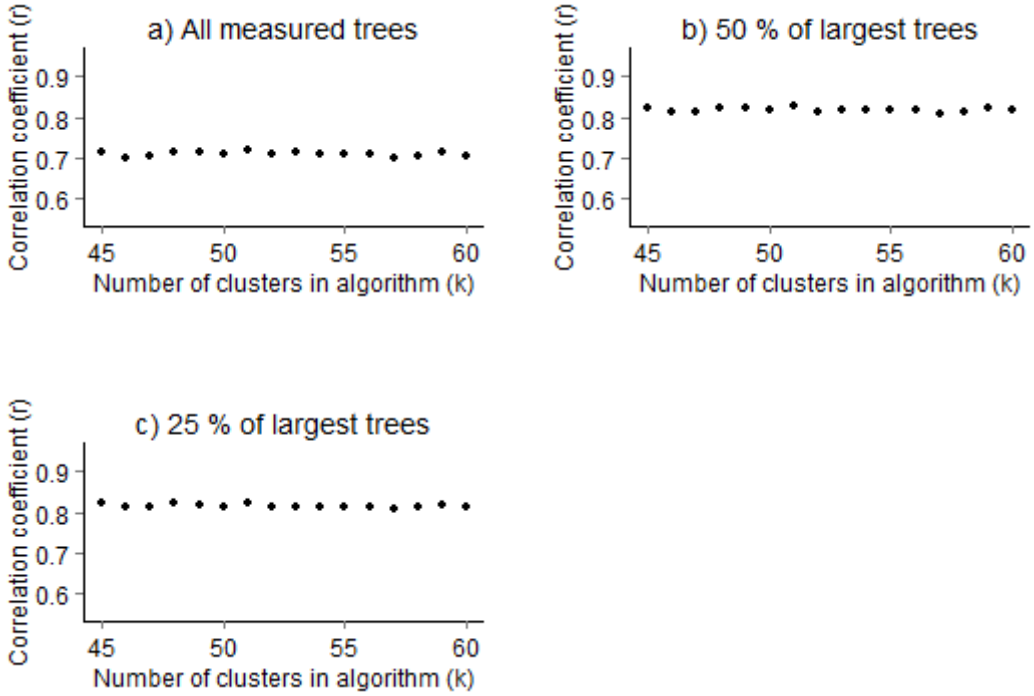


Figure 6-15 a – c. Effect of *k* on correlations between the Shannon–Wiener index from field data and from mean of 100 clustering results.

6.4 Modelling of species diversity measures

Linear regression models were fit to the relationships between diversity measures from field data and from the mean result of 100 clusterings. The coefficients of the models for species richness are shown in tables 6-1 to 6-3. The coefficients of determination (r^2) and the RMSE values do not seem to be affected by the number of clusters. In contrast, with all tree sizes the intercepts show a slight increasing trend and the slope mostly decreases with an increasing number of clusters. The intercept closest to 0 and the slope closest to 1 were obtained when 50 % of the largest trees were included and the number of clusters was 46.

Table 6-1. Coefficients of linear regression models for species richness with all measured trees.

X	intercept	slope	r^2	p-value	RMSE
Spr_mean_k45	4.99	1.01	0.49	< 0.001	3.46
Spr_mean_k46	4.93	1.03	0.50	< 0.001	3.44
Spr_mean_k47	5.03	1.01	0.48	< 0.001	3.49
Spr_mean_k48	5.17	0.97	0.48	< 0.001	3.49
Spr_mean_k49	4.97	1.00	0.49	< 0.001	3.46
Spr_mean_k50	5.17	0.96	0.48	< 0.001	3.49
Spr_mean_k51	5.08	0.96	0.49	< 0.001	3.46
Spr_mean_k52	5.15	0.95	0.49	< 0.001	3.48
Spr_mean_k53	5.19	0.94	0.48	< 0.001	3.49
Spr_mean_k54	5.04	0.95	0.51	< 0.001	3.41
Spr_mean_k55	5.15	0.93	0.49	< 0.001	3.46
Spr_mean_k56	5.24	0.92	0.48	< 0.001	3.49
Spr_mean_k57	5.16	0.92	0.48	< 0.001	3.49
Spr_mean_k58	5.10	0.92	0.49	< 0.001	3.45
Spr_mean_k59	5.03	0.92	0.50	< 0.001	3.42
Spr_mean_k60	5.24	0.90	0.49	< 0.001	3.48

Table 6-2. Coefficients of linear regression models for species richness with 50 % of largest trees.

X	intercept	slope	r^2	p-value	RMSE
Spr_mean_k45	1.16	0.95	0.54	< 0.001	2.96
Spr_mean_k46	1.14	0.96	0.54	< 0.001	2.95
Spr_mean_k47	1.23	0.94	0.53	< 0.001	2.99
Spr_mean_k48	1.35	0.91	0.53	< 0.001	2.99
Spr_mean_k49	1.22	0.93	0.53	< 0.001	2.98
Spr_mean_k50	1.34	0.90	0.53	< 0.001	2.99
Spr_mean_k51	1.28	0.90	0.54	< 0.001	2.97
Spr_mean_k52	1.33	0.89	0.53	< 0.001	2.98
Spr_mean_k53	1.34	0.89	0.53	< 0.001	2.98
Spr_mean_k54	1.28	0.88	0.55	< 0.001	2.93
Spr_mean_k55	1.32	0.87	0.54	< 0.001	2.97
Spr_mean_k56	1.37	0.87	0.53	< 0.001	2.97
Spr_mean_k57	1.36	0.86	0.53	< 0.001	3.00
Spr_mean_k58	1.29	0.86	0.54	< 0.001	2.96
Spr_mean_k59	1.23	0.86	0.55	< 0.001	2.93
Spr_mean_k60	1.49	0.83	0.52	< 0.001	3.01

Table 6-3. Coefficients of linear regression models for species richness with 25 % of largest trees.

X	intercept	slope	r^2	p-value	RMSE
Spr_mean_k45	0.72	0.61	0.54	< 0.001	1.97
Spr_mean_k46	0.75	0.61	0.53	< 0.001	1.99
Spr_mean_k47	0.69	0.62	0.54	< 0.001	1.96
Spr_mean_k48	0.85	0.58	0.52	< 0.001	2.00
Spr_mean_k49	0.68	0.61	0.54	< 0.001	1.96
Spr_mean_k50	0.83	0.58	0.52	< 0.001	1.99
Spr_mean_k51	0.78	0.58	0.54	< 0.001	1.98
Spr_mean_k52	0.81	0.58	0.53	< 0.001	1.99
Spr_mean_k53	0.80	0.58	0.54	< 0.001	1.98
Spr_mean_k54	0.79	0.57	0.55	< 0.001	1.95
Spr_mean_k55	0.80	0.57	0.54	< 0.001	1.98
Spr_mean_k56	0.82	0.56	0.54	< 0.001	1.96
Spr_mean_k57	0.84	0.55	0.52	< 0.001	2.00
Spr_mean_k58	0.76	0.56	0.54	< 0.001	1.96
Spr_mean_k59	0.75	0.56	0.55	< 0.001	1.95
Spr_mean_k60	0.86	0.54	0.53	< 0.001	1.97

Figure 6-16 illustrates the relationship of species richness measured in the field and obtained from clustering, when the number of clusters was 46, and 50 % of the largest trees were considered. The modelled species richness is clearly related to the measured one. Although the values do not exactly correspond to each other, there are no remarkable outliers. The plantation plots mostly have the lowest species richness values in both datasets.

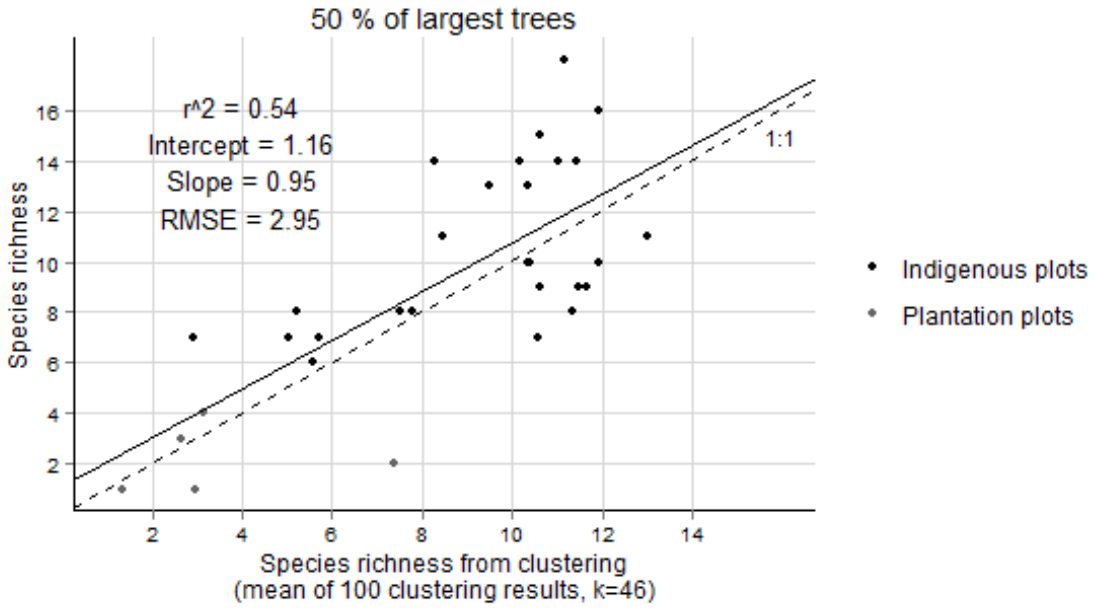


Figure 6-16. The relationship of species richness obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered.

For the two diversity indices, similar trends of the slope and intercept of the models were observed as with species richness when the number of clusters increased (not shown). The coefficients of determination were higher as with species richness, but figures 6-17 and 6-18 show that the assumptions for a linear regression were not met. As can be seen in the relationship for Simpson's index (figure 6-17), the low and high index values correspond reasonably well. But because most of the plots get high values, the data is not normally distributed. In addition, the residuals are not homoscedastic because they deviate from the regression line most in the middle.

Figure 6-18 illustrates the relationship between Shannon–Wiener index from clustering and from the field data, when the number of clusters was 46 and 50 % of the largest trees were considered. The high index values are reasonably well predicted by the clustering, but the few lower and middle values not as well. Also here, the assumptions of a normal distribution and homoscedastic residuals are not met.

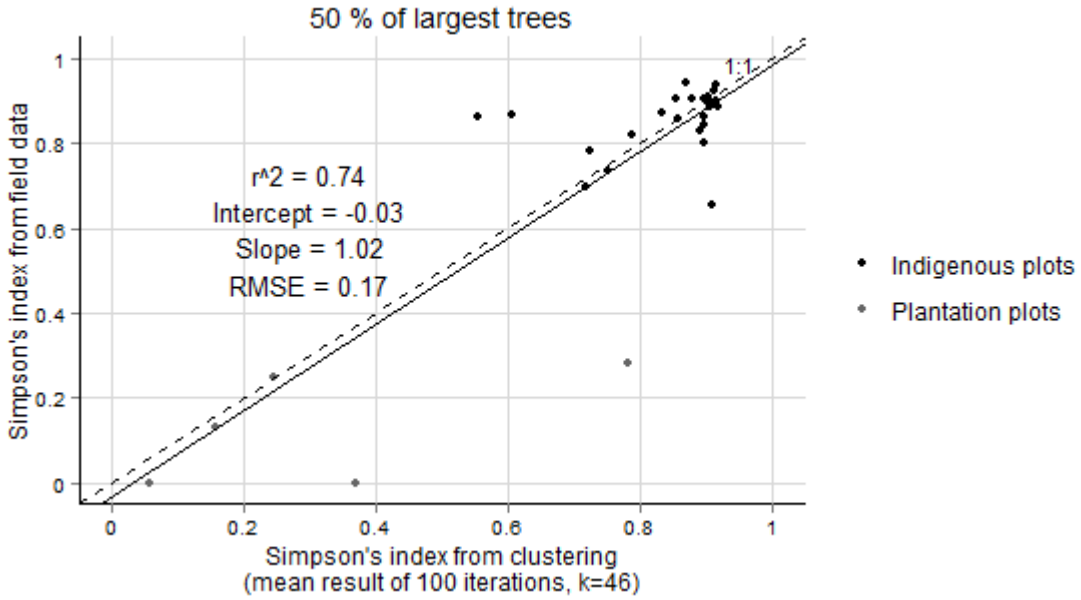


Figure 6-17. The relationship of Simpson's index obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered.

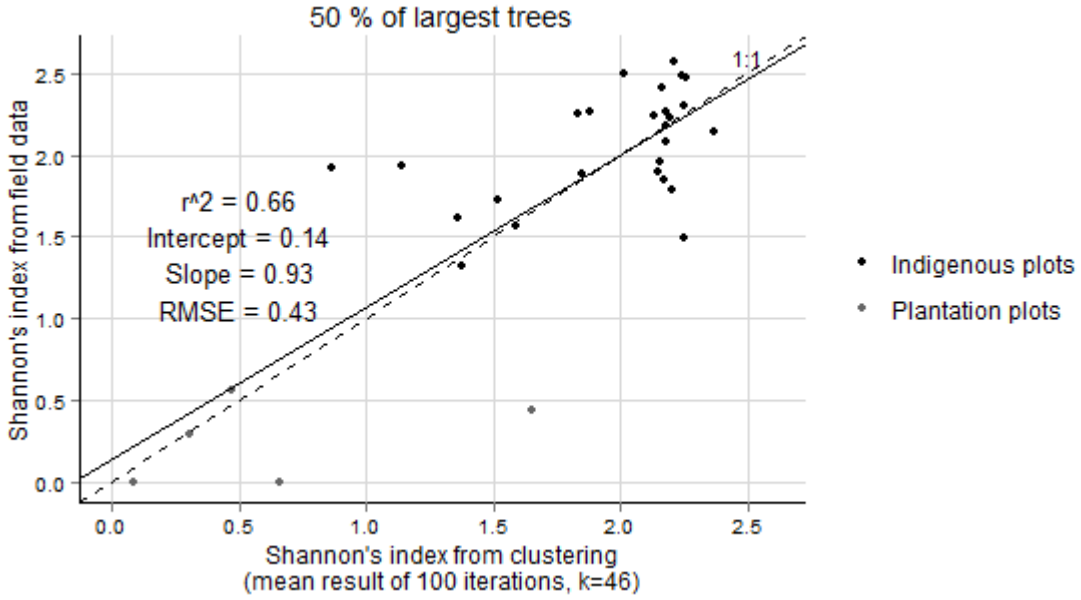


Figure 6-18. The relationship of the Shannon–Wiener index obtained from field data and from clustering, when the number of clusters was 46 and 50 % of the largest trees were considered.

6.5 Tree species richness map of Ngangao

The model for mapping tree species richness for the whole study area was based on the relationship in figure 6-16. Species richness for each pixel in the map was calculated as in the following equation:

$$\text{Species richness} = 0.96 \times \text{Spr\_50} + 1.14 \qquad \text{Equation 4.}$$

where Spr_50 was the number of clusters found in the 0.1 ha neighbourhood of the pixel. The final tree species map, which was calculated as the mean value of 5 maps created using this equation, is shown in figure 6-19.

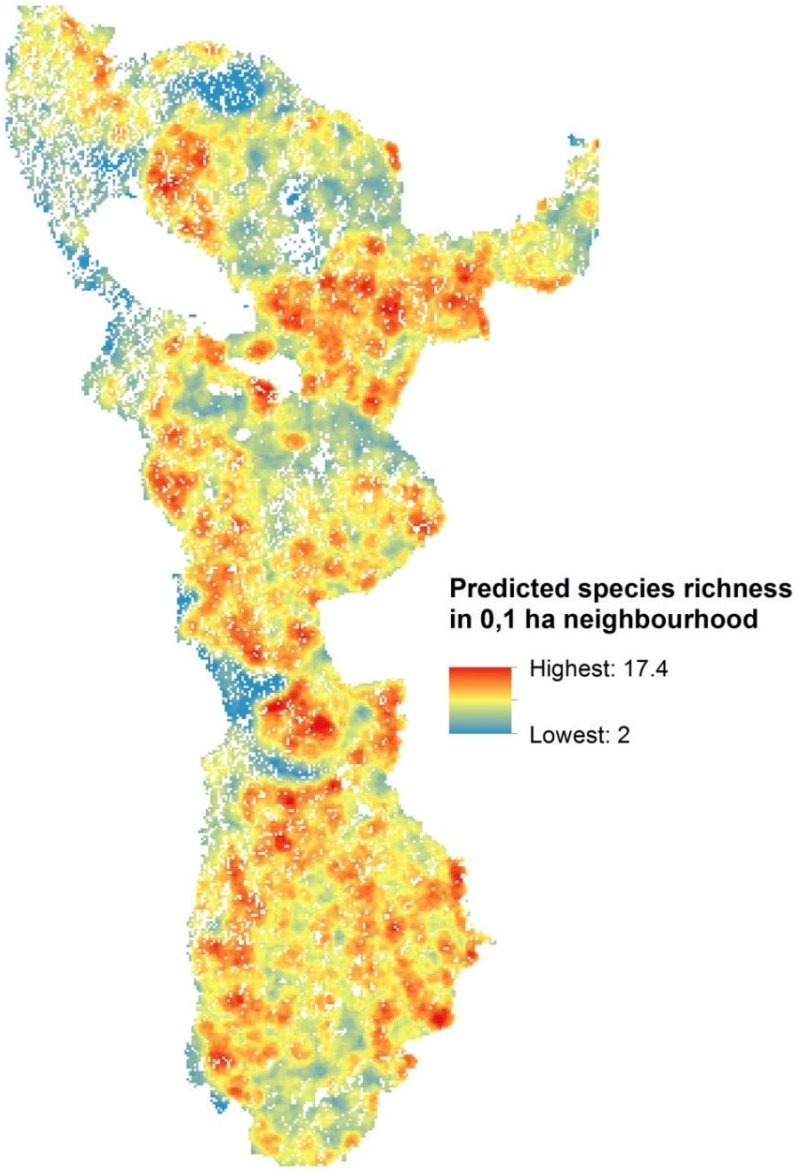


Figure 6-19. Tree species richness in the study area as the mean prediction of 5 clustering results.

The most prominent feature in the tree species richness map are the low-richness areas corresponding to plantation sites. Exotic plantation forests are well distinguished from the indigenous forest by their species richness. Also the Northwestern parts of indigenous forest have low tree species richness.

Otherwise the map is characterized by patches of high and low richness. More areas of high richness occur in the middle and Western parts of the forest and in certain areas in the North. Comparing to the canopy height model (figure 6-20), patches of lower richness often correspond to areas where trees are large, and the high species richness often occurs at sites where tree height is low.

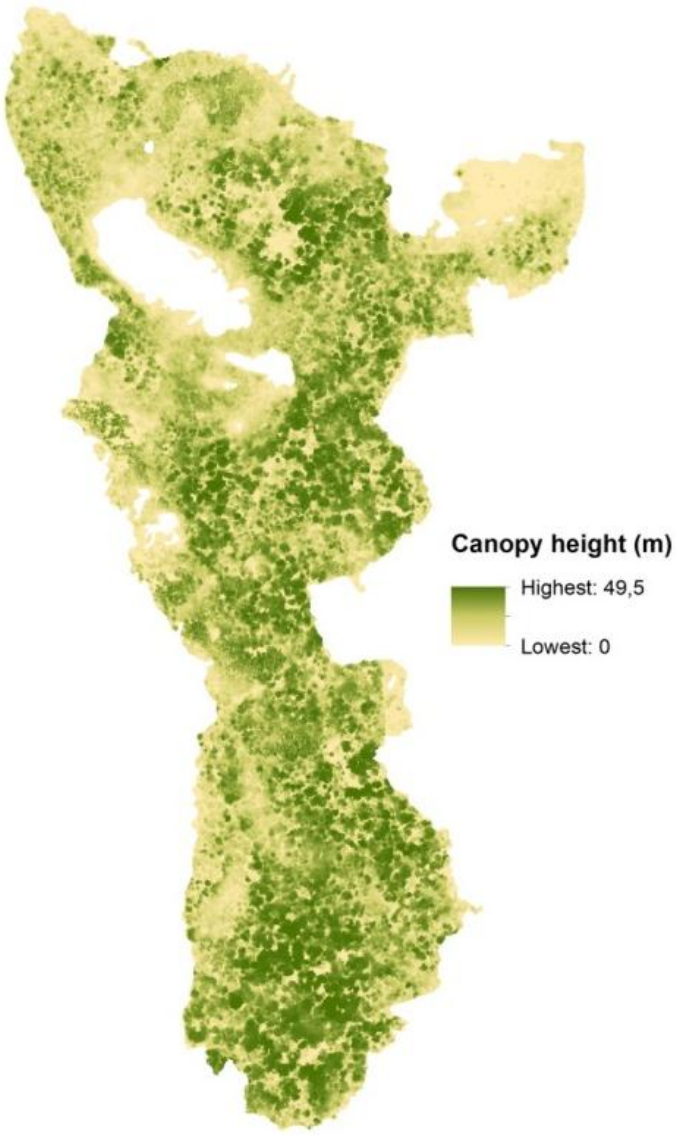


Figure 6-20. Canopy height model of the study area.

7 DISCUSSION

7.1 Performance of the approach in estimating tree species diversity

The first objective of the study was to assess the performance of the applied object-based clustering approach in estimating tree species diversity measures. The approach succeeded well in revealing diversity patterns, as the correlations between field-measured and clustering-derived diversity measures were high.

The approach performed especially well in predicting tree species richness, because it worked with similar accuracy for plots of varying species richness, and for indigenous and plantation forests. Species richness derived from clustering could explain approximately half of the variation of the field-measured species richness, which is a moderate accuracy. However, the predictions were so similar in magnitude with the field measurements of the larger half of the trees, that the clustering-derived predictions could almost be used as such. Here, the model coefficients were used for mapping tree species richness, but almost similar results would be obtained if the slope and intercept were 1 and 0, respectively. This means that even without the use of a model, the approach could estimate species richness of the larger trees with the RMSE of approximately ± 3 species.

With the two diversity indices, the performance of the approach could not be validated well for low and intermediate index values, because most of the indigenous plots got similarly high index values. Both indices could, however, distinguish well between low and high-diversity plots. Another study of tree species diversity in Ngangao similarly observed lower Shannon–Wiener index values in plantation sites than on indigenous sites (Omoro *et al.* 2010).

The second objective of the study was to predict spatial patterns of tree species richness in the study area. In the created map, species richness seemed to follow the patterns of tree height. The result was expected because species richness was estimated for a fixed area, and fewer trees fit on an area if they are large.

7.2 The approach in context of previous research in the field

While remote sensing may be used for modelling biodiversity via environmental variables (Gillespie *et al.* 2008; Turner *et al.* 2003), the approach presented in this study estimates tree species diversity directly, based on spectral variation of imaging spectroscopy data. Most studies that assess diversity from the spectral variation use measures such as the standard deviation of a set of pixels (Rocchini *et al.* 2010). With coarse spatial and spectral resolution data, this is a sensible approach.

However, high spatial resolution imaging spectroscopy data allow for new kind of approaches for biodiversity assessment. Here, species diversity was linked to the variation in reflectance even more directly, by assessing the differences between tree crowns. The approach takes advantage of the capability of imaging spectroscopy to detect spectral differences between tree crowns (Clark 2012). However, unlike in supervised classification of species, no training and validation data are needed with the clustering approach.

Two other studies have estimated canopy alpha diversity very recently using a clustering approach. Medina *et al.* (2013) studied a dry tropical forest in Puerto Rico with a similar AisaEAGLE sensor as used in this study. They applied a hierarchical agglomerative clustering to pure reflectance data, with non-vegetation pixels removed. The correlations between the Shannon–Wiener index values obtained from clustering and from field data were variable and even negative in some cases, but better results were obtained when spectral unmixing was applied after the clustering.

Féret & Asner (2014) used the Carnegie Airborne Observatory imaging spectrometer that covers the range from visible light to SWIR, to study areas in the Peruvian Amazon. After a principal component analysis (PCA), they applied a k-means clustering to randomly selected pixels across an image of selected PCs. The Shannon–Wiener index values were systematically underestimated, but correlations with field-measured Shannon–Wiener index values were high (around 0.83). In the same study, also beta diversity was estimated using Bray–Curtis dissimilarities.

The main differences of this study compared to these two are the object-based approach and the estimation of species richness and Simpson's index in addition to the Shannon–Wiener index. The estimation of all three biodiversity measures was successful, as already discussed.

The object-based approach also proved to be a very viable alternative to a pixel-based clustering.

Because the approach applied in this study clearly succeeded in delineating tree crowns and detecting spectral differences between them, it may even have potential for linking the reflectance properties of individual crowns to the corresponding tree species. This would be an interesting subject for further research.

7.3 Evaluation of factors that affected tree diversity measures

7.3.1 Optimal number of clusters

When the number of clusters in the clustering algorithm increased, the number of clusters (or species richness from clustering) on the plots increased as well. This result was expected because when the algorithm is set to find many clusters, also the data on the plots will more probably be divided to several clusters. This was also reflected by an increase in the Shannon–Wiener index values (figure 6-9), but not so clearly by the Simpson's index values (figure 6-8). This makes sense keeping in mind that the Simpson's index is not very sensitive to changes in rare species (Krebs 2014). Thus for example one additional cluster on a plot should not affect the Simpson's index as much as the Shannon–Wiener index.

The effect of the number of clusters was tested more carefully for the range of 45–60 clusters by averaging the clustering results of 100 iterations. The slight trends that the slope and intercept show with increasing number of clusters (tables 6-1 to 6-3) reflect the same fact, that in average more clusters are found on the plots when more clusters are searched for. However, no trends were observed with the correlation coefficients, which shows that the averaged solutions of the clustering algorithm were quite robust at this range.

In this study, the optimal number of clusters was determined based on best results with field data. An alternative would be to estimate the best number of clusters based on the data itself. A range of techniques exist for determining the optimal number of clusters (e.g. Mirkin 2011). If the optimal number of clusters could be determined in advance on a theoretical basis, the same method could be applied in other cases without having to find the best number of clusters empirically.

For example Medina *et al.* (2013) applied an information criterion to determine at how many clusters their hierarchical agglomerative clustering algorithm should stop. However, their criterion resulted in only 14–16 clusters per scene, which may be the reason why they did not achieve good results in correlation analyses with field data. The optimal of number of clusters defined from a theoretical basis is not necessarily the clustering solution that is searched for, because tree crowns do not necessarily form well-defined clusters by their spectral properties.

The results of this study suggest that the species accumulation curve may give indications of a suitable range for the number of clusters. The clustering results best matched the field diversity data when the number of clusters was around 45–60. Looking at the species accumulation curves (figures 4-5), the number of species will probably keep rising at a slowing rate when a larger area is sampled. The range 45–60 would not seem far from what the total number of species in the forest could be (with 50 % of the largest trees), although exact estimations depend very much on the chosen method and are therefore not recommended (Krebs 2014). However, the averaged clustering solutions were not affected very much by the exact number of clusters in the range, so the exact estimation does not seem important.

7.3.2 Optimal tree size

The size of trees in the field data affected the relationships between diversity measures from field data and from clustering. In general, the best relationships were observed when 50 % of the largest trees on the plots were considered. Considering only 25 % of the largest trees also gave better results than considering all measured trees (DBH ≥ 10 cm).

Dropping smaller trees from the evaluation was expected to improve the results, because it was observed in the field that not all measured trees reached the top of the canopy. Only tree crowns visible to the airborne sensor could contribute to the diversity measures that were based on clustering of the MNF transformed imagery.

Limiting the size of the trees under consideration could be done at least in two ways. Here, a percentage threshold was used where 50 % or 25 % of the largest trees on the plots in terms of DBH were considered. This caused the actual size limit to vary between the plots. Another way would be to set constant size limits of e.g. DBH ≥ 20 cm or 30 cm, as done for instance by Imai *et al.* (2014). If an optimal size limit was found, it could give implications of what

size of trees is relevant in field data collections for remote sensing applications. However, Imai *et al.* (2014) concluded that raising the standard DBH limit from 10 cm to 20 cm was not recommendable, because the average size of trees varied between plots. This was the case also in this study, and therefore a percentage limit was considered more suitable. In addition, even if a constant optimal size limit was found in this study, it would probably not be applicable in other forests.

Restricting the limit from 50 % to 25 % of the largest trees did generally not have much effect on the results. The reason for this is probably that setting a limit for the DBH did not result in exact selection of those trees that reached the canopy, no matter what the limit was. This was indicated by field observations that showed that the DBH did not determine whether a tree reaches the canopy. For example, trees of the shade-tolerant pioneer species *Tabernaemontana stapfiana* rarely reached the canopy even if their DBH was large. Also, when very large emergent trees occurred on the plots, their crowns covered most of other relatively large trees.

The suitability of tree size limits probably depended to some extent on the relationship between the number of trees and the number of segments on the plots (figure 6-4). This comparison showed that including either 50 % or 25 % of the largest trees in the field corresponded better to the number of segments than including all trees. However, the exact size limit did not seem to be important, because the individual plots showed no relationship between the number segments and the number of trees.

7.3.3 Sources of error in the relationships between field data and clustering results

The relationships between diversity measures from field and from clustering probably contained some error, because the match between the data from the two sources was not perfect. One aspect was the accuracy of the georeferencing, which was lower in areas that fell near the edges of flightlines. However, it affected only the in the areas where two flightlines overlap, and in the worst case it may have resulted in the absence of some smaller tree crowns.

Another issue was the difficulty in selecting the trees that reach the canopy, mentioned in the previous section. This probably resulted in more error, because it caused a mismatch in the trees that were measured in the field and the segmented crowns on the image. The only way to

overcome this discrepancy would be to record those trees in the field that reach the canopy level. This was also intended in the field work, but proven very difficult in practice. In forests such as this one, lianas and the multiple layers of the canopy often inhibit observing the highest canopy level from the ground.

Another source of error is the fact that the imagery had gaps with no data after shadow removal. Although shadow removal was mainly targeted at removing shaded portions of tree crowns, some larger gaps occurred where topography or other reasons caused larger shady areas. At least one plot was missing entire tree crowns because of the gaps. Minor error may have been caused also by the fact that field plots were circular on the image, but in reality the slope of the terrain made the plots seem rather ellipsoidal when viewed from above. Tree crowns also crossed plot borders, while field measurements were made only to the trees that had their trunk inside the plot. However, on relatively large plots such as these (0.1 ha) the number of trees is high enough for this effect to be small.

Species identification in the field is also disposed to errors, especially in species-rich tropical forests. It is possible that not all the trees were identified correctly. For instance, one very common species in this study was identified as *Macaranga capensis*, but Mbuthia (2003) listed *Macaranga conglomerata* as one of the most common species in the same study area. The number of unidentified trees was very low (3 %), but otherwise it is impossible to estimate the error in species identification.

Altogether, these factors result in the situation that the trees that produced the field-derived diversity measures were not exactly the same trees that the clustering-derived diversity measures were based on. However, they affected only the accuracy of the relationships which the models were based on, but not the clustering itself.

7.3.4 Factors that affected species discrimination

The capability of the approach to detect spectral differences between species was the other main source of error. A weakness in the applied methodology was that the spectral differences across the image had also other sources than the spectral differences between tree species. Indications of this could be seen in the clustering results with only ten clusters (figure 6-5 b).

The patterns in the clustering give reason to assume that shadows still played a role in the MNF transformed image. The patterns of the cluster 2 (blue) follow not only the distribution of coniferous plantation forests, but also the distribution of shadows due to topography – the Northwestern corner that lies on a west-facing slope differs from the rest of the indigenous forest. Therefore, also individual crowns split to several segments may have been assigned to several clusters, if the brightness between the segments differed much.

Shadow removal by applying a brightness threshold was a simple solution, and other methods for minimizing brightness variation could be considered for improving the approach. For instance, Féret & Asner (2014) applied a method called continuum removal transformation. Another solution may be calculating vegetation indices from the image, which are sensitive to differences in vegetation biochemistry and structure, but not to illumination conditions. However, this would not suit the clustering approach, which requires the units being comparable for dissimilarity calculations.

Nevertheless, the amount of shade in a tree crown may also be a means to discriminate between species, because it is affected by crown structure (Clark 2012). For example the coniferous forest parts have a very different canopy structure than the broadleaved parts, which probably affects the amount of shade. This possibly contributed to the good separation of the coniferous plantation parts in the final clustering. Therefore, eliminating shade completely from the imagery may cause loss of information relevant to the clustering.

Across-scene spectral variation was also added to the image by the artefacts caused by differences between flightlines. Especially in the clustering results of the Northern part of the forest it can be seen that one horizontal strip has more clusters of one colour than others. The reason is probably related to the atmospheric correction, which considerably diminished visible differences between flightlines, but it is not clear why this particular flightline still differed from the others after atmospheric correction.

Another factor that affects the feasibility of the approach is the question whether tree species are spectrally dissimilar enough that they can be distinguished. No spectral measurements were made for the species in this study area, but studies in tropical forests in Hawaii (Asner & Martin 2009), Australia (Asner *et al.* 2009), and Amazonia (Asner & Martin 2011) have shown that tree species often have unique spectral signatures on the leaf level. The spectral signatures have also been linked to the chemical properties of the leaves, and attempts have been made to use the information for species classification on the canopy level (Clark *et al.*

2005; Féret & Asner 2012; Féret & Asner 2013). Based on these studies, there is reason to assume that also the species of this tropical forest show spectral dissimilarities. However, the imaging spectroscopy data of this study did not cover the SWIR range, which has been shown useful in species discrimination (Clark *et al.* 2005; Asner *et al.* 2009).

The capability of the MNF transformation to detect spectral differences between species affected both the segmentation and the clustering. The results showed that where tree crowns were visually separable, the segmentation mostly made the same delineations for tree crowns as a human eye makes based on the first three MNF bands. Therefore the segmentation seems to have succeeded well. Keeping in mind that the algorithm can handle the information from 13 bands, whereas a human eye can see only three bands at a time, the algorithm probably performed well also in the areas where visual tree crown delineation was difficult.

In the clustering phase, segments were grouped together based on the similarity of their mean MNF values. Only for the plantation sites it was possible to evaluate how well a cluster represented a true species. The pine plantations were consistently assigned to one cluster even when the number of clusters was high. The crowns in the cypress plantations, however, were assigned to several clusters, partly the same as the pines (figure 6-6). This may be due to the varying degrees of topographic shade, and possibly to the visibility of the background through the sparse cypress canopy. Accordingly, the second cypress plot was an outlier in the scatterplots of the relationships between diversity measures from field data and from clustering.

Lianas and other epiphytes probably added confusion to the segmentation and clustering of the crowns. Epiphytes add to the spectral variation of the canopy, and may grow on top of several neighbouring trees (Clark 2012). Lianas were very common in the study site, so they probably affected the reflectance of the crowns, and added uncertainty to their segmentation. Another confusing factor may be phenological variation within the same species (e.g. flowering), which may have caused crowns of the same species to be assigned to different clusters. These factors are very difficult to control for, but their effect on tree species discrimination has been recognized (Clark 2012).

In general, the asset of the object-based MNF transformation approach applied in this study is that it takes into account all variation between tree crowns, including spectral effects depending on crown structure (such as shade and leaf angle distribution). In contrast, e.g. vegetation indices measure only the differences that they are designed to detect. In cases like

this, it is not known in advance which spectral features differentiate the species. Therefore it is an advantage to use a method that finds the distinguishing features from the available spectroscopic image itself.

The shortcoming of the MNF transformation approach is that we don't know exactly what causes the variation across the image, and therefore it is difficult to eliminate undesired variation such as shadows and image acquisition artefacts. However, diversity was assessed here at a local scale, which emphasizes local differences rather than overall variation across a large extent. Therefore the topographic shadows and flightline differences should not have much effect on the species richness estimates of the final map.

7.4 Diversity measures as indicators of conservation value

As brought up in the introduction, remote sensing provides possibilities for assessment of biodiversity that can be combined with carbon stock estimations. Such approaches are necessary for avoiding carbon-biodiversity tradeoffs in e.g. REDD+ projects (Phelps *et al.* 2012). However, it is worth consideration which measures of biodiversity are the most meaningful from a conservation perspective. At the same time, it has to be kept in mind that all research is limited to study only some aspects of the broad concept of biodiversity; in this case, it was tree species diversity.

The two diversity indices used here are designed to take into account the evenness in abundance of the species. However, their usefulness as descriptors of natural communities can be questioned, because natural communities are usually not even. Rather, they typically consist of few abundant species, while most species occur more occasionally (Gaston & Blackburn 2000). This was the case also in Ngangao, although the species composition of the forest has human influence. As the occurrence of rare species may well be one criterion for high conservation priority, these indices are probably not suitable measures of biodiversity if they "punish" for the occurrence of rare species.

However, here both indices were useful for making a difference between plantations and indigenous forest. The differences between index values between the two forest types were much larger than the difference in species richness. Therefore the indices succeeded in describing species diversity in a meaningful way: although some plantation plots had almost

as many species as some indigenous plots, they were dominated by one species to such extent that the overall diversity was low.

Species richness also successfully made a difference between plantations and indigenous forest. Otherwise, it is unclear whether locally species rich sites are the most important for conservation. For example, the map reveals areas of high species richness at the edges of cliffs and at other sites where trees are small. Tree species richness on a fixed area is of course affected by the size of the individuals – areas with large trees now may have lower species richness simply because they have fewer trees. Therefore a measure such as Fisher's alpha (Fisher *et al.* 1943) could be a good alternative for species richness, as it neutralizes the effect of sample size (Krebs 2014).

Effective biodiversity conservation networks include not only species rich sites, but those whose species compositions complement each other (Pressey *et al.* 1993; Howard *et al.* 1998). Therefore beta diversity, or the turnover in species composition between sites, would be another interesting measure for biodiversity estimation. In addition, there are indications that priority conservation sites selected based on tree species diversity represent priority conservation sites of other taxa as well, because of complementary site selection (Howard *et al.* 1998). Tree community composition, which beta diversity estimations are based on, has also been shown to be related to aboveground biomass measurements (Imai *et al.* 2014). Some of the abovementioned studies have already attempted to measure beta diversity with different approaches based on imaging spectroscopy (Higgins *et al.* 2014; Baldeck & Asner 2013; Féret & Asner 2014). Also the approach presented in this study may well be used for estimating beta diversity, once the challenge of systematic variation in reflectance across the scene, caused by topography and data acquisition artefacts, is overcome.

At the scale of one relatively small forest fragment, however, all three measures used here are good descriptors of the spatial patterns of tree species diversity. They describe well the low diversity of plantation sites compared to indigenous forest, and bring up small-scale variation in tree species richness which would be interesting to explain in detail in further research on the study area.

8 CONCLUSIONS

In this study, the use of airborne imaging spectroscopy was studied for estimating tree species diversity in a tropical montane forest. The performance of a clustering-based approach for estimating species richness and two biodiversity indices was assessed. Figure 8-1 illustrates the processing flow.

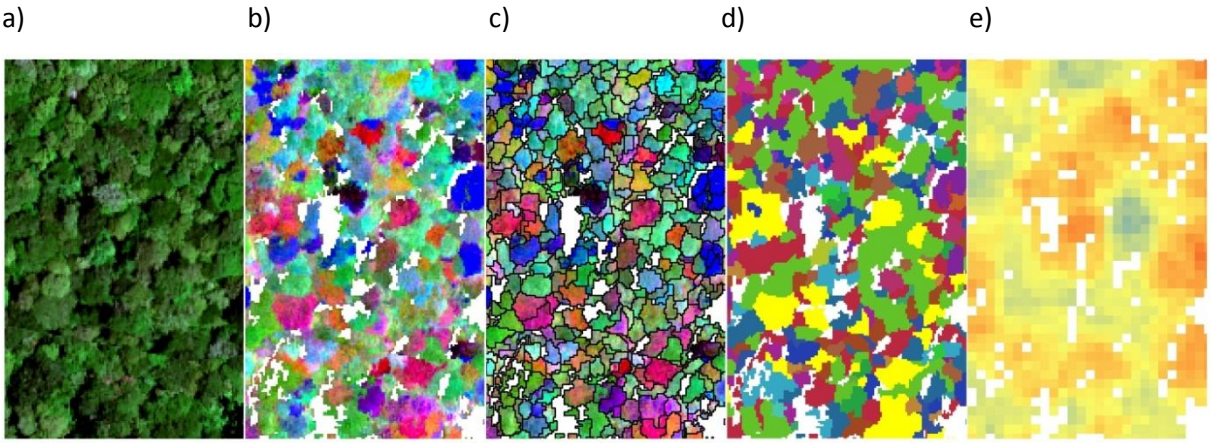


Figure 8-1. From left to right, the figure shows a fraction of the study area a) as a true-colour composition of the spectroscopic image, b) after shadow removal and MNF transformation, c) after spectral segmentation, d) with clustering results, e) with predicted species richness.

It was found that the approach succeeded well in revealing tree species diversity patterns with all three diversity measures. Despite factors that added error to the relationship between field-derived and clustering-derived diversity measures, high correlations were observed. Especially tree species richness could be well predicted based on the approach, and a tree species richness map was created for the study area.

A challenge in the approach was how to accurately link the field measurements to the remotely sensed measurements of the canopy. This issue was addressed by testing three different tree size limits. Another main challenge was the undesired variation in reflectance that the spectroscopic imagery exhibited. Much of it was removed by applying a shadow removal and an atmospheric correction, but some effects remained. However, the local-scale species diversity estimates were probably not affected by across-scene topographic shadows and data acquisition artefacts. If the challenges can be overcome, the approach has potential for estimating beta diversity as well.

The study builds on previous research in remote sensing of biodiversity, and presents a very direct approach for estimating tree species diversity from high-resolution imaging spectroscopy data. Understanding biodiversity patterns is essential for effective conservation strategies, and as the approach is based on remote sensing, it can well be combined with remote sensing based carbon stock estimations. With further development, the presented approach has potential for other interesting applications, such as estimation of beta diversity, and tree species identification by linking the reflectance properties of individual crowns to their corresponding species.

9 ACKNOWLEDGEMENTS

The study and its materials got financial support from the BIODEV research project (Building Biocarbon and Rural Development in West Africa), funded by the Ministry of Foreign Affairs in Finland. A travel fund for conducting the field work was granted by the Department of Geosciences and Geography. The Taita Research Station of the University of Helsinki in Wundanyi, Kenya, provided excellent maintenance and help during field work.

I am thankful to my supervisors prof. Petri Pellikka and Dr. Janne Heiskanen for the opportunity to write my thesis as part of an inspiring research team. I hope this is not the last time I can work with a person like Janne who I thank for excellent guidance and encouragement throughout the process. I am also grateful to the friendly and helpful staff of the Taita Research Station, especially Mr. Darius Kimuzi who did long days of field work with us and whose expertise in tree species identification made the study possible. I also thank Vuokko Heikinheimo for the long days of field work and for driving the car, as well as others doing their thesis on the Station for excellent company. I thank Rami Piiroinen and Tuure Takala for assistance in preprocessing the imaging spectroscopy data, and everyone involved in the remote sensing research team who made this study possible. At last, I thank my loved ones for support during the long process of making the thesis.

10 REFERENCES

Aerts, R., K. Thijs, V. Lehouck, H. Beentje, B. Bytebier, E. Matthysen, H. Gulinck, L. Lens & B. Muys (2011). Woody plant communities of isolated Afromontane cloud forests in Taita Hills, Kenya. *Plant Ecology* 212: 4, 639-649.

Asner, G. P., M. O. Jones, R. E. Martin, D. E. Knapp & R. F. Hughes (2008). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment* 112: 5, 1912-1926.

Asner, G. P. & R. E. Martin (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytologist* 189: 4, 999-1012.

Asner, G. P. & R. E. Martin (2009). Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment* 7: 5, 269-276.

Asner, G. P., R. E. Martin, A. J. Ford, D. J. Metcalfe & M. J. Liddell (2009). Leaf chemical and spectral diversity in Australian tropical forests. *Ecological Applications* 19: 1, 236-253.

Bajwa, S.G. & Kulkarni, S.S. 2012, "Hyperspectral data mining." in *Hyperspectral remote sensing of vegetation*, eds. P.S. Thenkabail, J.G. Lyon & A. Huete, Taylor & Francis Group, Boca Raton, pp. 93-120.

Baldeck, C. A. & G. P. Asner (2013). Estimating vegetation beta diversity from airborne imaging spectroscopy and unsupervised clustering. *Remote Sensing* 5: 5, 2057-2071.

Beentje, H. J. & N. Ndiang'ui (1988). An ecological and floristic study of the forests of the Taita Hills, Kenya. 1. Ecology of the forests. *Utafiti* 1: 2, 23-42.

Bunting, P. & R. Lucas (2006). The delineation of tree crowns in Australian mixed species forests using hyperspectral compact airborne spectrographic imager (CASI) data. *Remote Sensing of Environment* 101: 2, 230-248.

Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster, A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J. Lamarque, F. Leverington, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J. ViÃ© & R. Watson (2010). Global biodiversity: Indicators of recent declines. *Science* 328: 5982, 1164-1168.

Carlson, K. M., G. P. Asner, R. F. Hughes, R. Ostertag & R. E. Martin (2007). Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems* 10: 4, 536-549.

CBD 1992, *Convention on Biological Diversity*. Available: http://www.cbd.int/convention/text/ [2014, 17.11.].

Clark, M.L. 2012, "Identification of canopy species in tropical forests using hyperspectral data" in *Hyperspectral remote sensing of vegetatio*n, eds. P.S. Thenkabail, J.G. Lyon & A. Huete, Taylor & Francis Group, Boca Raton, pp. 423-445.

Clark, M. L. & D. A. Roberts (2012). Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensin*g 4: 6, 1820-1855.

Clark, M. L., D. A. Roberts & D. B. Clark (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environmen*t 96: 3–4, 375-398.

Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensin*g 21: 10, 2075-2087.

Duffy, J. E. (2009). Why biodiversity is important to the functioning of real-world ecosystems. *Frontiers in Ecology and the Environmen*t 7: 8, 437-444.

Féret, J. -. & G. P. Asner (2012). Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and LiDAR. *Remote Sensin*g 4:, 2457-2467.

Féret, J. & G. P. Asner (2014). Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Application*s 24: 6, 1289-1296.

Féret, J. -. & G. P. Asner (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. *Geoscience and Remote Sensing, IEEE Transactions o*n 51: 1, 73-84.

Fisher, R. A., A. S. Corbet & C. B. Williams (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12: 1, 42-58.

Gardner, T. A., N. D. Burgess, N. Aguilar-Amuchastegui, J. Barlow, E. Berenguer, T. Clements, F. Danielsen, J. Ferreira, W. Foden, V. Kapos, S. M. Khan, A. C. Lees, L. Parry, R. M. Roman-Cuesta, C. B. Schmitt, N. Strange, I. Theilade & I. C. G. Vieira (2012). A framework for integrating biodiversity concerns into national REDD+ programmes. *Biological Conservatio*n 154: 0, 61-71.

Gaston, K.J. & T.M. Blackburn (2000). *Pattern and process in macroecology.* pp. 377. Blackwell Science Ltd., Malden.

Gaston, K. J. (2000). Global patterns in biodiversity. *Natur*e 405: 6783, 220-227.

Ghosh, A., F. E. Fassnacht, P. K. Joshi & B. Koch (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformatio*n 26: 0, 49-63.

Gillespie, T. W., G. M. Foody, D. Rocchini, A. P. Giorgi & S. Saatchi (2008). Measuring and modelling biodiversity from space. *Progress in Physical Geography* 32: 2, 203-221.

Green, A. A., M. Berman, P. Switzer & M. D. Craig (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions o*n 26: 1, 65-74.

Hector, A. & R. Bagchi (2007). Biodiversity and ecosystem multifunctionality. *Natur*e 448: 7150, 188-190.

Higgins, M. A., G. P. Asner, R. E. Martin, D. E. Knapp, C. Anderson, T. Kennedy-Bowdoin, R. Saenz, A. Aguilar & S. Joseph Wright (2014). Linking imaging spectroscopy and LiDAR with floristic composition and forest structure in Panama. *Remote Sensing of Environmen*t 154: 0, 358-367.

Howard, P. C., P. Viskanic, T. R. B. Davenport, F. W. Kigenyi, M. Baltzer, C. J. Dickinson, J. S. Lwanga, R. A. Matthews & A. Balmford (1998). Complementarity and the use of indicator groups for reserve selection in Uganda. *Natur*e 394: 6692, 472-475.

Imai, N., A. Tanaka, H. Samejima, J. B. Sugau, J. T. Pereira, J. Titin, Y. Kurniawan & K. Kitayama (2014). Tree community composition as an indicator in biodiversity monitoring of REDD+. *Forest Ecology and Managemen*t 313: 0, 169-179.

Jones, H.G. & R.A. Vaughan (2010). *Remote sensing of vegetation - Principles, techniques, and applications.* 1. p. 384 s. Oxford University Press, Oxford.

Kaufman, L. & P.J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* pp. 342. John Wiley & Sons, Inc., New Jersey.

Krebs, C.J. (2014). *Ecological Methodology.* 3. ed. In preparation. Available: http://www.zoology.ubc.ca/~krebs/books.html [2014, 10.10.].

Kuenzer, C., M. Ottinger, M. Wegmann, H. Guo, C. Wang, J. Zhang, S. Dech & M. Wikelski (2014). Earth observation satellite sensors for biodiversity monitoring: Potentials and bottlenecks. *International Journal of Remote Sensin*g 35: 18, 6599-6647.

Lee, J. B., A. S. Woodyatt & M. Berman (1990). Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *Geoscience and Remote Sensing, IEEE Transactions o*n 28: 3, 295-304.

Leutner, B. F., B. Reineking, J. Müller, M. Bachmann, C. Beierkuhnlein, S. Dech & M. Wegmann (2012). Modelling forest alpha diversity and floristic composition - on the added value of LiDAR plus hyperspectral remote sensing. *Remote Sensin*g 4: 9, 2818-2845.

Lucas, K. L. & G. A. Carter (2008). The use of hyperspectral remote sensing to assess vascular plant species richness on Horn Island, Mississippi. *Remote Sensing of Environmen*t 112: 10, 3908-3915.

Lucas, R., P. Bunting, M. Paterson & L. Chisholm (2008). Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sensing of Environmen*t 112: 5, 2088-2103.

Maeda, E. E., J. Heiskanen, K. W. Thijs & P. K. E. Pellikka (2014). Season-dependence of remote sensing indicators of tree species diversity. *Remote Sensing Letter*s 5: 5, 404-412.

Mbuthia, K.W. 2003, *Ecological and ethnobotanical analyses for forest restoration in the Taita Hills, Keny*a, Department of Botany, Miami University, Oxford, Ohio.

Medina, O., V. Manian & J. D. Chinea (2013). Biodiversity assessment using hierarchical agglomerative clustering and spectral unmixing over hyperspectral images. *Sensor*s 13: 10, 13949-13959.

Milliken, W., D. Zappi, D. Sasaki, M. Hopkins & R. T. Pennington (2010). Amazon vegetation: How much don't we know and how much does it matter*? Kew Bulleti*n 65: 4, 691-709.

Mirkin, B. (2011). Choosing the number of clusters. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discover*y 1: 3, 252-260.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, da Fonseca, Gustavo A. B. & J. Kent (2000). Biodiversity hotspots for conservation priorities. *Natur*e 403: 6772, 853-858.

Nagendra, H., R. Lucas, J. P. Honrado, R. H. G. Jongman, C. Tarantino, M. Adamo & P. Mairota (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicator*s 33: 0, 45-59.

Oldeland, J., D. Wesuls, D. Rocchini, M. Schmidt & N. Jurgens (2010). Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity*? Ecological Indicator*s 10: 2, 390-396.

Omoro, L. A., P. E. Pellikka & P. Rogers (2010). Tree species diversity, richness, and similarity between exotic and indigenous forests in the cloud forests of eastern arc mountains, Taita Hills, Kenya. *Journal of Forestry Researc*h 21: 3, 255-264.

Palmer, M.W., Wohlgemuth, T., Earls, P., Arévalo, J.R. & Thompson, S.D. 2000, "Opportunities for long-term ecological research at the Tallgrass Prairie Preserve, Oklahoma", Pr*oceedings of the ILTER Regional Workshop: Cooperation in Long Term Ecological Research in Central and Eastern Europe*, 22–29 June 1999, pp. 123–128.

Palmer, M. W., P. G. Earls, B. W. Hoagland, P. S. White & T. Wohlgemuth (2002). Quantitative tools for perfecting species lists. *Environmetrics* 13: 2, 121-137.

Pellikka, P.K.E., Clark, B.J.F., Gosa, A.G., Himberg, N., Hurskainen, P., Maeda, E., Mwang'ombe, J., Omoro, L.M.A. & Siljander, M. 2013, "Agricultural expansion and its consequences in the Taita Hills, Kenya" in *Developments in Earth Surface Processes, Vol. 16. Kenya: a natural outlook*, eds. P. Paron, D.O. Olago & C.T. Omuto, 1st edn, Elsevier, Amsterdam, pp. 165-179.

Pellikka, P. K. E., M. Lötjönen, M. Siljander & L. Lens (2009). Airborne remote sensing of spatiotemporal change (1955–2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. *International Journal of Applied Earth Observation and Geoinformation* 11: 4, 221-232.

Pettorelli, N., K. Safi & W. Turner (2014). Satellite remote sensing, biodiversity research and conservation of the future. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 1643.

Phelps, J., D. A. Friess & E. L. Webb (2012). Win–win REDD+ approaches belie carbon–biodiversity trade-offs. *Biological Conservation* 154: 0, 53-60.

Pressey, R. L., C. J. Humphries, C. R. Margules, R. I. Vane-Wright & P. H. Williams (1993). Beyond opportunism: Key principles for systematic reserve selection. *Trends in Ecology & Evolution* 8: 4, 124-128.

Richter, R. & Schläpfer, D. 2011, *ATCOR4 user guide, version 6.0.2*, ReSe Applications, Wil, Switzerland.

Rocchini, D., N. Balkenhol, G. A. Carter, G. M. Foody, T. W. Gillespie, K. S. He, S. Kark, N. Levin, K. Lucas, M. Luoto, H. Nagendra, J. Oldeland, C. Ricotta, J. Southworth & M. Neteler (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics* 5: 5, 318-329.

Rocchini, D., C. Ricotta & A. Chiarucci (2007). Using satellite imagery to assess plant species richness: The role of multispectral systems. *Applied Vegetation Science* 10: 3, 325-331.

Schaepman, M. E., S. L. Ustin, A. J. Plaza, T. H. Painter, J. Verrelst & S. Liang (2009). Earth system science related imaging spectroscopy—An assessment. *Remote Sensing of Environment* 113, Supplement 1: 0, S123-S137.

Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel & J. V. Martonchik (2006). Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment* 103: 1, 27-42.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, The* 27: 3, 379-423.

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163:, 688.

Specim Ltd. 2013, *AisaEAGLE datasheet ver2-2013*, Spectral Imaging Ltd., Oulu.

Tan, P., Steinbach, M. & Kumar, V. 2006, "Cluster analysis: basic concepts and algorithms" in *Introduction to data mining*, eds. P. Tan, M. Steinbach & V. Kumar, 1st edn, Addison-Wesley, Instock, pp. 487-568.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling & M. Steininger (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* 18: 6, 306-314.

Vaglio Laurin, G., J. C. Chan, Q. Chen, J. A. Lindsell, D. A. Coomes, L. Guerriero, F. D. Frate, F. Miglietta & R. Valentini (2014). Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS ONE* 9: 6, e97910.