Geo-information Science and Remote Sensing

Thesis Report GIRS-2025

Species Classification in a Diverse Agroforestry System Using UAV-Based Remote Sensing

Assessing Hyperspectral-LiDAR integration for Mapping Tree and Shrub diversity in Food Forest Ketelbroek.

Emma Schalker

26-11-2025

WAGENINGEN
UNIVERSITY & RESEARCH

Foreword

Before the reader continues, I would like to state my utmost thanks and appreciation to my supervisors Kirsten de Beurs and Sander Mucher for their help, feedback and patience for the past year. Furthermore, I want to thank Wouter van Eck for his help with the acquisition of the fieldwork data and the guided tour though Ketelbroek. Finally, I want to thank my friends and family for their support and patience throughout the thesis writing process.

Species Classification in a Diverse Agroforestry System Using UAV-based Remote Sensing

Assessing Hyperspectral-LiDAR Integration for Mapping Tree and Shrub species in Food Forest Ketelbroek.

Emma Schalker

Registration number: 1073400

Supervisors:

Kirsten de Beurs

Sander Mucher

A thesis submitted in partial fulfilment of the degree of Master of Science

At Wageningen University and Research,

The Netherlands.

26-11-2025

Wageningen, The Netherlands

Thesis code number: GRS-82000_2019_0+
Thesis Report: GIRS-2025
Wageningen University and Research

Laboratory of Geo-Information Science and Remote Sensing

Abstract

Agroforestry systems are increasingly recognised for their potential to support climate resilient agriculture, biodiversity and carbon storage. While remote sensing techniques such as UAV imagery, satellite observations and AI based analysis are becoming more widely used in environmental monitoring, the application to food forests is limited. This study addressed this research gap by examining whether UAV-based hyperspectral imagery, combined with a LiDAR-derived Canopy Height Model (CHM), can classify tree and shrub species within the highly diverse Dutch food forest Ketelbroek. This research has answered the question: *To what extent can airborne hyperspectral data identify individual species, and how does a CHM improve this classification?*

Hyperspectral data were acquired over three rasters. Field samples from the 2.4ha food forest provided training and validation samples. After evaluating systematic reflectance differences, within the rasters, species spectral separability (via Jeffries Matusita distance), in six spatial resolutions, species classification was performed using Random Forest (RF) models with and without CHM. Four Principal Components, derived from a Principal Component Analysis (PCA), were used as spectral predictors.

Species separability ranged from 1.0 to 1.4 and improved when the pixel size was coarsened from 7 to 44cm. By simultaneously applying the CHM, the overall accuracy increased (0.68 to 0.82) of the RF models with significance confirmed by McNemar tests (<0.01). Structurally distinct species were consistently better classified (Douglas Fir, European Linden), while species with similar leaf chemistry or strong understory mixing remained challenging (e.g. Sweet Chestnut, Chocolate vine). The CHM most strongly benefited species whose spectral signatures overlapped but differed in height (e.g. Hazel, European Spindle).

The findings of this research show that integrating hyperspectral and structural data can meaningfully improve species mapping in complex forests, but performances are highly dependent on species and environment. Future research should prioritise balanced sampling, multi seasonal data, and spatially aware classifiers.

Table of Contents

1. Introduction

In the Netherlands, over half of the land surface is used for agricultural purposes, covering 2.2 million hectares (CBS, n.d.). In 2021, the total economic value of the Dutch agricultural sector was estimated at around €57 billion, or 6.7% of the gross domestic product (Berkhout et al., 2023). Due to a long history of agriculture and research, the Netherlands produces food in an extremely efficient manner (van Grinsven et al., 2019). However, research indicates that conventional agricultural techniques can result in permanent damage to the soil, water, biodiversity, human health and other ecosystem services. This damage is primarily linked to anthropogenic causes such as soil de-surfacing/ploughing, indiscriminate use of agrochemicals and lack of soil conservation practices, over-extraction of groundwater, and the use of heavy machinery (Almusaed, 2016; Osman, 2014; Tudi et al., 2021). An agricultural approach that prioritises ecosystem services, soil health, and biodiversity is seen as essential for the protecting the future food supply and the current environment (AZ, 2018; Sanders et al., 2015).

In 2024, the agricultural sector contributed around 17% of the total Dutch greenhouse gas emissions (LVVN, 2024). The Dutch Climate Agreement of 2019 states that the agricultural and land use sector must be climate-neutral by 2050. Actions necessary to reach climate neutrality will reduce (cattle) emission, artificial fertilisers, tillage and chemical use, and increase carbon sequestration (EZ&K, 2019).

Agroforestry is a form of agriculture that has the potential to achieve these targets while maintaining or in some cases even increasing the production potential of the land (Breidenbach et al., 2017; Hussain et al., 2021; Wendel et al., 2023). Agroforestry is an umbrella term for range of agricultural practices, where at least one woody species is incorporated to promote other (arable) species (Nair et al., 2021). There are many types of agroforestry - ranging from alley cropping to a highly mixed and layered food forests with over 200 woody species.

While agroforestry is described as a relatively new concept within the contemporary European agricultural sector, it is an ancient practice adapted from the tropics. The term and frameworks are mostly based on the traditions from the tropics, where agroforestry has been an integral part of the livelihoods for centuries (Nair et al., 2021). However, forms of agroforestry have also existed in parts of Europe, such as Spain with the Dehesa system (Joffre et al., 1999). In a four-decade long study, researchers found that mixed agricultural systems strongly contribute to ecosystem services, such assoil fertility, carbon sequestration, biodiversity conservation and food security (Nair et al., 2021). There is a growing recognition and integration of agroforestry into conventional agriculture in the northern hemisphere and is gaining more popularity in as a more prominent technique of farming in North America and Europe. (Dupraz et al., 2018).

In the Netherlands, the interest in agroforestry is also growing, with the establishment of over 100 individual food forests in the past 15 years. Over 500 ha of land in the Netherlands is now farmed following the principles of agroforestry (Ballemans, 2022). To support the shift to sustainable agriculture and to learn more about the effects of agroforestry on the environment, extensive research on its impact on biodiversity and climate is essential (AZ, 2018). However, this type of research is often time-consuming and expensive as it requires manual labour with fieldwork and monitoring. This results in high costs, and in some cases can prevent research from being conducted altogether (Thomas et al., 2019).

Fortunately, the increased accessibility of satellite and UAV data, and Artificial Intelligence (AI), has facilitated conducting large-scale and exact research in this field. These techniques play an

important role in geospatial research, exploration and monitoring. Remote Sensing allows for large-scale precise monitoring and sustainable managing of a forests. Over the past ten years, the amount of research done on UAV remote sensing and forests has increased significantly (Díaz-Delgado & Mücher, 2019). It has become evident that Remote Sensing is an essential tool in forest monitoring and management. (Dainelli et al., 2021).

However, there has been limited scientific research that uses these new techniques to study food forest, environmental impact and social impact on the Dutch environment (Wendel et al., 2023). This was picked up by organisations such as, Wageningen University and Research (WUR), the Dutch Food Forest Network (Netwerk Voedselbosbouw), and the National monitoring program for Food Forests (Nationaal Monitorings-programma Voedselbossen). These organisations are working to fill these research gaps. Understanding agroforestry systems and their potential for biodiversity and ecosystem services, can further facilitate the transition towards more sustainable agricultural practices.

An example of researchers' efforts to explore knowledge on food forests is the development of the integrated innovative biodiversity monitoring (IBM) tool. In 2023, Wageningen Environmental Research (WEnR) established a new research project in collaboration with the Ministry of Agriculture, Fisheries, Food Security and Nature (LVVN), project number KB-36-010-001. This project started tests in both a young and an old food forest to develop a modular scheme for detailed, evidence-based bio- and environmental monitoring of agroforestry ecosystems. Next to modern techniques such as eDNA, camera traps, audio moths, techniques AI-based species recognition and Remote Sensing are explored for the development of the framework of this research (IBM, 2024). The results of this research lay the foundation for a biodiversity inventory.

The old food forest Ketelbroek researched with the IBM is also the first official Dutch food forest (Veluw, 2013). Ketelbroek is located near Groesbeek close to the German border and started operating in 2009 when two entrepreneurs bought a former intensively cultivated corn field (Figure 1). Years later this field transformed into a luscious mixed forest with over 200 woody species (Figures 2 & 3, Appendix 1). Due to the forest's age, and the time it has had to develop, various plant and animal species have settled there. It makes it an excellent research location for observing measurable interactions and changes over time.

Figure 1: Aerial view of Ketelbroek in 2009, when the site was still a maize field in Groesbeek (NL). This image documents the start of the area before the agroforestry conversion.

Figure 2: Early-stage food forest structure in Ketelbroek (2015), six years after planting. This image shows the first visible stratification and species, illustrating the transition from monoculture to permaculture.

Figure 3: Ketelbroek food forest in 2023, showing a mature, structured agroforestry system with > 200 species. The image highlights the dense canopy and structural diversity developed in over 14 years. (Source: WENR)

2. Research questions and challenges

The aim of this MSc study is to evaluate how accurate airborne hyperspectral imagery, combined with a LiDAR-derived canopy height model, can classify/identify the individual (woody) plant species in the highly diverse Dutch food forest *Ketelbroek*, using Random Forest classifiers. To address this aim, the following central research question is posed:

In a 2.4 ha food forest containing over 200 woody species, to what extent can species be classified using airborne hyperspectral data, and what is the added value of a Canopy Height Model in the classification process?

To answer this question, this research is structured into three components.

1. Inter-strip variation and linear regression
    a. How much do overlapping areas between adjacent flightlines differ in reflectance values, and is there a consistent trend between overlapping pairs?
    b. Can linear regression reduce between-strip spectral differences?
2. Spectral separability of species
    a. What is the effect of the different spatial resolutions on species discrimination?
    b. Which species and species pairs are consistently confused, and which are well-separated spectrally?
3. Random forest classification and CHM integration.
        a. What are the overall accuracies (OA), and how does the inclusion of a CHM influence classification performance?
3. To what extent does the CHM improve classification accuracy at the species level (Precision, Recall, F1 mean)

This study explores to what extent UAV-based hyperspectral imagery (Nano Hyperspec with 270 spectral bands), combined with a Canopy Height Model (CHM) and a Random Forest classifier, can be used to identify trees at the species level. This research uses data from a 2.4-hectare Dutch food forest (*Ketelbroek*) containing approximately 200 plant species. Both the research approach and the field of research present multiple challenges that must be considered.

One important challenge is scale-dependent spectral variability (Fassnacht et al., 2016). Hyperspectral data with very high spatial resolution can increase within-crown spectral heterogeneity, adding considerable complexity to species classification. In contrast, coarser resolutions may lead to spectral mixing between adjacent species making it difficult to isolate individual spectral signatures. Therefore, determining the optimal spatial resolution is context-dependent and requires a study specific calibration. Furthermore, spectral noise from understory vegetation, soil, shadows, and overlapping crowns is strongly influenced by spatial resolution. At a very high spatial resolution, these background signals become more prominent, requiring object-based classification methods or spectral unmixing techniques to mitigate their impact.

Secondly, the high dimensionality of spectral data and the associated need for effective feature selection present another key challenge. When studying hundreds of bands, the dataset becomes high-dimensional, leading to increased computational demands and a higher risk of overfitting. A phenomenon known as the “Curse of dimensionality” (Taşkın et al., 2017). Dimensionality reduction techniques are therefore essential to simplify the data while retaining relevant information and variation. In addition, the large number of unique species in the study area increases the likelihood of overlapping spectral signatures, which can cause confusion during

classification. This makes the ratio between sample size and number of species an important factor in model performance.

Finally, spatial and sample size bias poses a challenge due to over- and under-representation of certain areas or species. In the case of Ketelbroek, over 65% of the species have fewer than ten samples, which limits their utility for robust training and testing. An adequate number of representative samples is essential to ensure reliable model validation and accurate classification performance (Foody, 2009; Song & Wang, 2023). Moreover, when species are not evenly distributed across space, their spectral signatures may become spatially clustered or diluted, increasing the risk of class averaging and the loss of unique spectral characteristics (He et al., 2018).

3. Data and methods

3.1 Study area

This research was conducted using data gathered in the Ketelbroek food forest (Figure 4), located near Groesbeek in the Netherlands (51°46'08.8"N 5°58'01.8"E). The study area spans approximately 2.4 ha and has been managed as a multi-layered agroforestry system that started in 2009. Prior to its conversion, the land was used for intensive agriculture, primarily maize cultivation.

The vegetation structure within Ketelbroek is highly diverse. The canopy layer reaches heights over 15 metres, while midstory shrubs range from 1 to 5 metres. Groundcover vegetation typically remains around 30 centimetres. More than 200 tree and shrub species have been recorded within the forest, of which roughly two-thirds are fruit-bearing species. The remaining species function as support species.

Figure 4: High resolution oblique image acquired with a DJI Zenmuse P1 on 31 May 2023. This image shows a landscape overview of canopy structure, local hydrology and the surrounding flat terrain. (made by Stan Los)

Topographically, the terrain is flat, with an average elevation around 13m above sea level. The site lies in a basin formed by the push moraine of Nijmegen and Groesbeek. The subsoil consists of sandy loam, characteristic of the Pleistocene fluvial terraces found in this part of Gelderland (Bodemdata, n.d.).

Hydrologically, a small permanent stream flows through the forest, supporting a resident population of Eurasian beavers (*Castor Fiber*). Their activities, such as the construction of small dams and foraging, have created localized clearings and water management features that contribute to heterogeneity in canopy cover and ground conditions.

The climate in Groesbeek is classified as temperate maritime (Cfb). Long term meteorological data indicate a mean annual temperature of 10.8 °C and mean annual precipitation of 832 mm (Klimaat Gelderland, n.d.).

3.2 Data acquisition

The data used in this research consists of hyperspectral imagery from airborne source and field reference data. An overview of the metadata is included in Appendix 1.

3.2.1 UAV- Borne Remote Sensing

Table 1: Technical overview of hyperspectral and LiDAR acquisition parameters collected on 31 May 2023 in Ketelbroek. Includes platform details, spectral ranges, resolutions and sensor specifications.

PARAMETERS	**HYPERSPECTRAL**	**LiDAR**
UAV platform	DJI Matrice 300 RTK	RIEGL VUX-SYS
Trajectory	Applanix APX-15 GNSS/IMU	
Flight altitude	80m	Max 120m
FOV V/H	30° / 360°	
CRS	RD NEW / EPSG:28992	
Sensors	Headwall Nano Hyperspec VNIR pushbroom spectrometer	Velodyne VLP-16 “Puck” LiDAR
Acquisition time	31 May 2023	31 May 2023
(Band) Range	400nm – 1000nm	-
Spectral resolution	2.2nm	-
Spatial resolution	0.074 m	0.1 m (after processing)
Spatial bands	640	-
Spectral bands	268	-

The hyperspectral imagery (Figure 5) was taken along 15 parallel flightlines (Figure 6) by a drone. The flightlines are North-East / South-West orientated with each adjacent flightline having a width of approximately 20m. Each direction was flown once. The hyperspectral images from the flightlines have an overlap of approximately 5 - 10m on either side. The imagery survey took place on a clear-sky day within 2 hours of the solar noon.

The raw and hyperspectral cubes of the imagery survey were processed in R (v 2024.12.0) as follows:

1. Radiometric calibration using a 3 x 3 m field reflectance panel
2. Spectral smoothing via a first-order Savitzky-Golay filter (window = 11 band)
3. Spatial down-sampling by a factor of 2
4. Reprojection to Amersfoort / RD New (EPSG:28992) using bilinear resampling – yielding a final GSD of 0.073995 m for both hyperspectral as RGB rasters.

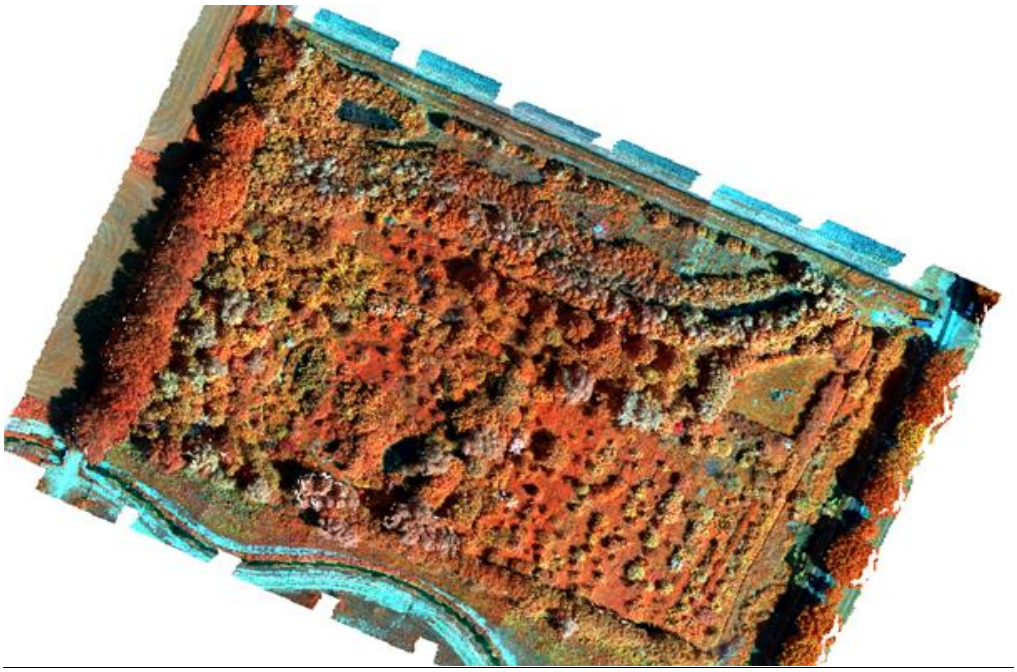


Figure 5: Hyperspectral UAV image of Ketelbroek collected on 31 May 2023 using a Nano Hyperspec sensor (5 cm GSD). This dataset forms the spectral basis for all subsequent analysis. (Source: WENR)

Figure 6: Overview of the 15 hyperspectral flightlines (NE – SW orientation) flown on 31 May 2023. The diagram illustrates overlap zones that are critical for evaluating radiometric differences.

For the LiDAR campaigns the LiDAR RiCopter is used, with a Reigl VUX-SYS light weight campact laser scanner of 3.5kg. Up to 350.000 measurements per second are taken. To enhance the detection of the canopy height, the full vegetation is measured on 31 May 2023 and terrain on 6 December 2023.

LiDAR point clouds were processed in ArcGIS Pro 3.4.2 and used to generate a Digital Terrain Model (DTM) and Digital Surface Model (DSM). The Canopy Height Model (CHM) was computed as CHM = DSM – DTM, produced at a pixel size of 0.10 m (Figure 7). For further analysis, the CHM was resampled to match the spatial resolution of the hyperspectral rasters.

Figure 7: Digital Surface Model (Left) and Digital Terrain Model (right) derived from LiDAR scans (May and December 2023). The subtraction forms the Canopy Height Model (CHM) used as structural input in the Random Forest classifier (Source: WENR)

3.2.2 Resolution

To evaluate scale effects on classification accuracy, rasters were aggregated by taking the mean of the aggregated pixels. In this report, I aggregated the data from 1 to 6, as you can see in table 2 below. In this research, when aggregation 6 is mentioned, I reference to a resolution of 0.444m.

Table 2: Spatial resolutions corresponding to six aggregation levels used in the study, ranging from 7 cm to 44cm. Showing how pixel size increases through aggregation.

Aggregation	Resolution (m)
1	0.074
2	0.148
3	0.221
4	0.300
5	0.370
6	0.444

3.2.3 Field reference data

A desktop analysis of the RGB orthomosaic was performed to label crowns in areas inaccessible on foot. These labels were transformed to a point dataset. In addition, I conducted two fieldwork days to gather information on the ground:

- 28 May 2024 (leaf-on): accessible spring vegetation.
- 29 January 2025 (leaf-off): improved understorey access.

During each visit, the owner and a research team recorded species identity and precise locations using RTK-GNSS coordinates (horizontal accuracy of ± 0.03 m) for individual trees and shrubs. Where possible, the sampling strategy aimed for maximum taxonomic spread and replication. All sample points are merged into one dataset. In appendix 2 (figure 20), the distribution of samples is illustrated. Before cleaning, the point datasets consisted of 1003 samples.

3.3 Data preparation

This section describes how the raw remote-sensing and field data were ingested, cleaned, and assembled into a pixel-level dataset used for classification model testing. A photographic description of the species that are included in the final dataset is included in Appendix 3 (Table 16).

3.3.1 Data cleaning and attribute augmentation

For this study, only flightlines 1-3 were used. Other flightlines were either in shadow or did not have sufficient tree samples for the analysis. Pixel-level reflectance values were extracted at each point from the three rasters and six aggregation levels and subsequently merged into the attribute table. Any values larger than 10000 DN were treated as measurement errors and therefore removed. Records with NA height values were also removed. Canopy height at each point was retrieved from the CHM raster and stored as height. Each species was labelled as: "Tree", "Shrub", or "Vine". Trees and a shrubs were separated using a three meter threshold (Allaby, 2012). After cleaning, 798 points remained in the dataset for this research, located in flightlines 1-3 (table 3).

Table 3: Overview of sample count, unique species, and species with ≥ 10 samples across rasters 1 – 3. These thresholds determine which species are used for classification.

Raster	**N total**	**N species**	**N species > 10**	**Vine/shrub/tree**
1	234	17	10	23 / 7 / 204
2	306	29	13	38 / 10 / 268
3	258	43	8	0 / 132 / 126

3.4 Spectral separability

3.4.1 Raster calculation

To identify inter-raster variability, the pixel-wise difference across each adjacent raster pair is computed. Raster 1 vs 2 → Pair 1, and Raster 2 vs 3 → Pair 2. For each pair the overlapping extent is determined with an intersect, and every second raster is resampled to the grid of the first raster. Within each band, the reflectance values are averaged, and the signed and relative difference is calculated (1 and 2).

(1) $$\Delta R = R_i - R_j^{resampled}(x, y)$$

(2) $$\Delta R_{rel} = \frac{\Delta R}{R_i} * 100$$

For this part, the spectral range is split into five groups to measure if the raster difference is the same over the wavelengths. Blue, Green and Red (in the visible spectrum) capture differences that can be seen with the eye, pigment related differences like chlorophyll in the leaves and carotenoids in flowers and fruits. The Red Edge bands can pick up even more detailed differences in chlorophyll content and gives a sharp reflectance increase on vegetation. These bands are generally seen as the most important for species discrimination. The NIR bands are mostly influenced by leaf and canopy structure and shows a strong plant/soil contrast (Hennessy et al., 2020).

- Blue: 400–500 nm
- Green: 500–590 nm
- Red: 590–680 nm
- Red Edge: 680–740 nm
- NIR: 740–1000 nm

3.4.2 Radiometric harmonization

To test if a per-band linear correction can remove a bias, a random sampling of 1000 pixels from pair 1 and pair 2 is taken. From this, a simple linear regression model (3) is fitted and plotted:

(3) $$\hat{y} = b_0 + b_1 x$$

The regression, intercept, slope, Root Mean Square Error (RMSE) and the coefficient of determination (R^2) is extracted. To visualise the regression line, for each of the five spectral regions, the central wavelength is selected (table 4).

Table 4: Five reference wavelengths selected for band-specific linear regression of reflectance bias.

Region	**Blue**	**Green**	**Red**	**Red Edge**	**NIR**
Wavelength (nm)	435.933	547.453	647.822	723.655	870.862

Based on the magnitude of residual bias after correction—particularly in the Green and NIR bands—and the spatially structured errors linked to canopy height; we assessed whether a per-band linear adjustment provided sufficient benefit. Where residuals remained large or spatially heterogeneous, strips were kept separate rather than merged into a harmonized mosaic.

3.4.3 Dimensionality reduction

To reduce dimensionality, a Principal Component Analysis (PCA) is applied for the Spectral Separability, and the Random Forest Classification (RF). For both methods, a PCA is performed on the full spectrum (~400nm – 1000nm), to reduce the spectral separability. Covariance PCA is used because all bands share the same reflectance units. This preserves the natural variance across wavelengths (Uddin et al., 2021). The first four PCs (explain over 99% of the spectral variance) are used both to compute the spectral separability and as inputs for the RF (with and without CHM). In Appendix 4 (figure 21), the cumulative eigenvalues and loadings are visualized.

3.4.4 Spectral separability function

To calculate the spectral separability, there are two common distance methods: the Jeffries Matusita (JM), and the Bhattacharyya distance. Both functions measure the overlap between two probability distributions, based on the class mean. The difference between the JM and the Bhattacharyya distance is JM is a non-linear transformation, and has a normalized range from 0 to 2, whereas the Bhattacharyya distance has no range. It is therefore easier to interpret the JM (Sen et al., 2019). Studies show that a higher number of class pairs with JM separability above 1.0, especially above 1.9, leads to improvement of overall accuracy (OA) (Table 5). Class pairs with JM separability < 1.0 reduce OA (Rehman et al., 2024; Wicaksono & Aryaguna, 2020a).

Table 5: Classification of JM-distance thresholds indicating poor, moderate, and high separability.

JM distance	Separability interpretation	Impact on accuracy
< 1.0	Poor	Low OA
1.0 – 1.9	Moderate	Improved OA
> 1.9	High	High OA

The separability analysis is based on the three cleaned point datasets and associated reflectance values from raster 1 – 3, over all 6 aggregations. Table 3 shows an overview of the samples used. Throughout this research, only species with N > 10 are used.

3.5 Random Forest

Random Forest (RF) is a machine learning method that combines many individual decision trees to perform classification. Each tree is trained on a bootstrap sample of the data, and at every tree split only a random subset of predictor variables is considered. This method reduces overfitting and makes RF well-suited for high dimensional and noisy datasets such as hyperspectral imagery (Belgiu & Drăguţ, 2016; Breiman, 2001).

RF was chosen for this study because this hyperspectral dataset contains 270 correlated bands. As mentioned, RF can identify spectral dimensions without requiring strict assumptions about the data distributions. Secondly, the data contains strong within-class variability and overlapping spectral signatures. Thirdly, the addition of the CHM adds a structural variable that RF can model effectively. Finally, RF performs reliably with relatively small and imbalanced training sets, making it an appropriate method for the limited sample set in this research.

3.5.1 Model setup and training

The RF classifier was implemented separately for each raster. Only species with ≥ 10 samples were included and have been randomly split into 60% training and 40% testing data set. Stratified random sampling was used to consider for class imbalance. The coarsest pixel size (0.44cm, aggregation 6) is used, and a PCA is applied to reduce dimensionality. To assess the influence of the CHM on each Raster, a RF analysis was performed twice; once on the PCA-only set, and once on the PCA + CHM.

For the Random Forest, the following metrics are used:

- Trees 500
- Mtry p/3
- Node size 1
- Importance Permutation based
- Cross validation grid 5-fold

3.5.2 Model evaluation

The model accuracy is assessed with the Overall Accuracy and Kappa. Kappa is a metric that compares Observed Accuracy with Expected Accuracy (random chance). The classification accuracy reported in this study is based on the point-level validation dataset. To determine whether adding the CHM improves the RF model, the McNemar test was used. This analyses whether the overall accuracy of the RF model improved. This statistical test is used in the case of binary outcome. In this case, the outcome was either a correct or incorrect classification. The test then can compare the number of cases where the two models disagree. The paired

prediction on the same dataset allows the statistical test to calculate the significance of the addition of the CHM (Wang et al., 2021).

It was expected that there will be an improvement in the overall accuracy of the model with the inclusion of the CHM as a 5th predictor. The null hypothesis (H_0) is as follows:

There is no significant difference in the classification performance between the model using only principal components (PCA-only), and the model using both principal components and the canopy height model (PCA + CHM). The number of test instances that are correctly classified by one model and misclassified by the other is symmetric.

The alternative hypothesis (H_1) states:

There is a significant difference in the classification performance between the PCA-only model and the PCA + CHM model. The number of test instances for which model makes a correct prediction while the other makes an incorrect prediction is asymmetric.

A significance level of α = 0.10 is applied, corresponding to a 90% confidence level.

3.5.3 Performance Analysis

To analyse the effect of the CHM on species level multiple metrics are used. The recall and precision, also referred to as Producer's and User's accuracy, were calculated from the confusion matrix. This analysis was necessary to quantify how including the CHM affects species-level discrimination and to characterize the types of errors the classifier made. Because species frequencies were imbalanced and several species were spectrally similar in the hyperspectral data, overall accuracy could be misleading. I therefore computed per-species Producer's accuracy (recall), which measures omission error (the proportion of reference instances of a species correctly mapped, (4)). The User's accuracy (precision) measures commission error (the reliability of the mapped level, (5)).

(4) $$Producer's\ accuracy = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

(5) $$User's\ accuracy = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

I reported their harmonic mean (also known as F1, (6)) to summarize the trade-off between missing trees and falsely labelling them. Taken together, these confusion matrix metrics allowed a fair comparison of models with and without CHM and indicated whether CHM increased detectability, improved label reliability, or both.

(6) $$F1 = 2 * \frac{User's * Producer's}{User's + Producer's}$$

4. Results

The results section is organised around the three research questions. First, we test how strongly flightline position affects raw reflectance and whether a simple linear regression can harmonise each adjacent strip (RQ1). Next, we examine how spectral region and spatial aggregation influence species separability (RQ2). Finally, we quantify the added value of the canopy-height model (CHM) in a Random Forest classifier (RQ3).

4.1 Inter-strip radiometric consistency

The first research question investigates the consistency between adjacent UAV hyperspectral flightlines and evaluates whether a linear regression model can harmonize reflectance values across overlapping areas. This is an important step, as systematic differences between rasters may bias subsequent classification or spectral analyses. Therefore, this section first quantifies the difference in reflectance values between the adjacent strips for the spectral regions: Blue, Green, Red, Red Edge, NIR. Both absolute (ΔR) and relative (ΔR_{rel}) differences are calculated (Table 6, Figure 8). Next, I assess whether per-band linear regression can reduce the difference in reflectance values by examining regression metrics such as slope, intercept, R^2 and RMSE across the full spectrum (Figures 9– 12).

4.1.1 Magnitude and pattern of the strip bias

How much do overlapping areas between adjacent flight lines differ in reflectance values, and is there a consistent trend between overlapping pairs?

Figure 8a shows that the ΔR and ΔR_{rel} is negative in every band, indicating that the second strip consistently returns higher reflectance values than the first. The magnitude of this absolute difference grows with wavelength, reaching its maximum in the Red Edge and NIR regions. This is to be expected since the reflectance values are higher in those spectral areas compared to the visible wavelengths (Fassnacht et al., 2016). Figure 8b presents the same comparison on a relative scale. Although bands show a decrease of R_1, relative to R_2, the largest percentage drops (-9% to -11%) occur in the visible red band, while the smallest change (-2% to -3%) is observed in the NIR.

Spatially, the difference is heterogeneous and changes depending on the surroundings: ΔR is larger in areas where there is a higher variety in Canopy Height. Whereas ΔR decreases in areas where the CHM shows little to no height difference (Appendix 5, Table 17).

Table 6: Mean absolute and relative reflectance differences for two overlapping strip pairs across five spectral regions.

	Pair 1		**Pair 2**	
N pixels	*175503*		*179000*	
	Mean difference (DN)	**Relative diff. (%)**	**Mean difference (DN)**	**Relative diff. (%)**
BLUE (400 – 500nm)	-15.59	-7.77	-14.72	-6.52
GREEN (500 – 590nm)	-43.35	-10.25	-44.90	-9.24
RED (590 – 700nm)	-34.52	-11.07	-31.69	-9.11
RED EDGE (700 – 740nm)	-73.55	-6.02	-93.44	-6.57
NIR (740 – 1000nm)	-62.32	-1.59	-153.85	-3.35

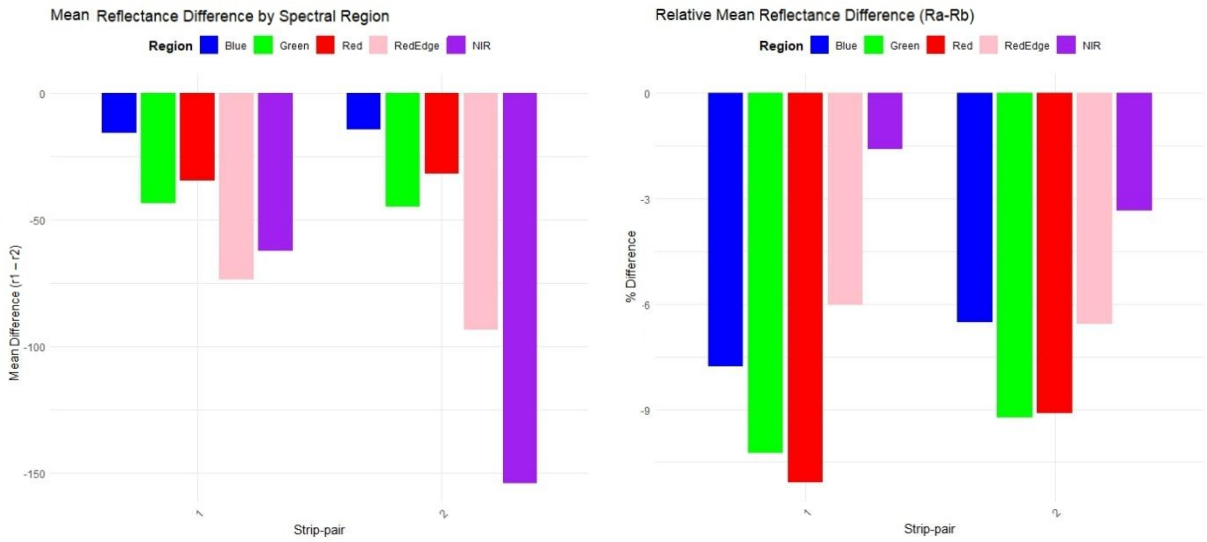


Figure 8: a) Mean absolute difference (left) and b) relative differences (right) between adjacent flightlines across five spectral regions. Higer discrepancies in Red Edge and NIR reveal wavelength dependent strip bias.

4.1.2 Linear harmonisation test

Can radiometric harmonization reduce this between-strip difference?

One thousand random pixels from the overlapping area between raster 1 and 2, and raster 2 and 3, were sampled. For every spectral region a representative wavelength (Table 4) was regressed (ordinary least squares) against its counterpart in the adjacent strip (Figures. 9 & 10). With a regression line (solid – red) and a 1:1 line (dashed – black). Both pairs follow a similar pattern. Pair 1 is visualised in the figures below. The regression plots of pair 2 are shown in Appendix 6 (figures 22 & 23). The regression diagnostics, including Intercept, Slope, R^2 and RMSE of pair 2 are provided in Appendix 7 (figures 24 & 25).

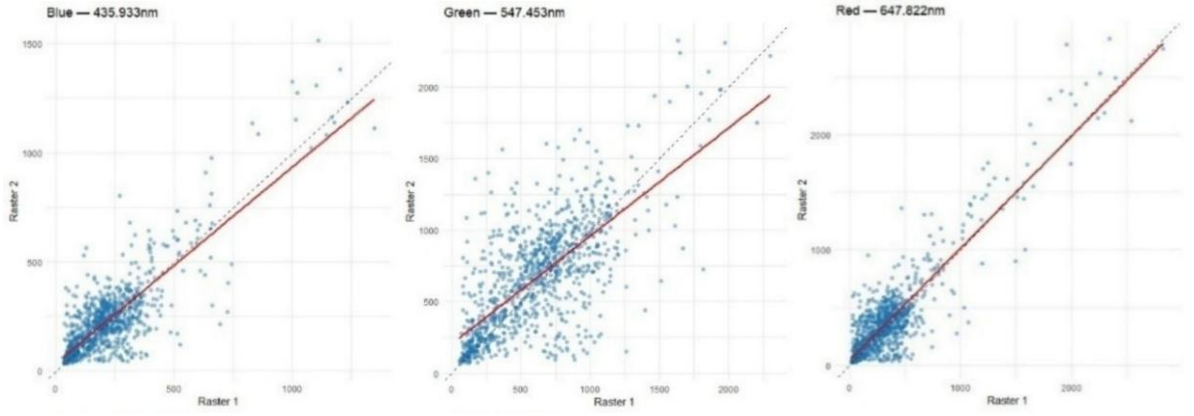


Figure 9: Linear regressions of Raster 1 vs Raster 2 reflectace for the Blue, Green and red bands. Deviations from the 1:1 line indicate inconsistent radiometry between flightlines.

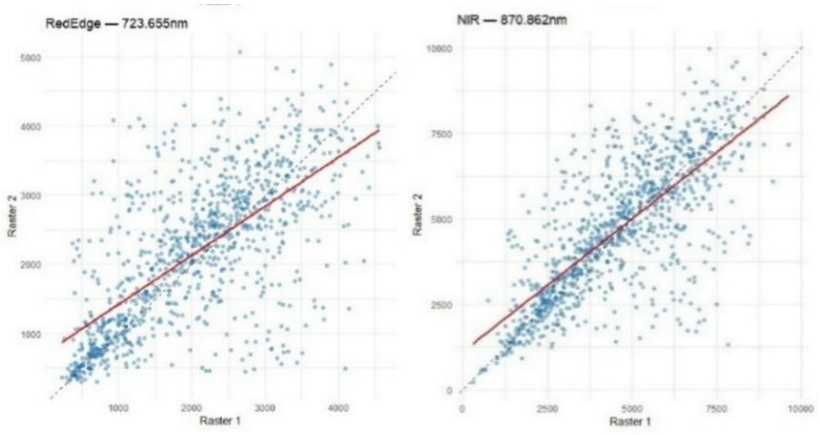


Figure 10: Linear regression of Raster 1 vs Raster 2 for Red Edge and NIR bands. Deviations from the 1:1 line indicate inconsistent radiometry between flightlines.

The linear regression metrics are visualised in Figures 11 and 12 and cover the full spectral range (400 – 1000nm).

The intercept is initially flat (~ 20) across 400 – 500 nm (blue), rises to a modest bump (250) around 540nm, dips almost back to zero near 680 nm, then steeply rises to ~ 980 at 720 nm (Red Edge) and stays above 900 throughout most of the NIR before tapering slightly beyond 950 nm.

The slope stays just around 0.90 in the blue, plunges to 0.75 at ~ 560 nm (green), and rises in until to a slope over 1.0 at ~ 590 nm. Plumets to ~ 0.71 around 730 nm and then climbs gradually to ~ 0.80 by 1000 nm.

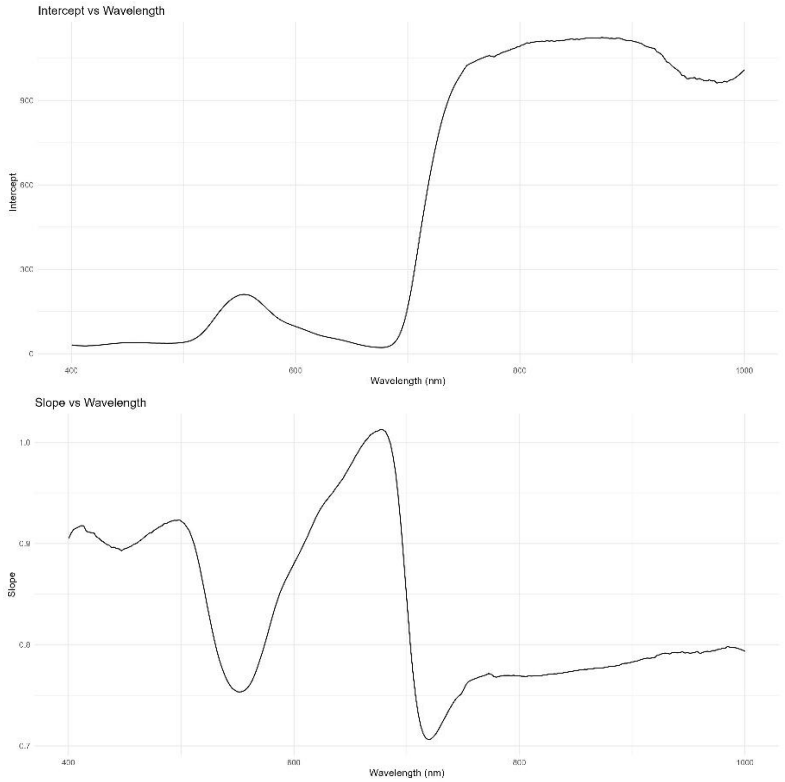


Figure 11: Intercept and Slope parameters of per-band regression across 400 – 1000nm. Sharp changes in the Red Edge and NIR reveal wavelength dependent scaling effects between strips.

The R^2 Shows two quality windows: ~0.72 in the Blue – Green (≈ 500 nm) and a stronger ~ 0.86 in the Red Edge (≈ 690 nm). R^2 slumps to ≈ 0.50 across 600 – 630 nm and again just beyond 720 nm, then levels out at ~0.55 – 0.60 from 800 nm to 1000 nm.

The RMSE Mirrors the intercept: ~ 350 – 400 through most of the visible, a shoulder of 600 near 540 nm, a local minimum (~ 300) at 700 nm, followed by a step-change to > 1100 from 760 nm onwards where errors plateau across the NIR.

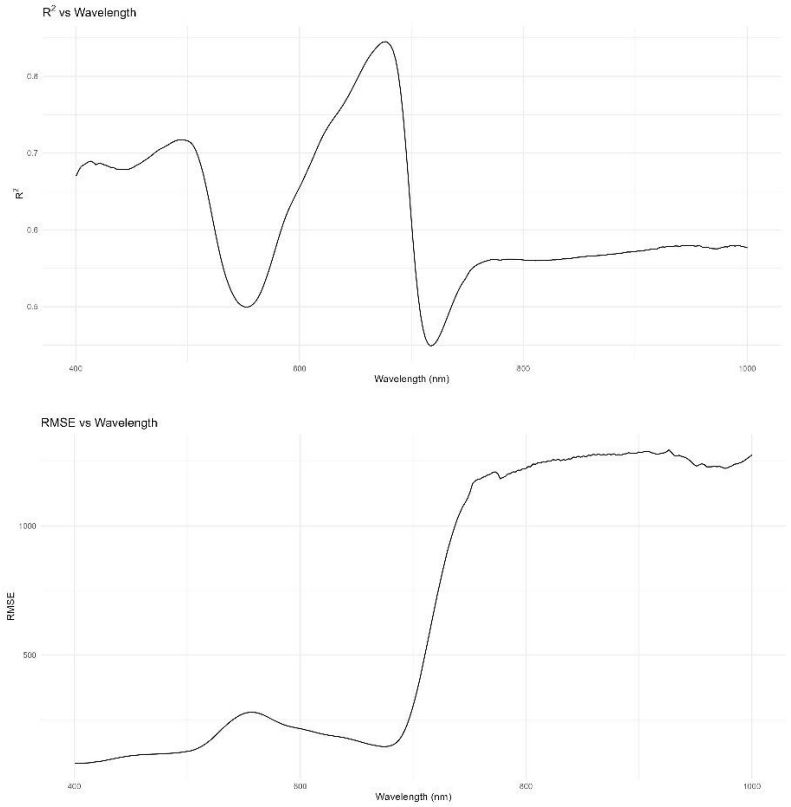


Figure 12: R^2 and RMSE parameters of per band regression across 400 – 1000nm. Goodness of fit metric emphasises inconsistent regression performance and large NIR errors that justify keeping flightlines separate.

The intercept rise sharply in the Red Edge and remains high throughout the NIR. This vertical shift in reflectance between flightlines could be partially corrected by a linear regression. However, the slope values deviate from 1 which indicates wavelength dependent scaling inconsistencies. The slope drops below 0.8 and remains below 0.8 in the NIR. This implies that reflectance differences are not uniform and applying a linear correction could distort real spectral variation (Villaescusa-Nadal et al., 2019).

Furthermore, the R^2 shows inconsistent fit quality. The strong agreement in the Red Edge contrasts with the poor R^2 in the Green and the NIR and limits the confidence in regression-based harmonisation across the full spectral range. This is confirmed by the RMSE values, which exceed 1100 in the NIR, indicating large structure/driven residuals even after correction (Ariza et al., 2018).

While a linear regression model can reduce systematic bias, it does not correct for multiple differences. Also, linear regression cannot resolve the non-linear, spatially structured residuals

that is visualised by the low R^2 and high RMSE. For these reasons, and to avoid overfitting, I chose to keep the flightlines separated throughout this research rather than merge them into a radiometric harmonized mosaic (Kizel & Vidro, 2023).

4.2 Spectral Separability

The JM distance per class shows the distance values to cluster around 1.0 – 1.4, suggesting moderate separability among the species that have the highest values and a poor separability for all species below 1.0 (Table 8 & 9). The moderate separability suggests that there may be a significant distinction between mean spectral signatures. However, enough overlap remains in signatures that misclassifications can occur. Below (figure 13), the spectral signatures are visualised. This shows a substantial within-class variance and may explain why there are no species with a high spectral separability.

4.2.1 Aggregation contribution

Figure 13 summarises the mean Jeffries – Matusita (JM) distance for all "frequent samples" (≥ 10 samples) across three rasters and six aggregation levels. From this the separability increases with a higher aggregation (lower resolution). Around aggregation 5 (spatial resolution: 37cm), the mean JM distance seems to flatten (Chen et al., 2023). Coarser pixels help with the spectral separability, but up to a certain point. Therefore, we use a resolution of 44cm (aggregation 6) for the rest of the study. As the pixel size increases, the difference between the rasters decreases.

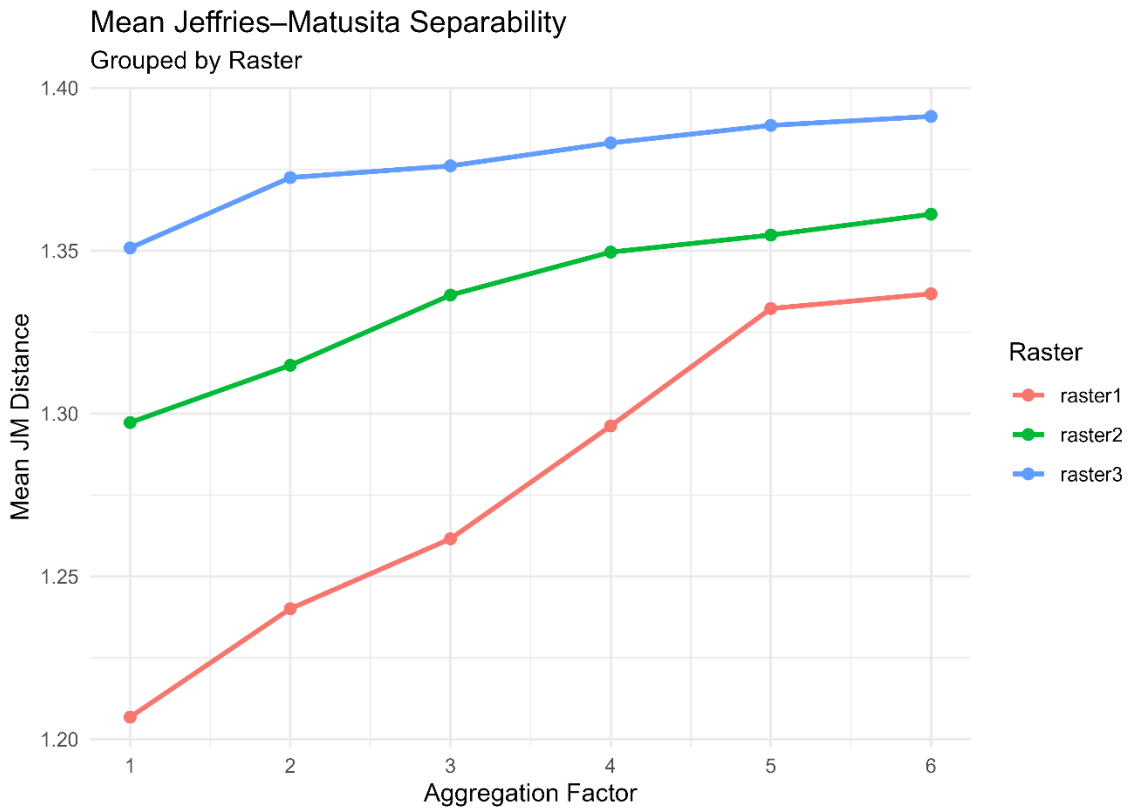


Figure 13: Mean Jeffries Matusita separability aggregation level for the three rasters. Separability increases as resolution becomes coarser, plateauing around aggregation 5 – 6.

4.2.2 Species level

In this section, the separability scores per species are analysed. Table 7 shows the number of samples in each dataset, the number of unique species (with at least 10 samples), and the number of species pairs that can be made for the spectral separability matrix.

Table 7: Overview of: number of samples, unique species, and resulting species pairs for spectral separability calculations for each raster.

Raster	N samples	N Species	N Pairs
1	202	10	45
2	258	13	78
3	159	8	28

In the figures below (fig 14, 15, 16), the mean JM distance per species is plotted over the six aggregations. Since the JM distance is based on a species pair, the mean is used to express the overall separability for each species. The mean is based on all pairs that is made per species and averaged. With the increase in resolution, the spectral separability increases. This is true for all three rasters. The mean JM distance stagnates around aggregation 5. However, since in most of the cases aggregation 6 results in the highest mean JM distance, aggregation 6 is seen as the most successful resolution. Interestingly in raster 3, Autumn Olive, Bollwiller Pear and Trifoliate Orange do not seem to be affected by the increase in aggregation.

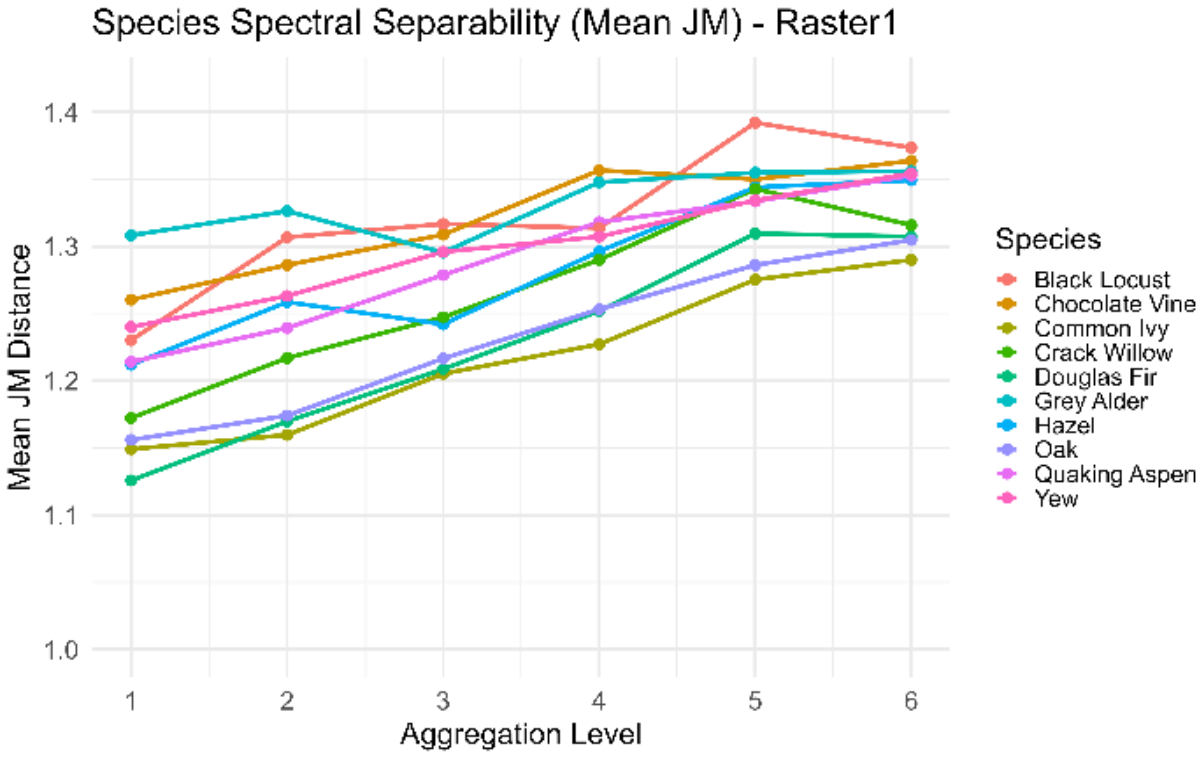


Figure 14: Species level mean JM distances for Raster 1, showing how aggregation reduces spectral noise and increases pairwise separability. Most species benefit form coarser resolution, with separability stabilising near 44cm.

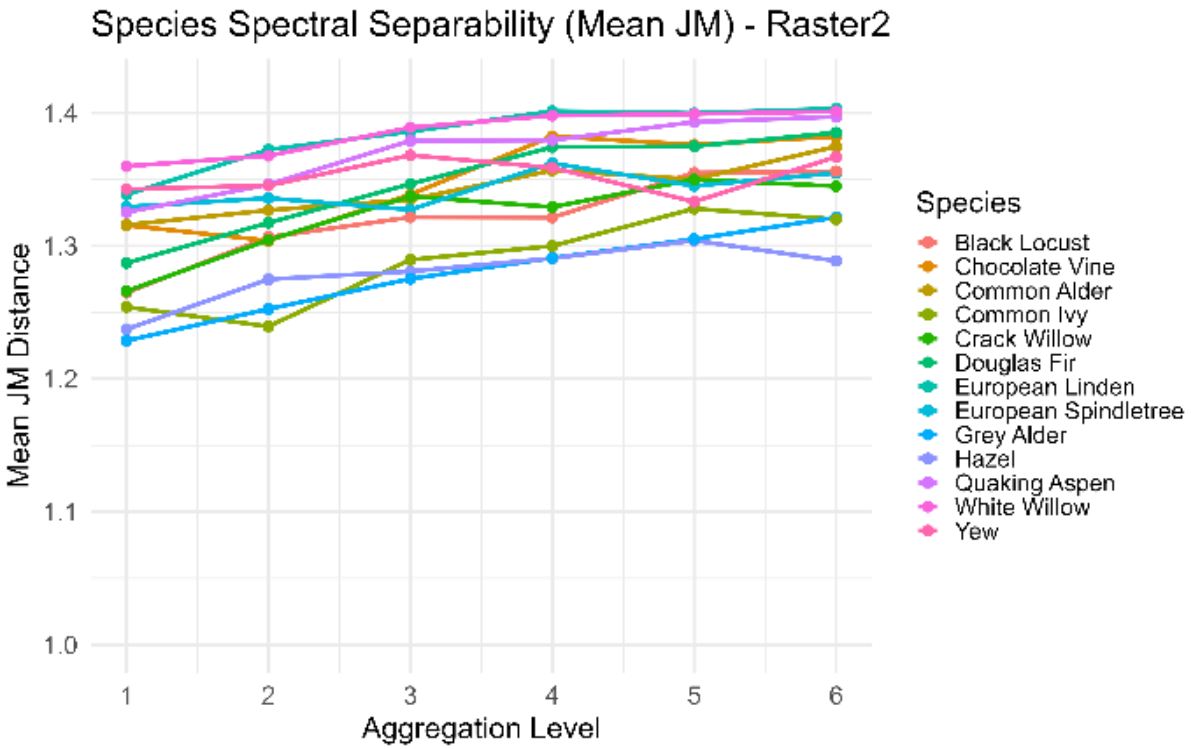


Figure 15: Species level mean JM distances for Raster 2. Trends show a slow but consistent improvement with aggregation, though species differ in response.

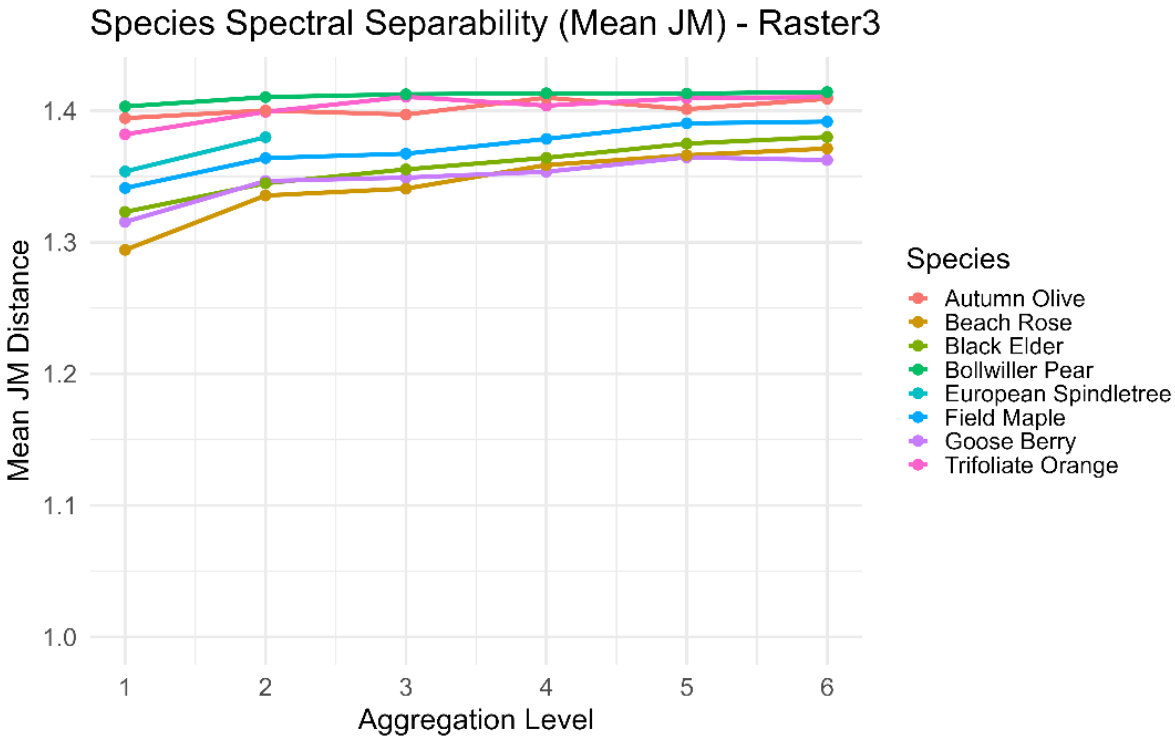


Figure 16: Species level mean JM distances for Raster 3. Showing generally a slight increase in JM values,

In the case of aggregation 6, table 8 highlight the species that have the highest and lowest mean JM distance. The difference between the highest and lowest performing species based on the mean is relatively small. By taking the average of all pairs per species, the gap gets very small. There are no species that are very distinctively having a low spectral separability.

Table 9, on the other hand, shows that when looking at the pairs specifically, there are differences that are clearer. This table is also based on aggregation 6. Douglas fir, hazel, and Gooseberry are mentioned most frequently in the lowest JM distance category for raster 1, 2 and 3 respectively. These values are all around 1. The species pair with the highest JM distance do not have species that are frequently mentioned, besides the Bollwiller pear in raster 3.

Table 8: Species with the highest (Left) and lowest (right) average JM separability at aggregation 6. Highlighting subtle separability differences within each raster.

Raster	Species	MEAN JM		Species	MEAN JM
1	Black Locust	1.374		Common Ivy	1.290
	Chocolate Vine	1.364		Oak	1.305
	Grey Alder	1.356		Douglas Fir	1.307
2	European Linden	1.403		Hazel	1.289
	White Willow	1.401		Common Ivy	1.320
	Quacking Aspen	1.397		Grey Alder	1.321
3	Bollwiller Pear	1.414		Gooseberry	1.362
	Trifoliate Orange	1.410		Beach Rose	1.371
	Autumn Olive	1.409		Black Elder	1.380

Table 9: Pairwise JM distances showing the highest (left) and lowest (right) JM separability at aggregation 6. These extremes indicate which pairs are most and least spectrally distinct.

Raster	Pair	JM Distance		Pair	JM Distance
1	Black Locust – Quaking Aspen	1.411		Common Ivy – Douglas Fir	1.133
	Common Ivy – Yew	1.409		Common Ivy – Quacking Aspen	1.171
	Grey Alder – Quacking Aspen	1.408		Crack Willow – Douglas Fir	1.221
2	Black Locust – Chocolate Vine	1,414		Grey Alder – Hazel	1.033
	Douglas Fir – European Linden	1,414		Crack Willow – Hazel	1.039
	Chocolate Vine – Crack Willow	1,414		Common Ivy – Hazel	1.188
3	Autumn Olive – Bollwiller Pear	1,414		Black Elder – Gooseberry	1.285
	Bollwiller Pear – Trifoliate Orange	1,414		Beach Rose – Gooseberry	1.295
	Black Elder – Bollwiller Pear	1,414		Field Maple – Gooseberry	1.358

4.3 Random Forest classifier with Canopy Height Model

In the following section, a Random Forest (RF) classifier is applied to the three raster datasets. Per raster, two RF classifiers are applied. One where the data is transformed with a Principal Component Analysis (PCA), and one with a PCA and the addition of a Canopy Height Model (CHM) as an extra predictor. Four Principal Components (PCs) are used for the RF. The overall accuracy is calculated and the difference between the two models are analysed. The significance is expressed with a McNemar test. In the second part, the influence of the CHM is analysed per species. The coarsest resolution of 44cm (aggregation 6) is used.

4.3.1 Overall performance

The overall accuracy explains what percentage of the labels our model got correct. Table 10 shows overall accuracy from the two models and the difference between the models. For all three rasters, the overall accuracy and kappa increases. Accuracy trends across all six aggregation levels for each raster are provided in Appendix 8 (Table 20 and Figure 26).

Table 10: Overall accuracy and the corresponding Kappa values for two RF model variants (PCA – only and PCA + CHM) per raster. CHM improves accuracy in all cases, most strongly in Raster 3.

Raster	**PCA - only**		**PCA + CHM**		**ΔOA**	**ΔOA**
	OA	**Kappa**	**OA**	**Kappa**	**Absolute**	**Relative**
1	0.69	0.64	0.70	0.66	+ 0.01	+ 1.4%
2	0.60	0.56	0.65	0.63	+ 0.05	+ 8,3%
3	0.68	0.62	0.86	0.87	+ 0.18	+ 26,5%

To analyse whether the addition of the CHM is increasing the prediction performance of the RF model, a McNemar test used on all three rasters. McNemar tests how many times the PCA +CHM was correct, and how many times the PCA was correct. From this the χ^2 is calculated, and the p-value expresses whether the difference is significant at a confidence interval of 90% and 95%. In table 11 below, the results of the McNemar tests are depicted. Both raster 2 and 3 are significantly improved after the addition of the CHM. However, this is not the case for raster 1. From the incorrect predicted classes, only class 1 was improved by the CHM. This difference is not enough to express significance and not enough to have a chi squared value. The H_0 is rejected for raster 2 and 3. The H_0 cannot be rejected for raster 1.

Because the Kappa considers the chance of agreement, based on what is observed, it is a stricter test than OA. Kappa, in table 10, increases slightly for Raster 1 (0.64 – 0.66), significantly more for Raster 2 (0.56 – 0.63) and the most in Raster 3 (0.62 – 0.86). Indicating that the CHM improved beyond-chance agreement for two of the three rasters. These findings are consistent with the McNemar tests (table 11).

Table 11: McNemar test results assessing whether CHM significantly improves RF prediction accuracy. Improvements are significant in Rasters 2 and 3.

Raster	**χ^2**	**p-value**	**Significant at 90%**	**Significant at 95%**
1	0	1	NO	NO
2	12,5	<0.01	YES	YES
3	4,9	0.027	YES	YES

4.3.2 Class-specific performance

To assess the per-class performance of the classification models, three metrics were computed for each species: precision, recall, and F1-score. These metrics were compared between models trained with four principal components (PCA-only) and models augmented with canopy height information (PCA + CHM). Tables 12, 13 and 14 visualise the resulting confusion matrices and performance metrics. In these tables, columns represent the true species labels, while rows represent the species predicted by the RF model. When a cell contains two values, the left value corresponds to the PCA-only model, and the right value corresponds to the PCA + CHM model. Allowing both results to be displayed in a single matrix. The rightmost columns report the per-class precision for both methods, while the bottom rows present the recall and F1 scores for each species.

Across all three rasters, species such as Douglas Fir, Trifoliate Orange, Bollwiller Pear, and European Linden consistently scored high across all three metrics (F1 ≥ 0.90), indicating strong spectral distinctiveness and effective classification. Their performance remained stable regardless of the inclusion of CHM, suggesting that spectral information alone was sufficient.

In contrast, species like *Common Beech*, *Common Ivy*, and *Thorny Olive* showed persistently low values across all metrics (F1 ≤ 0.50), indicating that they are either spectrally confused with other species or poorly represented in the training data. For these classes, the addition of CHM only marginally improved classification.

Importantly, several species demonstrated a significant improvement with the addition of CHM. *Hazel* showed the most dramatic gain, with its precision increasing from 0.25 to 0.75, and F1-score from 0.18 to 0.55. Similarly, *Gooseberry* improved from 0.50 to 0.90 in F1-score, and *European Spindletree* from 0.33 to 0.67. These results suggest that canopy height adds important structural information that helps distinguish species with similar spectral profiles.

However, the CHM was not uniformly beneficial. *Chocolate Vine*, for instance, showed a decrease in recall (0.75 to 0.5) and F1-score (0.86 to 0.67). A similar drop was observed for *Sweet Chestnut*. These declines may reflect the structural ambiguity of these species (e.g., low-stature vines or multi-layered crowns), which can confuse the height-based model input.

Overall, the results demonstrate that CHM enhances the classifier's ability to distinguish certain species, especially those with vertical structure not captured in the spectral data. Yet, the effect is species-specific: while some species benefit substantially from the added predictor, others may experience performance losses due to increased confusion or structural noise.

In addition to overall performance improvements, class-specific metrics reveal clear patterns of over- and underestimation. Species such as *Thorny Olive*, *Common Beech*, and *Crack Willow* exhibited low precision scores, suggesting they were frequently predicted even when incorrect—an indication of overestimation. For example, *Thorny Olive* showed a large precision increase from 0.00 to 0.50 after the inclusion of the CHM, while recall remained low, indicating continued false positives. Conversely, several species showed signs of underestimation, characterized by low recall but moderate-to-high precision. *Common Alder* and *European Spindletree* were rarely detected in the PCA-only model, with recall scores of 0.00 and 0.25, respectively, but improved with the inclusion of CHM. *Common Ivy* consistently had low recall (0.33) despite a perfect precision score (1.00), indicating the model rarely identified it correctly, though it was highly confident when it did. These patterns highlight the differential response of species to the

classification process, with some consistently misclassified due to confusion with other classes or insufficient spectral or structural distinctiveness.

The results of the random forest classification are visualised in figures 17, 18 and 19. The legend shows each species and the percentage of pixels that belong to each species.

Table 12: Confusion matrix of Raster 1, summarizing classification accuracy for eleven tree species. Rows represent the true species and columns represent the predicted species. To allow direct comparison between the two RF models within one table, each matrix cell is split into two sub-columns: the left sub-cell shows the number of pixels predicted by the PCA model, while the right sub-cell shows predictions from the PCA + CHM model. Precision (right) and Recall (bottom) values are calculated separately for both models. Colors in precision and recall indicate performance, with higher values reflecting better agreement between true and predicted labels.

True → Predicted ↓	Chocolate Vine		Hazel		Grey Alder		Common Ivy		Douglas Fir	Quacking Aspen		Oak	Black Locust		Crack Willow		Yew		Total		Precision	
																					PCA	PCA+CHM
Chocolate Vine	3	2	0		0		0		0	0		0	0		0		0		3	2	1	1
Hazel	0	1	3	4	2	1	0	1	0	0		0	0		0		0		5	7	0,6	0,57
Grey Alder	0		0		2	3	0		0	0		0	0		0		1		3	4	0,67	0,75
Common Ivy	0		0		0		2		0	1		1	1	0	0		0		5	4	0,4	0,5
Douglas Fir	0		0		0		2		8	0		0	2		2	4	0		14	16	0,57	0,5
Quacking Aspen	0		0		0		1		0	10	9	1	0		0		0		12	11	0,83	0,82
Oak	1	0	1	0	0		1	0	0	0	1	15	0		0	2	1	0	19	18	0,79	0,83
Black Locust	0		0		0		0		0	0		0	1	2	3	0	0		4	2	0,25	1
Crack Willow	0		0		0		0		0	0		1	0		5	4	0		6	5	0,83	0,8
Yew	0	1	0		0		0		2	0		0	0		0		4	5	6	8	0,67	0,63
Total	4		4		4		6		10	11		18	4		10		6		60			
Recall PCA	0.75		0.75		0.5		0.333		0.8	0.909		0.833	0.25		0.5		0.667					
Recall PCA+CHM	0.5		1		0.75		0.333		0.8	0.818		0.833	0.5		0.4		0.833					
F1 PCA	0.86		0.67		0.57		0.36		0.67	0.87		0.81	0.25		0.63		0.67					
F1 PCA+CHM	0.67		0.73		0.75		0.40		0.62	0.82		0.83	0.67		0.53		0.71					

Table 13: Confusion matrix of Raster 2, summarizing classification accuracy for eleven tree species. Rows represent the true species and columns represent the predicted species. To allow direct comparison between the two RF models within one table, each matrix cell is split into two sub-columns: the left sub-cell shows the number of pixels predicted by the PCA model, while the right sub-cell shows predictions from the PCA + CHM model. Precision (right) and Recall (bottom) values are calculated separately for both models. Colors in precision and recall indicate performance, with higher values reflecting better agreement between true and predicted labels.

True → Predicted ↓	C. Vine	S. Chestnut	Hazel	T. Olive	E. Linden	C. Alder	G. Alder	E. Spindle	C. Beech	C. Ivy	D. Fir	Q. Aspen	B. Locust	W. Willow	C. Willow	Yew	Total	Precision	
																		PCA	PCA+CHM
C. Vine	6	0	0	1	0	0	0	1	0	2	0	0	0	0	0	1	11	0.55	0.55
S. Chestnut	0	4	0	0	0	0	0	0	1 | 3	0	0	0 | 2	0	0	0	0	5 | 9	0.8	0.44
Hazel	0	0	1 | 3	0	0	0	0	0 | 1	0	0	0	0	0	0	3 | 0	0	4	0.25	0.75
T. Olive	0	0	0	0 | 1	0	0	3 | 0	0	0	1 | 0	0	0	0	0	0	0 | 1	4 | 2	0	0.5
E. Linden	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	5	1	1
C. Alder	0	0	0	0	0	0 | 1	0	0	0	0	0	0	0	0	0	0 | 1	1*	1
G. Alder	0	0	2 | 0	0	1	0	3	1 | 0	0	0	0	0	0	0	0 | 2	0	7 | 6	0.43	0.5
E. Spindle	0	0	0	0	0	1 | 0	0	1 | 2	0	0	0	0	0	0	0	0	2	0.5	1
C. Beech	0	0	0	2	0	0	0 | 1	1 | 0	1	0	0	1 | 0	0	0	0	0	5 | 4	0.2	0.25
C. Ivy	0	0	0	0	0	0	0	0	0	1 | 2	0	0	0	0	0	0	1 | 2	1	1
D. Fir	1 | 0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	8 | 7	0.88	1
Q. Aspen	0	0	0	1 | 0	0	0	0	0	2 | 0	1	0	4 | 3	0	0	0	0	0	8 | 4	0.5	0.75
B. Locust	0	0	1 | 0	0	0	0	2	0	0	0 | 1	0	0	4	0	0	0	0	7 | 7	0.57	0.57
W. Willow	1 | 2	0	0	0	0	0	0	0	0	0	0	1	0	15	0	0	0	17 | 18	0.88	0.83
C. Willow	0	0	3 | 4	0	0	3	0	0	0	1 | 0	0	0	0	0	0	3 | 4	1 | 0	0	11 | 13	0.27	0.31
Yew	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	1	1
Total	8	4	7	4	6	4	8	4	4	6	7	6	4	15	6	6	69		
Recall PCA	0,75	1	0.143	0	0.833	0	0.375	0.25	0.25	0.167	1	0.667	1	1	0.5	0.667			
Recall PCA+CHM	0.75	1	0.429	0.25	0.833	0.25	0.375	0.5	0.25	0.333	1	0.5	1	1	0.667	0.667			
F1 PCA	0.63	0.89	0.18	0	0.91	0	0.40	0.33	0.22	0.29	0.93	0.57	0.73	0.94	0.35	0.80			
F1 PCA+CHM	063	0.62	0.55	0.33	0.91	0.40	0.43	0.67	0.25	0.50	1	0.6	0.73	0.91	0.42	0.8			

Table 14: Confusion matrix of Raster 3, summarizing classification accuracy for eleven tree species. Rows represent the true species and columns represent the predicted species. To allow direct comparison between the two RF models within one table, each matrix cell is split into two sub-columns: the left sub-cell shows the number of pixels predicted by the PCA model, while the right sub-cell shows predictions from the PCA + CHM model. Precision (right) and Recall (bottom) values are calculated separately for both models. Colors in precision and recall indicate performance, with higher values reflecting better agreement between true and predicted labels.

True → Predicted ↓	Trifoliate Orange	Thorny Olive	Autumn Olive	Gooseberry	Beach Rose	Black Elder	Bollwiller Pear	Field Maple	Total	Precision	
										PCA	PCA+CHM
Trifoliate Orange	4	0	0	0	0	0	0	0	4	1	1
Thorny Olive	0	1 | 2	1	0 | 1	0	1 | 0	0	0	3	0.33	0.5
Autumn Olive	0	1	3	0	0	0	0	0	4	0.75	0.75
Gooseberry	0	0	0	5 | 9	4 | 1	0	0	1 | 0	10	0.5	0.9
Beach Rose	0	2 | 1	0	3 | 0	9 | 12	0	0	1 | 0	15	0.6	1
Black Elder	0	0	0	1 | 0	1 | 2	9 | 10	0	0	11	0.82	0.83
Bollwiller Pear	0	0	0	1 | 0	0	0	4	0	5	0.8	1
Field Maple	0	0 | 1	0	0	1 | 0	1	0	6 | 8	8	0.75	0.8
Total	4	4	4	10	15	11	4	8	49		
Recall PCA (4PCs)	1	0.25	0.75	0.5	0.6	0.818	1	0.75			
Recall PCA+CHM	1	0.5	0.75	0.9	0.8	0.909	1	1			
F1 PCA	1	0.29	0.75	0.50	0.60	0.82	0.89	0.75			
F1 PCA+CHM	1	0.5	0.75	0.9	0.89	0.87	1	0.89			

4.3.3 RF Visualisation

Raster 1

Using the reduced dataset at a resolution of 44cm with 4 PCs + CHM, the raster 1 classification visual shows a very sparse class inventory relative to the training set. Only three of the ten species are represented in the map. Among the missing species are:

- Chocolate vine
- Hazel
- Grey Alder
- Douglas Fir
- Black locust
- Crack willow
- Yew

The legend indicates that Quaking Aspen and Oak together account for around 99% of all classified pixels, with Common Ivy only 0.2%. Visually, the predictions organise in contiguous blocks that follow tree crown shapes rather than producing a salt-and-pepper effect. Seven trained classes are absent in the visual, which means the map is effectively dominated by two broadleaf tree categories with the rare occurrence of an understory climber. As a result, this figure shows that, under the chosen setting and variables, the RF produced a highly skewed class distribution in Raster 1 despite a broader training inventory, and that the spatial patterns do coincide with crown geometry.

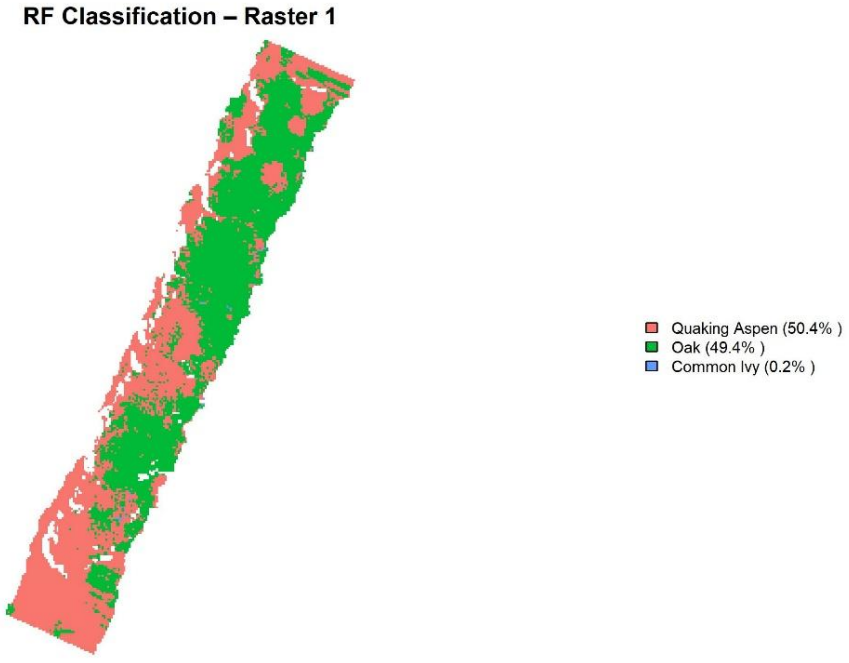


Figure 17: Random Forest species classification for Raster 1 (44cm, 4PCs + CHM). Only three of ten species appear, showing strong dominance by two broadleaf species, Quacking Aspen and Oak, and minimal representation of understory species.

Raster 2

Under the same conditions (4 PCs + CHM, 44cm resolution), the second Raster map shows a richer legend compared to figure 18. Eight of the thirteen species trained on are reported in the classification. The missing species are:

- Hazel
- European Spindle
- European linden
- Yew
- Black Locust

Nonetheless, the distribution remains highly uneven, with Quaking Aspen reporting nearly 50% of all classified pixels, and White Willow around 30%. This leaves the remaining six mapped species to fill comparatively small and scattered patches. In contrast to Raster 1, the spatial pattern is more speckled and fragmented, with multiple small clusters and edge pixels rather than large, crown-filling blocks. As a result, Raster 2 shows that the RF model can express a higher diversity in classes, but the effective map still shows that there are still species that dominate the classification. The class distribution remains skewed.

Figure 18: Random Forest species classification of Raster 2 (44cm, 4PCs + CHM). Eight of thirteen species identified, showing fragmented spatial patterns. The high amount of Quacking Aspen and White Willow reflects spectral dominance and class imbalance.

Raster 3

For Raster 3, the RF model (4 PCs + CHM, 44cm) maps seven of the eight trained species. Only one trained class is not present in the output legend, which is the European Spindletree. Compared to Rasters 1 and 2, the mapped classes seem to be a little bit more evenly distributed, but there is still a skewedness in the model. The images show a grained mosaic of shrub and small trees with numerous small clusters and short runs of contiguous pixels. Just like Raster 2, this image seems more scattered than Raster 1, but there is still no excessive salt-and-pepper effect at the crown scale. As a result, Raster 3 shows that the classifier can produce a nearly complete class coverage (7 of 8 present) in this strip and the spatial distribution shows a heterogeneous, layered vegetation structure, which is expected of the area.

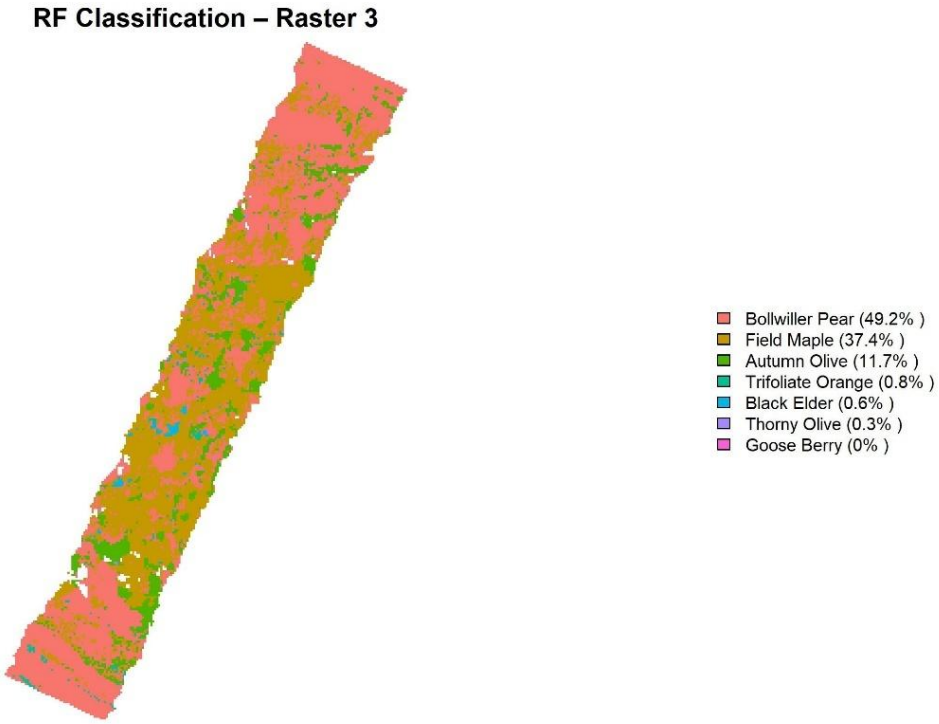


Figure 19: Random Forest species classification of Raster 3 (44cm, 4PCs + CHM). Showing seven of eight species in a mosaic like pattern that reflects Ketelbroek mixed structure.

5. Discussion

In this part, I evaluate the key findings of this research and will address the three research questions. First, there is an assessment of the quality and representativeness of the UAV-derived hyperspectral and canopy-height data, and the patterns and biases between flightlines. The spectral separability results, quantified by the Jeffries Matusita distance, are placed within the broader remote-sensing literature on spatial aggregation and band importance. Next, the effect of integrating a CHM into a PCA based RF classifier is discussed by examining overall accuracy gains and shifts in species specific performance. Finally, I will reflect on methodological considerations, acknowledge the study's limitations in sampling and model generality, and outline recommendations for the improvement of species mapping in complex agroforestry environments.

5.1 Data availability and quality

5.1.1 Sampling

In this study, I initially adopted a sampling strategy designed to capture as much of the variety in species present in Ketelbroek, which is around 200 species within 2.4 ha. Keeping in mind that much of the forest is inaccessible and that targeting known hotspots of diversity would maximize area coverage. In practice, however, this broad approach produced many unique species records but too few replicates per species to support a robust model training. As previously mentioned, over 65% of the recorded species in the first three rasters have less than 10 samples (see Table 3). Studies show that, in diverse systems, estimates of richness and classification accuracy depend on having sufficient samples of each species (Gimaret-Carpentier et al., 1998; Jeliazkov et al., 2022). When most species occur only once or twice in the dataset, statistical power reduces and machine learning models struggle to distinguish class boundaries (Ahrens et al., 2016).

Therefore, in retrospect, a more balanced or stratified sampling design would tackle the class imbalance and resulted in a higher accuracy (Feng et al., 2019; Nguyen & Chen, 2024). A minimum count per species should be set for each targeted species. This makes the dataset more balanced and ensures enough samples per species. For this research, this is attempted by removing all species with a count below 10. This threshold was chosen because this resulted in approximately 10 species to train on in per raster. If the threshold was higher, only three species per raster remained, which is not deemed enough. A threshold of 10 was in this case the trade-off between number of unique species and enough samples to train and test. Consequently, only a subset of the total recorded species could be included in the classification analysis. Depending on the raster, between eight and thirteen species met the minimum threshold of ten samples per class (table 3). While this subset represents the most abundant species, it does not cover the full range of diversity of Ketelbroek.

Another method that captures class imbalance well is by augmenting data. By generating synthetic samples, it is possible to capture the essence of a species without having samples of it. Data augmentation can help increase the sample size and is therefore a useful tool. However, it requires careful implementation to avoid overfitting or introducing noise (Kumar et al., 2024; Shorten & Khoshgoftaar, 2019). Since there is already a lot of noise to deal with in the current data set, I chose not to work with augmented data.

5.1.2 Inter-strip consistency and linear regression

When looking at the data strip by strip, the reflectance values have a range of values with a common trend. Adjacent strips show a consistent brightness shift (with each adjacent strip, the reflectance values increase), with the largest absolute differences in the red-edge and NIR (see Table 6, Figure 8). It is likely that is due to the rising sun and therefore increasing the angle at which the sunlight hits an object (Hashimoto et al., 2019). The flight time of the UAV was on a clear day at noon. The sun is at its highest around 2 o'clock in the afternoon.

When comparing overlapping areas from different flightlines, systematic differences in reflectance values are an issue that is found in various research. These systematic differences are often attributed to bidirectional reflectance distribution function (BRDF) effects, differences in solar illumination, and sensor viewing geometry. This can vary between flightlines and especially impact mosaicking and classification (Jia et al., 2024). The BDRF effect is inherent to UAV sensors and leads to variations in observed reflectance depending on the sun-target-sensor geometry. This effect can cause systematic differences between the overlapping areas of flightlines (Z. Wang & Liu, 2016). These differences in reflectance values are often most seen in the visible spectrum and less in the NIR (Perry et al., 2000). In addition, variations in atmospheric conditions and solar zenith angles between flightlines can further contribute to reflectance inconsistencies. However, this is corrected for prior to this research.

The linear regression was applied as a quick approach to harmonize the individual raster bands (Figures 9 – 12 Appendix 8 and 9). The intention was to make the flightlines more comparable so that a full mosaic could be generated. However, the per-band regressions only partially reduced the differences, and inconsistently across wavelengths. This indicates that the approach is not robust for a harmonization method. This limitation shows a weakness of simple least-squares regression when pixels are spatial and/or spectral not perfectly aligned. As Liu and Scaglione (2025) demonstrate, standard linear regression methods become unreliable under such conditions because they assume a one-to-one correspondence between the in- and outputs. When those relationships are misaligned, the regression fails to capture the underlying structure of the data. Similarly, De Jager and Fox (2013) show that pixel-level regression methods are better suited to describe local or temporal trends across rasters than as a correction method. For this reason, the rasters in this research are treated individually, allowing underlying spectral trends to remain visible and interpretable, while also reducing computing time compared to a large-scale mosaic.

5.2 Spectral separability

The spectral separability of the species pairs located in the three rasters in Ketelbroek range between a low moderate to moderate separability (1.0 – 1.4). However, this range is typically accepted for highly diverse environments due to similar vegetation and mixed pixels. In studies of invasive species and wetland habitats, JM values for challenging pairs often fall within or just above this range, especially when using standard hyperspectral data in dense, mixed environments. For example, mangrove pairs showed JM values of 0.93 – 1.12 in green periods and 1.18 – 1.34 in senescence periods, with higher values only achieved by combining phenological stages or advanced feature engineering (Tian et al., 2020). Similarly, benthic habitat mapping found that most class pairs had $S < 1.9$, and only a few exceed this threshold, directly impacting classification accuracy (Wicaksono & Aryaguna, 2020b).

5.2.1 Relevance of aggregation

Increasing the pixel size (lowering the spatial resolution from 7cm to ~44cm) improved the mean JM separability and stabilized its variance up to a threshold (Figs. 13 – 16) (Chen et al., 2023). At 7cm, pixels were so fine that they often sampled sub-crown elements and shadow, resulting in an increase in within-species variance and weakening the separability. Coarsening the resolution to ~44cm averaged this within-crown variability while still sampling a single crown, producing spectral signatures that were more representative of each species (Peña et al., 2013). This noise reduction explains the observed gain in separability. There is, however, a maximum to this. Once pixels become large enough to span multiple crowns, mixed pixels reduce separability. The optimal pixel size is case specific.

Dalponte et al. (2013) examined the effect of spatial resolution (range from 40cm to 1.5m) on tree species classification and class separability. They found that using a resolution of 40cm increased classification accuracy by about 20% relative to the coarser resolution of 1.5m. The results of my research show the same scale dependence, although I started much finer (7cm) aggregating to 44cm. In agriculture studies, optimum resolutions near 10cm are reported, but those studies typically involve monocultures or homogeneous grasslands (Lu & and He, 2018; Zeng et al., 2024). Overall, the optimal pixel size depends on crown size and scene heterogeneity. Very coarse pixels mix multiple canopies and reduce accuracy, while extreme fine pixels over-resolve within crown variation. Intermediate pixels that roughly match crown dimensions perform best. In the case of my research, that optimum is ~44cm.

Although the optimal mean JM separability occurs at 44cm, the variance of JM across species pairs decreases as pixel size is coarsened (lower spatial resolution). Aggregation averages within-crown variability and reduces the influence of rare or outlying spectra, so pairwise spectral differences appear more uniform and JM values cluster more tightly pairs (D'Amico et al., 2024; Piiroinen et al., 2017). In my data, reducing the resolution from 7cm to 44cm increases the mean JM separability and reduces its variation across species pairs. This tighter range comes from noise being averaged out, not from weaker discrimination.

5.2.2 Species

Viewed across the three rasters, the separability is shaped by both the food-forest ecology and by the sensors. A possible explanation for the nine pairs with the lowest JM distances could be the similar biochemical and structural make-up at the time of the image acquisition (late-spring, full leaf-on). Grey alder – hazel illustrates this most clearly: both are two medium broadleaves and measured when chlorophyll and water content are near their seasonal maximum. In that state, red absorption, a steep red-edge and a bright NIR plateau converge, while LiDAR captures rounded crowns with overlapping heights. Adding the CHM helps only marginally because the crowns are alike (Hovi et al., 2017; Shi et al., 2021). Common Ivy (read Ivy) complicates the situation even further. As a climbing plant, Ivy is recorded not just as an understory but also as vegetation located high in host trees. Ivy – Douglas fir and Ivy – Quaking Aspen therefore mix vine and host signals within single pixels, and the evergreen status of Ivy erases phenological contrast in a summer scene. Two shrub – shrub comparisons: Black Elder – Gooseberry and Beach Rose – Gooseberry, show a similar compression. They are dense, 1–3 m mounds with many thin twigs and therefore read almost identically to LiDAR, while leaf-scale differences (pinnate vs. lobed; rugose vs. smooth) may be lost at 44 cm under mixed illumination. Even Field Maple – Gooseberry, seemingly a tree–shrub contrast, collapses along crown edges were overhanging maple leaves

and shrub foliage share pixels. In short, where leaf chemistry converges, crown architecture resembles, and vertical mixing is common, JM distances sink toward ~1.03–1.36 (Table 8 and 9).

At the other end of the spectrum, the highest JM values (~ 1.41, Table 8 and 9) indicate very strong class separation. These cases can be explained by differences in both leaf properties and canopy structure. In Raster 1 for example, Black Locust and Quacking Aspen are both deciduous trees, but they are structurally and visually different. The Black Locust has compound leaves and open crowns, the Quacking Aspen has simple, fluttering leaves and a denser canopy. These differences affect both the wavelength signal and CHM distribution, resulting in a stronger separability. Similarly, Common Ivy and Yew show high separation for the opposite reason of Ivy's usual confusion with other species. In this case, Ivy remains as a shaded vine in the understory, while Yew forms a dense, needle leaf canopy above it. This vertical layering reduces the spectral mixing and enhances the contrast between the needle and broad leaves (see Appendix 4).

In Raster 2, the pairs Black Locust – Chocolate Vine and Chocolate Vine – Crack Willow represent clear vertical stratification that occupy different height zones. The LiDAR height data amplifies the differences that are already visible in the reflectance. Douglas Fir and European Linden form a conifer and broadleaf pair, where darker evergreen leaves and lighter deciduous phenology add to the structural and spectral contrast (see Appendix 4).

Finally, Raster 3 contains three pairs where both phenology and structure differ. Autumn Olive and Bollwiller Pear contrast a silvery hairy leaved shrub with a smoother orchard canopy. Bollwiller Pear and Trifoliate Orange differ through the thick oily trifoliate leaves, which affect the absorption (see Appendix 4).

The extremes in the JM distances show a pattern. The lowest JM values likely result from species that appear similar in summer, due to similar biochemistry, shrub layers that look alike and strong spectral mixing. The highest JM values occur when differences in leaf structure are picked up, and crown position are in different directions making them more unique. Improving separability will depend on data and methods that further reduce similarities. For example, adding seasonal variables, detailed 3D information on vegetation structure, and applying object-based analysis could improve the JM values.

5.3 Random Forest

The Random Forest (RF) experiments were designed to test whether adding the Canopy Height Model (CHM) to hyperspectral components improves species mapping in a layered (food) forest. At the raster level, the effect is clear in two of the three cases: overall accuracy (OA) and Kappa (Table 10) rise from 0.60 to 0.65 (raster 2) and 0.68 to 0.86 (raster 3), with McNemar tests significant at 95% confidence (Table 11). Raster 1 shows only a marginal OA gain and no significant difference. These results are consistent with the species level patterns visible in the confusion matrices (Tables 12 – 14). Together, these patterns indicate that a CHM can add information if there is vertical stratification, but also offering little benefit where classes already separate spectrally or where structures overlap (H. Wang et al., 2021).

There are two design choices that shaped the magnitude of the RF metric gains. First, I trained and tested at a resolution of 44 cm because separability increases with aggregation and stabilizes near this scale. Larger pixels capture canopy reflectance and suppress within-crown noise without exceeding crown size. This decision increases mean JM and, and therefore, the ceiling for RF accuracy. Second, I kept the flightlines separate instead of mosaicking them, to avoid

radiometric differences and BRDF effects between rasters. This approach reduces variation that is unrelated to species and makes structural patterns in the data easier to detect.

5.3.1 Species level

At the species level, the RF is comparable to the separability analysis. Species that were spectrally distinctive generally achieved higher and more stable scores regardless of CHM (See Appendix 12 – 14). For example: Douglas Fir, European Linden, Trifoliate Orange and Bollwiller Pear scored F1 ≥ 0.90. While Common Beech, Common Ivy and Thorny Olive remain a weak F1 (≤ 0.50), suggesting spectral confusion and/or understory mixing. Most importantly, several problem classes improved dramatically when CHM was added: Hazel (F1 ≈ 0.18→0.55), Gooseberry (F1 ≈ 0.50→0.90), and European Spindletree (F1 ≈ 0.33→0.67). A possible explanation for this is that these species may tend to have overlapping spectral signatures but differ enough in height and/or crown density, and therefore the CHM resolves ambiguity. In contrast, Chocolate Vine and Sweet Chestnut decline in F1 after CHM is introduced, which is consistent with the structure of these species being liana and low crown layers, respectively. This causes height (CHM) to increase noise instead of reducing.

5.3.2 Link to spectral separability

By linking the outcomes of the RF to the spectral separability scores, I try to strengthen the interpretation of both results. The lowest-JM pairs such as, Grey Alder–Hazel (≈ 1.03), Crack Willow–Hazel (≈ 1.04), Common Ivy–Douglas Fir (≈ 1.13), and Common Ivy–Quaking Aspen (≈ 1.17) (see Table 9), may be a result of a similarity in leaf chemistry at the time of measurement, and the mixing of pixels (climbers, understory). In addition, it is these species that show lower RF scores and/or are sensitive to CHM. On the other hand, the highest-JM pairs like, Douglas Fir – European Linden, Black Locust – Chocolate Vine, and Bollwiller Pear – Trifoliate Orange (≈ 1.414) are also among the easier to classify (see Table 9). The combination of divergent leaf structures (needle/broad leaf, single/compound leaf) and minimal/no overlapping canopy structures results in these higher results. The pairwise extreme of the spectral separability therefore can foreshadow RF behaviour at the species level.

When comparing my results to available literature, in diverse, mixed canopies, JM values clustering around 0.8 – 1.4 are common. Furthermore, to acquire high accuracies the research typically requires either structural information (LiDAR) or multi-date acquisitions (leaf-off/leaf-on), or both, for species differentiation (Hologa et al., 2021; Michałowska & Rapiński, 2021). My findings are consistent with that pattern: hyperspectral imagery and a CHM improve OA in two of three rasters, and the largest gains occur where shrub–tree mixtures are strongest. Prior research reports similar benefits from multi-sensor use and emphasizes the additional value of leaf-off data for separating deciduous species.

In sum, the RF results are generally consistent with the separability analysis and broadly in line with prior studies: the addition of structure (CHM) to the RF model improves its performance where spectra alone are ambiguous. However, the benefit is species-specific and depends on the scene geometry. For (food) forest mapping, this could mean that reliable species classification is possible for structurally and optically distinct species, and for difficult understory pairs once temporal (leaf-off) and structural (richer LiDAR features) information are added.

5.3.3 Classification visualization

The RF classification maps (Figures 13 – 15) visualise how the trained model translates spectral and structural inputs into spatial patterns across the three flightlines. These maps are not intended as literal depictions of the full species composition of Ketelbroek, but rather as a diagnostic view of model behaviour when applied to a filtered, class-imbalanced dataset. The uneven representation of species in the training data is clearly reflected in the spatial outputs. The dominant classes occupy large, contiguous patches, while rare or spectrally ambiguous are either underrepresented or absent. This skewed distribution highlights the sensitivity of the RF model to sample imbalance and spectral overlap (Feng et al., 2019). These are two factors that are intrinsic to species rich environments and complex canopy layering (Fassnacht et al., 2016).

In Raster 1 (figure 17), only three of the ten trained species appear in the final classification, with Quacking Aspen and Oak collectively accounting for nearly all predicted pixels. Their dominance likely comes from the large crowns and distinctive spectral-structural signatures, which results in the model to generalise beyond limited samples (Hologa et al., 2021). The absence of the majority of the species indicates that, at the current resolution (44cm) and with the restricted training dataset, the classifier prioritises broader structural cues over finer spectral distinctions (Michałowska & Rapiński, 2021). In Raster 2 (figure 18), class richness is improved, where eight of the thirteen species are predicted. However, the pattern remains fragmented and uneven, with Quacking Aspen and White Willow dominating the area. The increased patchiness and noise suggests spectral confusion and overlapping canopies (He et al., 2018). On the other hand, Raster 3 (figure 19) produces a more balanced image, mapping seven of the eight species reported in the area. Here, the model benefits from a better combination of canopy heterogeneity, CHM variation, and class representation (Michałowska & Rapiński, 2021).

Together, these three maps illustrate the strengths and limitations of integrating hyperspectral and LiDAR derived canopy height information in a (food) forest. Where there is vertical stratification and distinct crown architecture, the CHM clearly improves species discrimination. However, where understory and overstory overlap or species share similar visual properties, confusion remains. The dominance of a few species also shows that pixel-based RF classification struggles to capture ecological complexity when class balance and sample density are low. Nevertheless, the fact that the RF classifier can produce crown-scale patterns rather than random noise demonstrates that the combined hyperspectral – CHM approach successfully shows meaningful ecological structures.

5.4 Limitations

This study was subject to several limitations, both technological and ecological, that influenced the scope and interpretation of the results. Methodologically, the most persistent constraint was the limited and uneven sampling across species. Although the sampling strategy aimed to capture the widest range of species in Ketelbroek, the resulting limited dataset contained a strong imbalance in samples. This limited dataset means that results primarily reflect the dominant species, and not the rarer species that make up most of the forest's overall diversity (Foody, 2009; Gimaret-Carpentier et al., 1998). Approximately two-thirds (Table 3) of the recorded species had fewer than ten usable samples, reducing statistical robustness and constraining the RF model's ability to learn class boundaries. This imbalance forced a trade-off between taxonomic diversity and model reliability, ultimately limiting the number of species that could be included in the classification. Class imbalance also increases over- and under prediction effects, which is

reflected in the uneven spatial patterns seen in the RF visualisations (Figure 17 – 19) (Feng et al., 2019).

From a data acquisition point, several technical limitations affected spectral quality and consistency. The hyperspectral results of all fifteen flightlines show high variation between each raster. This is probably caused by Bidirectional Reflectance Distribution (BDRF) effects, illumination differences, and sensor geometry (Jia et al., 2024; Z. Wang & Liu, 2016). Although a linear correction per band was tested, the residual errors remained, leading to local brightness changes between the rasters. Treating flightlines separately avoided the spread of these biases. However, it restricted the analysis to sections of the research area (rasters 1 – 3) and reduced the sample size. In addition, the high spatial resolution (7 – 44cm) shows a higher variation within individual tree crowns, especially where overlapping canopies and shadows occurred. These factors are to be expected to UAV-based imagery of complex forests and show the difficulty in achieving spectral uniformity and vertically layered vegetation (Banerjee et al., 2020).

Algorithmic limitations further influenced the results. The RF classifier is well suited for high dimensional and nonlinear data, but it remains a pixel-based model without spatial awareness. As a result, it does not account for contextual information such as crown shape or neighbourhood texture, which can improve discrimination in heterogeneous areas. The model's structure also makes it prone to overfitting when sample size per class is low or spectral variance is high (Feng et al., 2019). Alternative approaches, such as Support Vector Machines (SVM) for margin-based separation, or Convolutional Neural Networks (CNNs) and other deep learning methods capable of leveraging spatial context, could potentially overcome some of these constraints. However, these methods demand larger, more balanced datasets and more extensive computational resources than were available for this research (Hologa et al., 2021; Raczko & Zagajewski, 2017).

Ecologically, several contextual factors shaped the limits of the analysis. Data was collected during late spring, in full leaf-on conditions. Meaning that chlorophyll and water content is at its peak across all species. This reduces spectral variability, making deciduous species less separable (Fassnacht et al., 2016; Tian et al., 2020). Furthermore, parts of the forest were inaccessible due to dense vegetation and wet terrain. This ultimately resulted in spatial sampling bias toward open or easily reachable areas. Even though the infield sampling is combined with a desktop analysis of an RGB image, the representation across the complete forest is limited. The CHM smoothed some of these effects by introducing structural diversity, but it could not fully capture fine scale vertical complexity.

These technical and ecological limitations restrict the generalizability of the results beyond the specific conditions of Ketelbroek. While the research demonstrated that UAV hyperspectral imagery and LiDAR derived data can classify species in a highly diverse agroforestry system, the accuracies remain very context dependent. Nevertheless, by identifying these uncertainties, this research shows where the methods perform reliably and where future improvements are needed to go towards more promising results across sites and seasons.

5.5 Recommendations

During this research, several directions for future research emerged in both methodological design and practical monitoring of (agro) forestry systems. First, future research should aim for balanced sampling across species. Establishing a high enough minimum number of samples per species, or making use of augmented data, can help improve a class imbalance and stabilise the performances of the RF model (Feng et al., 2019; Foody, 2009; Jeliazkov et al., 2022). In addition, by linking field data to automatically detected tree crowns, more pixels can be labelled for each

known species, giving a better image on the composition of a tree (Piiroinen et al., 2017). Integrating these systematic protocols within national monitoring programs, such as the Nationaal Monitoringsprogramma Voedselbossen, could also help standardize data quality across sites in the Netherlands (Wendel et al., 2023).

Secondly, a multi-temporal and -sensor approach is strongly recommended. Repeating acquisitions during different seasons (especially leaf-on and leaf-off phases) introduces a high contrast that could drastically improve discrimination among deciduous and evergreen species (Fassnacht et al., 2016; Tian et al., 2020). Other structural metrics such as crown density indices, gap fraction, and leaf shapes could further improve separability between shrubs, vines and trees. Also, the addition of vegetation indices could result in a better representation of canopy traits that are relevant to species classification (Hologa et al., 2021; Rehman et al., 2024).

Thirdly, future research should explore object-based and more spatially aware classification methods. Transitioning from pixel-based RF models toward object-based image analysis, or convolutional neural networks (CNN) can reduce noise and incorporate spatial context such as crown geometry and patterns (He et al., 2018). However, these methods should be combined with careful feature selection to prevent overfitting and maintain interpretability (Michałowska & Rapiński, 2021). Considering feature selection, while PCA effectively reduced the dimensionality and noise in the data of this research, it remains an unsupervised method where class separability is not necessarily optimised. Therefore, combining PCA with supervised band selection techniques could improve interpretability and potentially increase the classification accuracy (Paul & Chaki, 2022; Taşkın et al., 2017).

This research demonstrates that UAV sensors and open-source analytical tools can already produce meaningful insight in the structure of tree species biodiversity in Ketelbroek. To translate these methods in routine monitoring, future projects should develop systematic protocols that can be adapted to varying forest types, spectral sensors and data resolutions. This would enable people, from researcher to landowners, to evaluate biodiversity and structural change consistently across years and regions.

6. Conclusion

This research aimed to evaluate how accurately airborne hyperspectral imagery, combined with the LiDAR derived Canopy Height Model can classify (woody) plant species in the highly diverse Dutch food forest Ketelbroek, using Random Forest classifiers. Working with over 200 species, strong structural layering, imbalance in samples, and large within crown spectral variation presented a challenging case. The findings show the potential and the limitations of this combined remote sensing approach.

The first research question focussed on analysing the differences between the three flightlines, whether there is a linear trend present, and if linear regression could solve these between strip differences. The analysis showed that there is a systematic brightness difference between adjacent rasters with the next one always being brighter. This occurred especially in the Red Edge and the NIR. These differences are likely caused by BRDF effects and subtle illumination changes during the time of hyperspectral data acquisition. Although linear regression reduced a small part of the bias, inconsistencies remained too big across all wavelengths. Therefore, this is not a suitable solution to harmonise the different rasters into one mosaic. Treating flightlines separately was the most reliable approach in preserving spectral integrity in this case.

The second research question investigated the effect of different spatial resolutions on species discrimination and focussed on species and species pairs that are either well spectrally separated or consistently confused. Across all rasters, Jeffries Matusita distances indicated (low) moderate separability. The introduction of coarser pixel size (from 7cm to ~44cm) significantly improved separability. At approximately 44cm, spectral signatures became more representative of whole crowns, capturing more of the characteristics of species. Species pairs with the lowest separability typically shared similar leaf traits, canopy forms or height. The most separable species pairs differed more strongly in either leaf structure or height.

The third research question focussed on Random Forest classification and the integration of a Canopy Height Model. Across two of the three rasters, adding the CHM substantially improved the model's accuracy, supported by the McNemar test. Height information was especially valuable for species that are spectrally similar, but structurally different, such as Hazel, Gooseberry and European Spindle. On the other hand, species with low or no structural height differences saw reduced performance. This shows that a Canopy Height Model is not always beneficial. The classification maps further showed that Random Forest can reproduce meaningful ecological patterns but remains sensitive to class imbalances and spectral mixing.

Overall, this study shows that the use of airborne hyperspectral data is a promising method for biodiversity assessment in a complex agroforestry system. With the use of Random Forest, several woody species are distinguished. Hyperspectral data alone resulted in a classification that is somewhat reliable, and the addition of a Canopy Height Model significantly improved species prediction for two of the three rasters. However, both datasets were limited in the ability to separate species in the understory, and species that are similar in structure and spectrum. This indicates that airborne hyperspectral data can identify species only to a moderate extent in complex forest environments and that CHM provides improvements in some cases but does not fully resolve the classification challenges. Even within these constraints, the findings illustrate how UAV-based remote sensing can contribute to monitoring frameworks for (emerging) agroforestry systems in the Netherlands. To further improve the assessment of biodiversity it is advised to make use of balanced sampling strategies, multi seasonal acquisitions, richer structural metrics, and spatially aware classifiers.

7. Use of generative AI statement

In this thesis, AI is used as a sparring partner and as a feedback tool to improve the quality of the texts. The scientific approach, methods, and content presented in this thesis is the work of the author, who takes full responsibility for the analysis and text.

8. References

Ahrens, D., Fujisawa, T., Krammer, H.-J., Eberle, J., Fabrizi, S., & Vogler, A. P. (2016). Rarity and Incomplete Sampling in DNA-Based Species Delimitation. *Systematic Biology*, *65*(3), 478–494. https://doi.org/10.1093/sysbio/syw002

Allaby, M. (2012). *A Dictionary of Plant Sciences*. Oxford University Press.

Almusaed, A. (2016). *Landscape Ecology: The Influences of Land Use and Anthropogenic Impacts of Landscape Creation*. BoD – Books on Demand.

Ariza, A., Robredo Irizar ,Marina, & and Bayer, S. (2018). Empirical line model for the atmospheric correction of sentinel-2A MSI images in the Caribbean Islands. *European Journal of Remote Sensing*, *51*(1), 765–776. https://doi.org/10.1080/22797254.2018.1482732

AZ, M. van A. Z. (2018, September 8). *Visie Landbouw, Natuur en Voedsel: Waardevol en Verbonden - Beleidsnota - Rijksoverheid.nl* [Beleidsnota]. Ministerie van Algemene Zaken. https://www.rijksoverheid.nl/documenten/beleidsnota-s/2018/09/08/visie-landbouw-natuur-en-voedsel-waardevol-en-verbonden

Ballemans, M. (2022, April 5). Voedselbos kaart. *Puur Permacultuur*. https://puurpermacultuur.nl/voedselbos-kaart/

Banerjee, B. P., Raval, S., & Cullen, P. J. (2020). UAV-hyperspectral imaging of spectrally complex environments. *International Journal of Remote Sensing*, *41*(11), 4136–4159. https://doi.org/10.1080/01431161.2020.1714771

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Berkhout, P., Van Der Meulen, H., & Ramaekers, P. (2023). *Staat van Landbouw, Natuur en Voedsel: Editie 2023*. Wageningen Economic Research. https://doi.org/10.18174/641747

Breidenbach, J., Dijkgraaf, E., Nijpels-Cieremans, R., & Strijkstra. (2017, May). Voedselbossen van belang voor biodiverseit. *De levende natuur*, *3*(118), 5.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

CBS. (n.d.). *Hoe wordt de Nederlandse bodem gebruikt? - Nederland in cijfers 2020 | CBS* [Webpagina]. Hoe wordt de Nederlandse bodem gebruikt? - Nederland in cijfers 2020 | CBS. Retrieved 15 October 2024, from https://longreads.cbs.nl/nederland-in-cijfers-2020/hoe-wordt-de-nederlandse-bodem-gebruikt

Chen, J., Chen, Z., Huang, R., You, H., Han, X., Yue, T., & Zhou, G. (2023). The Effects of Spatial Resolution and Resampling on the Classification Accuracy of Wetland Vegetation Species and Ground Objects: A Study Based on High Spatial Resolution UAV Images. *Drones*, *7*(1), 61. https://doi.org/10.3390/drones7010061

Dainelli, R., Toscano, P., Di Gennaro, S. F., & Matese, A. (2021). Recent Advances in Unmanned Aerial Vehicles Forest Remote Sensing—A Systematic Review. Part II: Research Applications. *Forests*, *12*(4), Article 4. https://doi.org/10.3390/f12040397

Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D., & Næsset, E. (2013). Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(5), 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

D'Amico, G., Nilsson ,Mats, Axelsson ,Arvid, & and Chirici, G. (2024). Data homogeneity impact in tree species classification based on Sentinel-2 multitemporal data case study in central Sweden. *International Journal of Remote Sensing*, *45*(15), 5050–5075. https://doi.org/10.1080/01431161.2024.2371082

De Jager, N. R., & Fox, T. J. (2013). Curve Fit: A pixel-level raster regression tool for mapping spatial patterns. *Methods in Ecology and Evolution*, *4*(8), 789–792. https://doi.org/10.1111/2041-210X.12068

Díaz-Delgado, R., & Mücher, S. (2019). Editorial of Special Issue "Drones for Biodiversity Conservation and Ecological Monitoring". *Drones*, *3*(2), 47. https://doi.org/10.3390/drones3020047

Dupraz, C., Lawson, G. J., Lamersdorf, N., Papanastasis, V. P., Rosati, A., & Ruiz-Mirazo, J. (2018). Temperate agroforestry: The European way. *Temperate Agroforestry Systems*, 98–152. https://doi.org/10.1079/9781780644851.0098

EZ&K, M. van E. Z. en. (2019a-06-28T14:10). *Klimaatakkoord—Publicatie—Klimaatakkoord* [Publicatie]. Ministerie van Economische Zaken en Klimaat. https://www.klimaatakkoord.nl/documenten/publicaties/2019/06/28/klimaatakkoord

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, *186*, 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Feng, W., Huang, W., & Bao, W. (2019). Imbalanced Hyperspectral Image Classification With an Adaptive Ensemble Method Based on SMOTE and Rotation Forest With Differentiated Sampling Rates. *IEEE Geoscience and Remote Sensing Letters*, *16*(12), 1879–1883. https://doi.org/10.1109/LGRS.2019.2913387

Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, *30*(20), 5273–5291. https://doi.org/10.1080/01431160903130937

Gimaret-Carpentier, C., Pélissier, R., Pascal, J.-P., & Houllier, F. (1998). Sampling strategies for the assessment of tree species diversity. *Journal of Vegetation Science*, *9*(2), 161–172. https://doi.org/10.2307/3237115

Hashimoto, N., Saito, Y., Maki, M., & Homma, K. (2019). Simulation of Reflectance and Vegetation Indices for Unmanned Aerial Vehicle (UAV) Monitoring of Paddy Fields. *Remote Sensing*, *11*(18), 2119. https://doi.org/10.3390/rs11182119

He, L., Li, J., Liu, C., & Li, S. (2018). Recent Advances on Spectral–Spatial Hyperspectral Image Classification: An Overview and New Guidelines. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(3), 1579–1597. https://doi.org/10.1109/TGRS.2017.2765364

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. *Remote Sensing*, *12*(1), 113. https://doi.org/10.3390/rs12010113

Hologa, R., Scheffczyk, K., Dreiser, C., & Gärtner, S. (2021). Tree Species Classification in a Temperate Mixed Mountain Forest Landscape Using Random Forest and Multiple Datasets. *Remote Sensing*, *13*(22), 4657. https://doi.org/10.3390/rs13224657

Hovi, A., Raitio, P., & Rautiainen, M. (2017). A spectral analysis of 25 boreal tree species. *Silva Fennica*, *51*(4). https://doi.org/10.14214/sf.7753

Hussain, K., Ilyas, A., Bibi, I., & Hilger, T. (2021). Sustainable Soil Loss Management in Tropical Uplands: Impact on Maize-Chili Cropping Systems. *Sustainability*, *13*(11), Article 11. https://doi.org/10.3390/su13116477

Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W. E., & Henle, K. (2022). Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology*, *28*(12), 3754–3777. https://doi.org/10.1111/gcb.16114

Jia, W., Pang, Y., & Tortini, R. (2024). The influence of BRDF effects and representativeness of training data on tree species classification using multi-flightline airborne hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *207*, 245–263. https://doi.org/10.1016/j.isprsjprs.2023.11.025

Joffre, R., Rambal, S., & Ratte, J. P. (1999). The dehesa system of southern Spain and Portugal as a natural ecosystem mimic. *Agroforestry Systems*, *45*(1), 57–79. https://doi.org/10.1023/A:1006259402496

Kizel, F., & Vidro, Y. (2023). An unmixing-based BRDF correction in spectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *118*, 103161. https://doi.org/10.1016/j.jag.2022.103161

Kumar, T., Brennan, R., Mileo, A., & Bendechache, M. (2024). Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions. *IEEE Access*, *12*, 187536–187571. https://doi.org/10.1109/ACCESS.2024.3470122

Liu, H., & Scaglione, A. (2025). *Shuffled Linear Regression via Spectral Matching* (No. arXiv:2410.00078). arXiv. https://doi.org/10.48550/arXiv.2410.00078

Lu, B., & and He, Y. (2018). Optimal spatial resolution of Unmanned Aerial Vehicle (UAV)-acquired imagery for species classification in a heterogeneous grassland ecosystem. *GIScience & Remote Sensing*, *55*(2), 205–220. https://doi.org/10.1080/15481603.2017.1408930

LVVN. (2024). *Broeikasgasuitstoot landbouw*. https://www.staatvanlandbouwnatuurenvoedsel.nl/kerncijfers/broeikasgasemissie-ipcc/

Michałowska, M., & Rapiński, J. (2021). A Review of Tree Species Classification Based on Airborne LiDAR Data and Applied Classifiers. *Remote Sensing*, *13*(3), 353. https://doi.org/10.3390/rs13030353

Nair, P. K. R., Kumar, B. M., & Nair, V. D. (2021). *An Introduction to Agroforestry: Four Decades of Scientific Developments*. Springer International Publishing. https://doi.org/10.1007/978-3-030-75358-0

Nguyen, K. A., & Chen, W. (2024). Enhancing Cover Management Factor Classification Through Imbalanced Data Resolution. *Environments*, *11*(11), 250. https://doi.org/10.3390/environments11110250

Osman, K. T. (2014). Soil Resources and Soil Degradation. In K. T. Osman (Ed.), *Soil Degradation, Conservation and Remediation* (pp. 1–43). Springer Netherlands. https://doi.org/10.1007/978-94-007-7590-9_1

Paul, A., & Chaki, N. (2022). Supervised data-driven approach for hyperspectral band selection using quantization. *Geocarto International*, *37*(8), 2312–2322. https://doi.org/10.1080/10106049.2020.1822929

Peña, M. A., Cruz ,Pablo, & and Roig, M. (2013). The effect of spectral and spatial degradation of hyperspectral imagery for the Sclerophyll tree species classification. *International Journal of Remote Sensing*, *34*(20), 7113–7130. https://doi.org/10.1080/01431161.2013.817712

Perry, E. M., Warner ,T., & and Foote, P. (2000). Comparison of atmospheric modelling versus empirical line fitting for mosaicking HYDICE imagery. *International Journal of Remote Sensing*, *21*(4), 799–803. https://doi.org/10.1080/014311600210588

Piiroinen, R., Heiskanen, J., Maeda, E., Viinikka, A., & Pellikka, P. (2017). Classification of Tree Species in a Diverse African Agroforestry Landscape Using Imaging Spectroscopy and Laser Scanning. *Remote Sensing*, *9*(9), Article 9. https://doi.org/10.3390/rs9090875

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, *50*(1), 144–154. https://doi.org/10.1080/22797254.2017.1299557

Rehman, A. U., Zhang, L., Sajjad, M. M., & Raziq, A. (2024). Multi-Temporal Sentinel-1 and Sentinel-2 Data for Orchards Discrimination in Khairpur District, Pakistan Using Spectral Separability Analysis and Machine Learning Classification. *Remote Sensing*, *16*(4), Article 4. https://doi.org/10.3390/rs16040686

Sanders, M. E., Westerink, J., Migchels, G., Korevaar, H., Geerts, R. H. E. M., Bloem, J., Alebeek, F. A. N. van, Schotman, A. G. M., Melman, T. C. P., Plomp, M., Muskens, G. J. D. M., & Och, R. A. F. van. (2015). *Op weg naar een natuurinclusieve duurzame landbouw* (p.). Alterra. https://library.wur.nl/WebQuery/wurpubs/492292

Sen, R., Goswami, S., & Chakraborty, B. (2019). Jeffries-Matusita distance as a tool for feature selection. *2019 International Conference on Data Science and Engineering (ICDSE)*, 15–20. https://doi.org/10.1109/ICDSE47409.2019.8971800

Shi, Y., Wang, T., Skidmore, A. K., Holzwarth, S., Heiden, U., & Heurich, M. (2021). Mapping individual silver fir trees using hyperspectral and LiDAR data in a Central European mixed forest. *International*

Journal of Applied Earth Observation and Geoinformation, *98*, 102311. https://doi.org/10.1016/j.jag.2021.102311

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Song, G., & Wang, Q. (2023). Species classification from hyperspectral leaf information using machine learning approaches. *Ecological Informatics*, *76*, 102141. https://doi.org/10.1016/j.ecoinf.2023.102141

Taşkın, G., Kaya, H., & Bruzzone, L. (2017). Feature Selection Based on High Dimensional Model Representation for Hyperspectral Images. *IEEE Transactions on Image Processing*, *26*(6), 2918–2928. https://doi.org/10.1109/TIP.2017.2687128

Thomas, F. M., Vesk, P. A., & Hauser, C. E. (2019). Field Data Collection Can Be Expensive, Time Consuming, and Difficult; Particularly in Remote or Hard To Get To Places Like 'The Mallee' in Semi-Arid Australia. *The Bulletin of the Ecological Society of America*, *100*(1), e01473. https://doi.org/10.1002/bes2.1473

Tian, J., Wang, L., Yin, D., Li, X., Diao, C., Gong, H., Shi, C., Menenti, M., Ge, Y., Nie, S., Ou, Y., Song, X., & Liu, X. (2020). Development of spectral-phenological features for deep learning to understand Spartina alterniflora invasion. *Remote Sensing of Environment*, *242*, 111745. https://doi.org/10.1016/j.rse.2020.111745

Tudi, M., Daniel Ruan, H., Wang, L., Lyu, J., Sadler, R., Connell, D., Chu, C., & Phung, D. T. (2021). Agriculture Development, Pesticide Application and Its Impact on the Environment. *International Journal of Environmental Research and Public Health*, *18*(3), Article 3. https://doi.org/10.3390/ijerph18031112

Uddin, Md. P., Mamun, Md. A., & Hossain, Md. A. (2021). PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification. *IETE Technical Review*, *38*(4), 377–396. https://doi.org/10.1080/02564602.2020.1740615

van Grinsven, H. J. M., van Eerdt, M. M., Westhoek, H., & Kruitwagen, S. (2019). Benchmarking Eco-Efficiency and Footprints of Dutch Agriculture in European Context and Implications for Policies for Climate and Environment. *Frontiers in Sustainable Food Systems*, *3*. https://doi.org/10.3389/fsufs.2019.00013

Veluw, K. van. (2013, September). *eerste voedselbos in Nederland: Nieuw perspectief voor duurzame voedselproductie | Groenekennis*. https://library.wur.nl/WebQuery/groenekennis/2101477

Villaescusa-Nadal, J. L., Franch, B., Roger, J.-C., Vermote, E. F., Skakun, S., & Justice, C. (2019). Spectral Adjustment Model's Analysis and Application to Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(3), 961–972. https://doi.org/10.1109/JSTARS.2018.2890068

Wang, H., Seaborn, T., Wang, Z., Caudill, C. C., & Link, T. E. (2021). Modeling tree canopy height using machine learning over mixed vegetation landscapes. *International Journal of Applied Earth Observation and Geoinformation*, *101*, 102353. https://doi.org/10.1016/j.jag.2021.102353

Wang, Z., & Liu, L. (2016). Correcting Bidirectional Effect for Multiple-Flightline Aerial Images Using a Semiempirical Kernel-Based Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(9), 4450–4463. https://doi.org/10.1109/JSTARS.2016.2597855

Wendel, B., Rooduijn, B., & Disselhorst, E. (2023, June). *Voedselbossen: Bodem, biodiversiteit, biomassa, business & beweging*. Monitoringvoedselbossen. https://www.monitoringvoedselbossen.nl/

Wicaksono, P., & Aryaguna, P. A. (2020a). Analyses of inter-class spectral separability and classification accuracy of benthic habitat mapping using multispectral image. *Remote Sensing Applications: Society and Environment*, *19*, 100335. https://doi.org/10.1016/j.rsase.2020.100335

Wicaksono, P., & Aryaguna, P. A. (2020b). Analyses of inter-class spectral separability and classification accuracy of benthic habitat mapping using multispectral image. *Remote Sensing Applications: Society and Environment*, *19*, 100335. https://doi.org/10.1016/j.rsase.2020.100335

Zeng, T., Wang, Y., Yang, Y., Liang, Q., Fang, J., Li, Y., Zhang, H., Fu, W., Wang, J., & Zhang, X. (2024). Early detection of rubber tree powdery mildew using UAV-based hyperspectral imagery and deep learning. *Computers and Electronics in Agriculture*, *220*, 108909. https://doi.org/10.1016/j.compag.2024.108909

Appendix 1: Woody species recorded in Ketelbroek.

Table 15: Comprehensive list of all woody species recorded in Ketelbroek during fieldwork and desktop analysis. The table includes scientific and Dutch names, representing > 200 species. This list shows the ecological diversity underlying the classification challenge.

Fruit bearing woody species

		Scientific name	**Dutch name**
1		Acer saccharum	Suikeresdoorn
2		Actinidia arguta	Kiwibes
3		Actinidia deliciosi	Kiwi
4		Actinidia kolomikta	Straalstempelkiwi
5		Akebia quinata	Schijnaugurk
6		Akebia trifoliata	Driebladig schijnaugurk
7		Amelanchier spp.	Krentenboompje
8		Arália elata	Duivelswandelstok
9		Araucaria araucana	Slangenden
10		Arbutus unedo	Aardbeienboom
11		Aronia prunifolia	Appelbes
12		Asimina triloba	Pawpaw
13		Atriplex canescens	Struikzoutmelde
14		Broussonetia kazinoki	Japanse papierboom
15		Broussonetia papyrifera	Papiermoerbei
16		Camelia sinensis	Camelia sinensis
17		Caragana arborescens	Siberische erwtenstruik
18		Carya illinoinensis	Pecan
19		Carya illinoinensis x laciniosa	Hican
20		Carya illinoinensis x ovata	Hican
21		Carya laciniosa	Koningsnoot
22		Carya ovata	Hickory bitternoot
23		Carya ovata x cordiformis	Hickory hybride
24		Carya ovata x laciniosa	Hickory hybride
25		Castanea henryi	Parelkastanje
26		Castanea mollissima	Chinese tamme kastanje
27		Castanea mollissima x sativa	Tamme kastanje hybride
28		Castanea sativa	Tamme kastanje
29		Castanea sativa x crenata	Tamme kastanje hybride
30		Celtis australis	Europese netelboom
31		Celtis occidentalis	Zwepenboom
32		Cephalotaxus harringtonia	Knoptaxus
33		Cercis siliquastrum	Judasboom
34		Citrus trifoliata	Winterharde citroen
35		Chaenomeles Japonica	Japanse sierkwee
36		Chaenomeles cathayensis	Chinese kwee
37		cornus alba	Witte kornoelje
38		Cornus kousa	Chinese kornoelje
39		Cornus kousa x capitata	Porlock
40		Cornus mas	Gele kornoelje
41		Cornus officinalis	Japanse kornoelje
42		Corylus avellana	Hazelaar
43		Corylus sieboldiana	Japanse hazelnoot
44		Corylus x colurnoides	Trazel
45		Crataegus azarolus	azarooldoorn
46		Crataegus ellwangeriana	Amerikaanse vruchtmeidoorn
47		Crataegus mexicana	Mexicaanse vruchtmeidoorn
48		Crataegus mollis	Canadese vruchtmeidoorn
49		Crataegus pinnatifida major	Chinese vruchtmeidoorn
50		Crataegus tanacetifolia	Turkese vruchtmeidoorn
51		Crataegomespilus dardarii d'Asnieresii	Asnieresii meidoornmispel

		(Crataegus laevigata x Mespilus germanica)	
52		Crataegomespilus grandiflora (Crataegus laevigata x Mespilus germanica)	Grootbloemige meidoornmispel
53		Cudrania Tricuspidata	Che fruit
54		Cydonia oblonga	Kweepeer
55		Decaisnea fargesii	Augurkenstruik
56		Diospyros **kaki**	Kaki
57		Diospyros spp (D. virginiana x kaki)	Hybride kaki
58		Diospyros virginiana	Amerikaanse persimoen
59		Elaeagnus pungens	Stekelige olijfwilg
60		Elaeagnus multiflora	Langstelige olijfwilg
61		Elaeagnus umbellata	Schermbloemige olijfwilg
62		Elaeagnus x ebbingei	Zilverbes
63		Elaeagnus x quicksilver (E. commutata x angustifolia)	Quicksilver olijfwilg
64		Ficus carica	Vijg
65		Ginkgo biloba	Japanse notenboom
66		Gleditsia triacanthos	Valse christusboom
67		Halesia carolina	Sneeuwklokjesboom
68		Hibiscus syriacus	Hibiscus (Altheastruik)
69		Hippophae rhamnoïdes	Duindoorn
70		Juglans ailanthifolia	Japanse walnoot (hartnoot)
71		Juglans cinerea	Witte walnoot
72		Juglans nigra	Zwarte walnoot
73		Juglans regia	Walnoot
74		Lonicera caerulea	Honingbes
75		Malus domestica	Appel
76		Mespilus germanica	Mispel
77		Morus alba x rubra ‘Capsrum’	Capsrum-moerbei
78		Morus nigra	Zwarte moerbei
79		Myrica gale	Wilde Gagel
80		Myrica pensylvanica	Wasgagel
81		Oemleria cerasiformis	Indianenpruim
82		Paulownia tomentosa	Anna Paulownaboom
83		Phyllostachys spp	Reuzenbamboe
84		Pinus koraiensis	Koreaanse den
85		Prunus armeniaca	Abrikoos
86		Prunus avium	Zoete kers
87		Prunus cerasifera	Kroosjespruim
88		Prunus cerasus	Zure kers (morel)
89		Prunus domestica	Pruim
90		Prunus dulcis	Amandel
91		Prunis persica	Perzik
92		Prunus salicina	Japanse pruim
93		Prunus spinosa	Sleedoorn
94		Prunus tomentosa	Nanking cherry
95		Prunus spp	Complexe pruimen
96		Pyrus communis	Peer
97		Pyrus communis x pyrifolia	Champagnepeer
98		Pyrus pyrifolia	Aziatische zandpeer (nashi)
99		Pyrus x Sinkiangensis (P. bretschneideri x communis)	Xinjiangpeer
100		Quercus ilex	Steeneik
101		Rhus aromotica	Welriekende sumac
102		Rhus glabra	Fluweelboom
103		Rhus typhina	Fluweelboom
104		Ribes divaricatum	Worcesterbes
105		Ribes nigrum	Zwarte bes
106		Ribes odoratum	Buffelkrent
107		Ribes rubrum	Rode bes
108		Ribes uva-crispa	Kruisbes
109		Ribes × nidigrolaria (R. nigrum × R. uva-crispa)	Jostabes

110		Rosa rugosa	Rimpelroos
111		Rubus spp	Braam
112			Tayberrie
113			Chinese klimbraam
114			Framboos
115		Rubus iraneus	Moeraskruipframboos
116		Rubus Nepalensis	Himalya kruipframboos
117		Rubus parviflorus x Rubus idaeus	Dorman Red
118		Rubus phoenicolasius	Japanse wijnbes
119		Rubus sechuaensis	Zsechuansisbraam
120		Rubus tricolor	Kruipframboos
121		Sambucus canadensis	Canadese vlier
122		Sambucus nigra	Vlier
123		Sassafras albidum	Sassafras
124		Schisandra chinensis	Vijfsmakenbes
125		Shepherdia argentea	Buffelbes
126		Sorbocrataegus 'Ivan's Belle' (Sorbus aucuparia x Crataegus laevigata)	Ivan's Belle
127		Sorbopyrus auricularis	Peerlijsterbes
128		Sorbus aucuparia x Crataegus sanguinea	Granatjana lijstebes
129		Sorbus devoniensis	Devon meelbes
130		Sorbus domestica	Peervormige lijsterbes
131		Sorbus torminalis	Elsbes
132		Sorbus X ‘Burka’	Burka-lijsterbes
133		Staphylea bumalda	Japanse pimpernoot
134		Staphylea colchica	Kaukasische pimpernoot
135		Staphylea pinnata	Pimpernoot
136		Stauntonia hexaphylla	‘aardappelpruim’
137		Tilia cordata	Winterlinde
138		Tilia platyphyllos	Zomerlinde
139		Tilia x europaea	Hollandse Linde
140		Toona sinensis	Uiensoepboom
141		Torreya californica	Californische nootmuskaattaxus
142		Torreya nucifera	Kaya
143		Ulmus laevis	Fladderiep
144		Viburnum dentatum	Sneeuwbal dentatum
145		Viburnum furcatum	Sneeuwbal furcatum
146		Viburnum lentago	Schapenbes
147		Viburnum nudum	Sneeuwbal
148		Viburnum plicatum	Japanse sneeuwbal
149		Viburnum prunifolium	Zwarte haagdoorn
150		Vitis vinifera	Druif
151		Wisteria sinensis	Blauwe regen
152		Xanthoceras sorbifolium	Chinese bloeiende kastanje
153		X Pyrocydonia danielii	Pyrocydonia danielii
154		X Pyronia veitchii	Pyronia veitchii
155		Zanthoxylum alatum	Nepalese peper
156		Zanthoxylum bungeanum	Peperboom bungeanum
157		Zanthoxylum giraldii	Peperboom giraldii
158		Zanthoxylum schinifolium	Szechuanpeper schinifolium
159		Zanthoxylum simulans	Szechuanpeper
160		Ziziphus jujuba	Chinese dadelpruim
Supporting wooden species			
161		Acer campestre	Veldesdoorn
162		Acer pseudoplatanus	Gewone esdoorn
163		Alnus cordata	Italiaanse els
164		Alnus glutinosa	Zwarte els
165		Alnus incana	Witte els
166		Alnus rubra	Rode els

167		Betula pendula	Ruwe berk
168		Betula pubescens	Zachte berk
169		Carpinus betulus	Haagbeuk
170		Clematis vitalba	Clematis vitalba
171		Cornus sanguinea	Rode kornoelje
172		Crataegus laevigata	Tweestijlige meidoorn
173		Euonymus europaeus	Kardinaalsmuts
174		Fagus sylvatica	Gewone beuk
175		Fraxinus excelsior	Gewone es
176		Hedera helix	Hedera helix
177		Ligustrum vulgare	Liguster
178		Lonicera periclymenum	Kamperfoelie
179		Malus sylvestris	Wilde appel
180		Pseudotsuga menziesii	Douglasspar
181		Populus deltoides	Balsempopulier
182		Populus nigra	Zwarte populier
183		Populus × Canadensis	Canadapopulier
184		Populus tremula	Ratelpopulier
185		Prunus padus	Vogelkers
186		Pyrus pyraster	Wilde peer
187		Quercus petreae	Wintereik
188		Quercus robur	Zomereik
191		Rhamnus cathartica	Wegedoorn
192		Robinia pseudoacacia	Acacia
193		Rosa canina	Hondsroos
194		Rosa corymbifera	Heggenroos
195		Rosa rubiginosa	Egelantier
196		Rosa villosa	Viltroos
197		Salix alba	Schietwilg
198		Salix caprea	Boswilg
199		Salix fragilis	Kraakwilg
200		Salix spp	Overige wilgen
201		Sorbus aucuparia	Lijsterbes
202		Styrax officinalis	Storax
203		Taxus baccata	Venijnboom
204		Viburnum opulus	Gelderse roos
205		Pinus sylvestris	Grove den
206		Salix aurita	Geoorde wild
207		Prunus cerasifera	Paarse kroosjespruim
208		Craetegus viridus	Groene meidoorn
209		Prunus armeniaca x salicina	Abrikoos met japanse pruim
210		Berberis coreana	Koreaanse zuurbes
211		Maclura pomifera	Osagedoorn
212		Corylus avellana	Roodbladige hazelaar
213		Elaegnus angustifolia	Olijfwilg
214		Rubus subg batohamnus	Prachtbraam
215		Urtica doica	Grote brandnetel
216		Berberis coreana	Koreaanse zuurbes
217		Maclura pomifera	Osagedoorn
219		Elaeagnus angustifolia	Olijfwilg
220		Rubus subg batothamnus	Prachtbraam
300		Urtica doica	Grote brandnetel

As noted on 28 May 2024 in food forest Ketelbroek.

Appendix 2: Spatial sampling distribution.

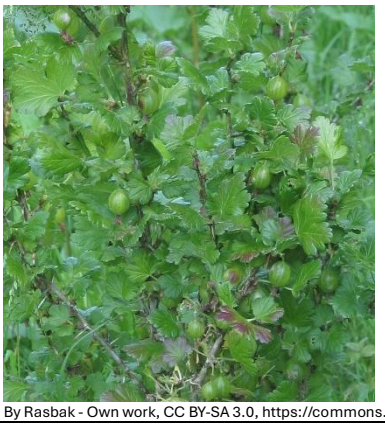
Figure 20: Spatial distribution map of reference points collected during fieldwork and desktop analysis. This figure illustrates spatial coverage and sample density used for species classification.

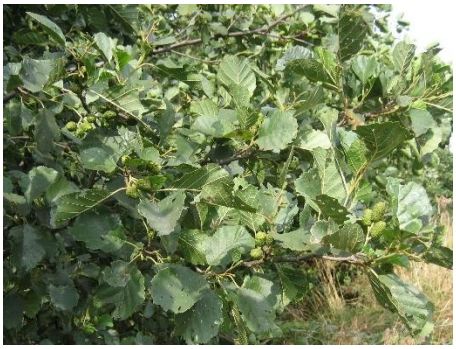
Appendix 3: Photographic species overview.

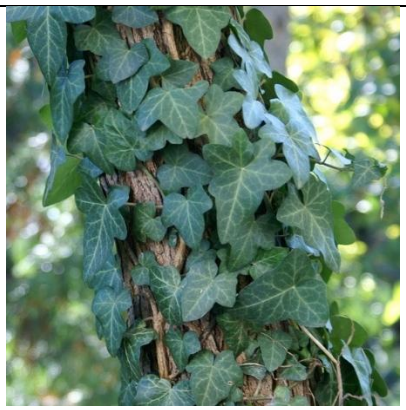
Table 16: Photo-illustrated overview of representative species included in the study. Images highlight key phenological traits relevant for spectral interpretation and classification.

Name	Photo
Chocolate vine - Akebia Quinata	 CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=120155

Sweet Chestnut - Castanae Sativa	 Flowering Sweet chestnut tree (Castanea sativa) by Evelyn Simak, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=144664100
Trifoliate Orange - Citrus Trifoliata	 By Daderot - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=170444637
Hazel - Corylus Avellana	 By AudreyMuratet - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=96537312
Thorny Olive - Elaeagnus pungens	 By KENPEI - KENPEI's photo, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1308280
Autumn Olive - Elaeagnus Umbellata	 By R. A. Nonenmacher - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=107094261

Gooseberry - Ribes Uva-Crispa	 By Rasbak - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=848387
Beach Rose - Rosa Rugosa	 By Robert Flogaus-Faust - Own work, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=148153167
Black Elder - Sambucus Nigra	 By Sambucus nigra by Bob Harvey, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=158447805
Bollwiller Pear - Pyrus Communis 'Bollwiller'	 By Abraham - Own work, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=149833344
European Linden - Acer Europaeus	

	 By Chris Light - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=84850391
Common Alder - Alnus Glutinosa	 Own work assumed (based on copyright claims)., CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=430484
Grey Alder - Alnus Incana	 By Sten Porse - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=459811
European Spindle - Euonymus Europaeus	 By R. A. Nonenmacher - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=36481774
Common Beech - Fagus Sylvatica	 By Tournasol7 - Own work, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=169787673
Common Ivy - Hedera Helix	

	 By chery - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=1282155
Douglas Fir - Pseudotsuga Menziesii	 CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=907873
Quacking Aspen - Populus Tremula	 By Rudolphous - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=91904378
Oak - Quercus Robur	 By Rudolphous - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=91905062
Black Locust - Robinia Pseudoacacia	

	 By Mark14 - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=7207386
White Willow - Salix Alba	 By Josep Gesti - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=116937189
Crack Willow - Salix Fragilis	 By Krzysztof Ziarnek, Kenraiz - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=99427373 ‘
Yew - Taxus Baccata	 By Original uploader was Kpjas at pl.wikipedia - CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1394796

Appendix 4: PCA outputs.

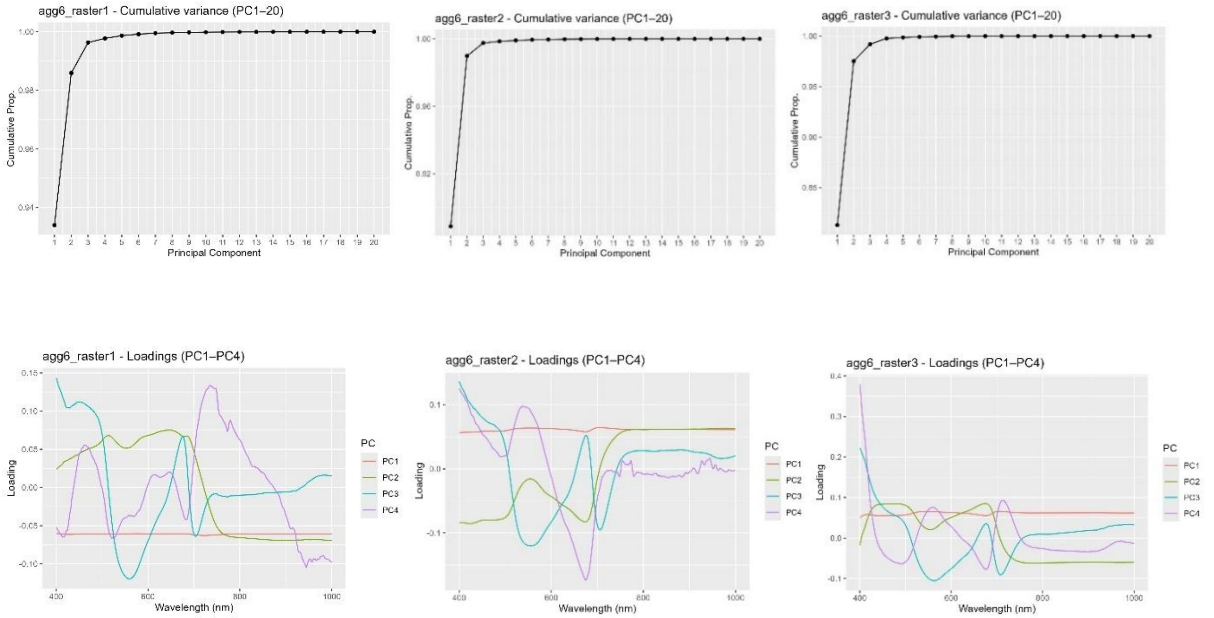


Figure 21: Principal Component Analysis results showing eigenvalues and loadings across all spectral bands. This figure illustrated how most spectral variance is captured by the first four PCs (top), and how the loadings is divided within the PCs (bottom).

Appendix 5: CHM height statistics.

Table 17: Mean canopy heights for species in Raster 1 based on CHM measurements. Values reflect varying crown structures and vertical stratification.

Name	**Veg type**	**Mean height (M)**	**N**
Quacking Aspen	tree	13.2	29
Quercus robur	Tree	13.1	45
Grey Alder	Tree	8.9	11
Crack Willow	Tree	8.8	25
Douglas Fir	Tree	7.8	26
Common Ivy	Vine	7.5	15
Black Locust	Tree	6.6	11
Hazel	Tree	6.5	12
Chocolate Vine	Vine	5.5	11
Taxus baccata	Shrub	3.3	11

Table 18: Mean canopy heights for species in Raster 2 based on CHM measurements. Values reflect varying crown structures and vertical stratification.

Name	**Veg type**	**Mean height (m)**	**N**
White Willow	Tree	12.7	28
Quaking Aspen	Tree	12.3	15
Black Alder	Tree	11.6	11
Crack Willow	Tree	9.9	16
Common Beech	Tree	9.8	10
Grey Alder	Tree	9.7	21
Oak	Tree	9.1	10
European Linden	Tree	7.7	15
Douglas Fir	Tree	7.6	18
Hazel	Tree	7.4	18
Hedera helix	Vine	7.3	15
Chocolate Vine	Vine	6.7	21
Black Locust	Tree	6.6	11
European Spindle	Tree	4.8	11
Yew	Shrub	3.4	11
Thorny Olive	Shrub	3.1	10

Table 19: Mean canopy heights for species in Raster 3 based on CHM measurements. Values reflect varying crown structures and vertical stratification.

Name	**Veg Type**	**Mean height (m)**	**N**
Bollwiller Pear	Tree	8.0	12
Field Maple	Tree	5.6	21
Black Elder	Tree	3.8	28
Autumn Olive	Shrub	3.5	11
Thorny Olive	Shrub	3.4	10
Gooseberry	Shrub	2.4	27
Trifoliate Orange	shrub	2.5	12
Beach Rose	shrub	2.0	38

Appendix 6: Linear regression plots. Pair 2.

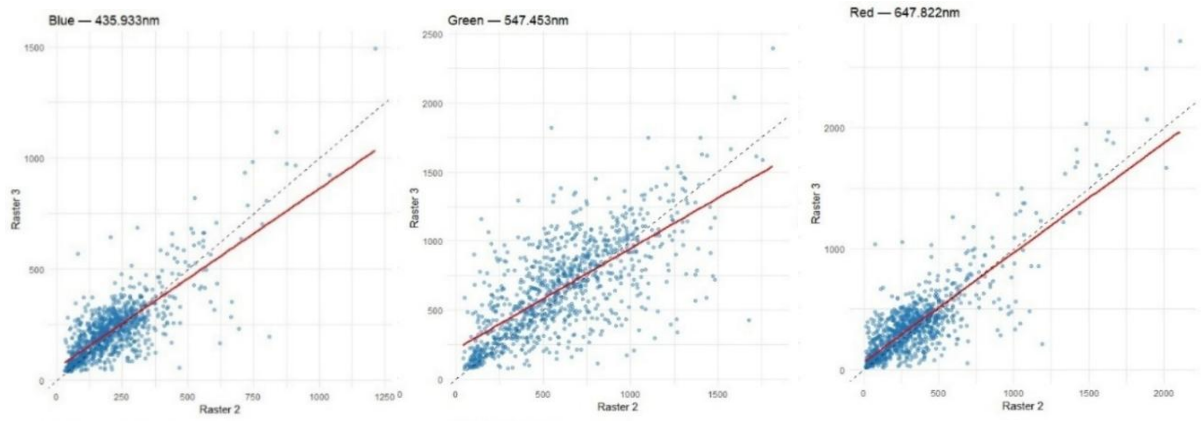


Figure 22: Linear regressions of Raster 2 vs Raster 3 reflectace for the Blue, Green and red bands. Deviations from the 1:1 line indicate inconsistent radiometry between flightlines.

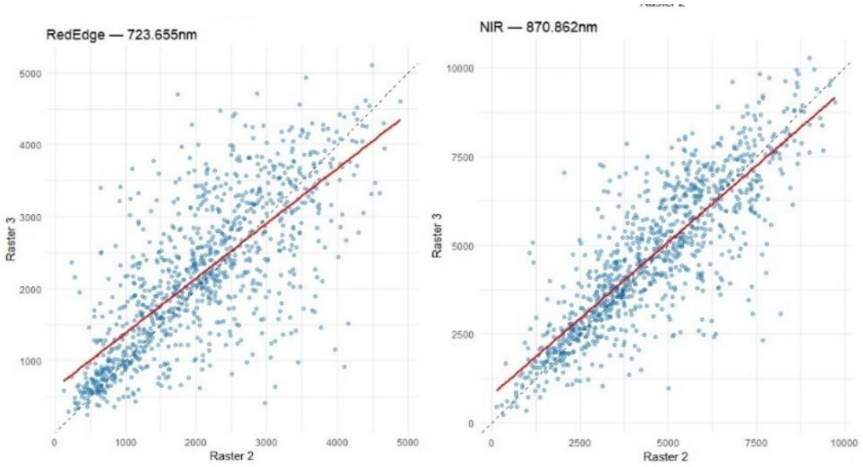


Figure 23: Linear regressions of Raster 2 vs Raster 3 reflectace for the Red Edge and NIR bands. Deviations from the 1:1 line indicate inconsistent radiometry between flightlines.

Appendix 7: Intercept, Slope, R2 and RMSE. Pair 2.

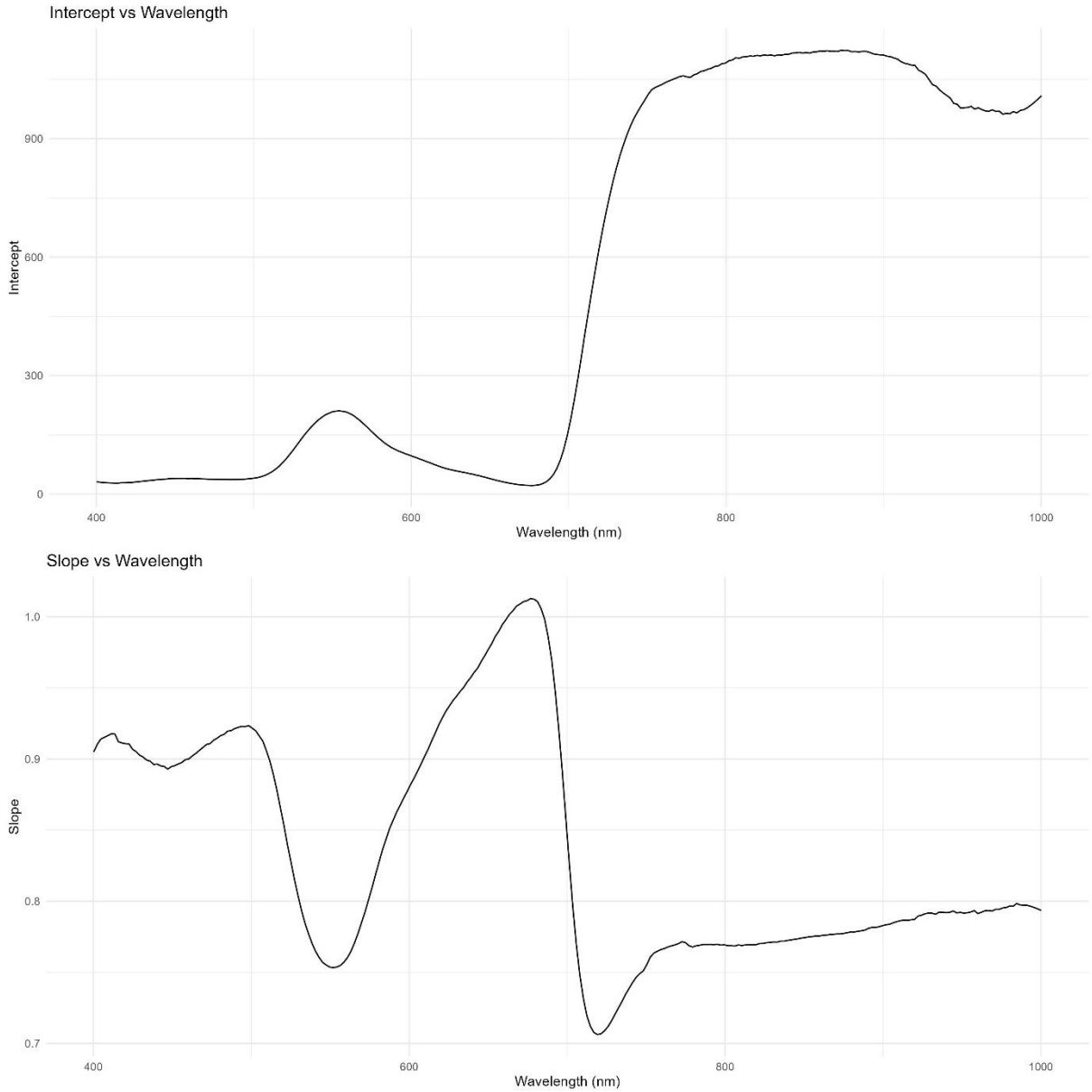


Figure 24: Intercept and Slope parameters for pair 2 of per-band regression across 400 – 1000nm. Sharp changes in the Red Edge and NIR reveal wavelength dependent scaling effects between strips.

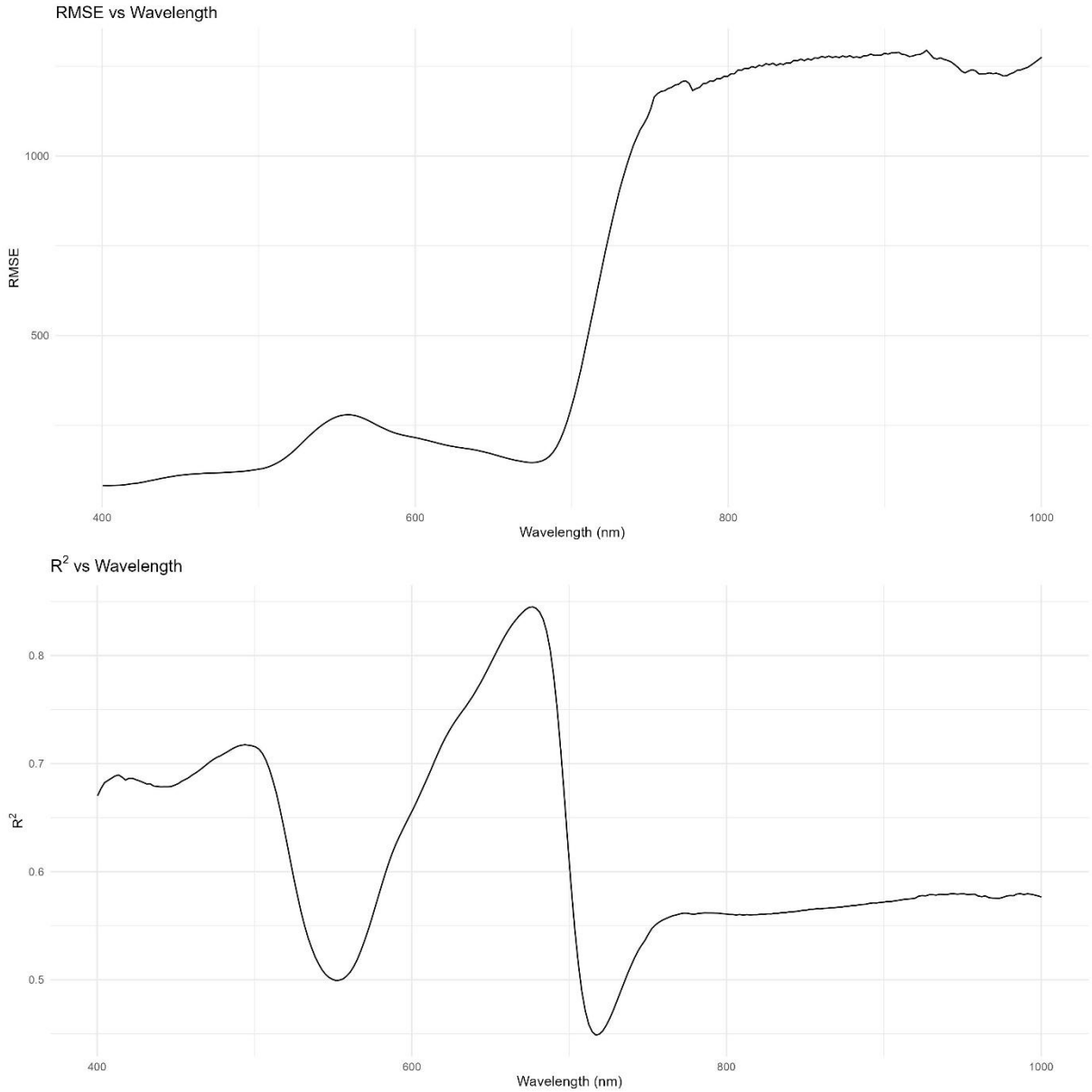


Figure 25: R^2 and RMSE parameters for pair 2 of per band regression across 400 – 1000nm. Goodness of fit metric emphasises inconsistent regression performance and large NIR errors that justify keeping flightlines separate.

Appendix 8: Overall accuracy PCA vs PCA + CHM, for all aggregations and rasters.

Table 20: Overall accuracy for PCA-only and PCA + CHM models across all six aggregation levels for each raster. Results demonstrate strong resolution dependence and CHM benefits.

Raster	Agg	Accuracy	Accuracy + CHM	Abs. Diff	Rel. Diff (%)
1	1	0.430	0.608	0.177	41.18
	2	0.456	0.646	0.190	41.67
	3	0.506	0.650	0.114	22.50
	4	0.584	0.779	0195	33.33
	5	0.705	0.795	0.090	12.73
	6	0.662	0.779	0.117	17.65
2	1	0.410	0.552	0.143	34.88
	2	0.514	0.648	0.133	25.93
	3	0.510	0.630	0.120	23.53
	4	0.556	0.657	0.101	18.18
	5	0.673	0.755	0.082	12.12
	6	0.687	0.697	0.010	1.47
3	1	0.609	0.875	0.266	43.59
	2	0.656	0.891	0.234	35.71
	3	0.688	0.906	0.219	31.82
	4	0.734	0.891	0.156	21.28
	5	0.672	0.891	0.219	32.56
	6	0.683	0.817	0.133	19.51

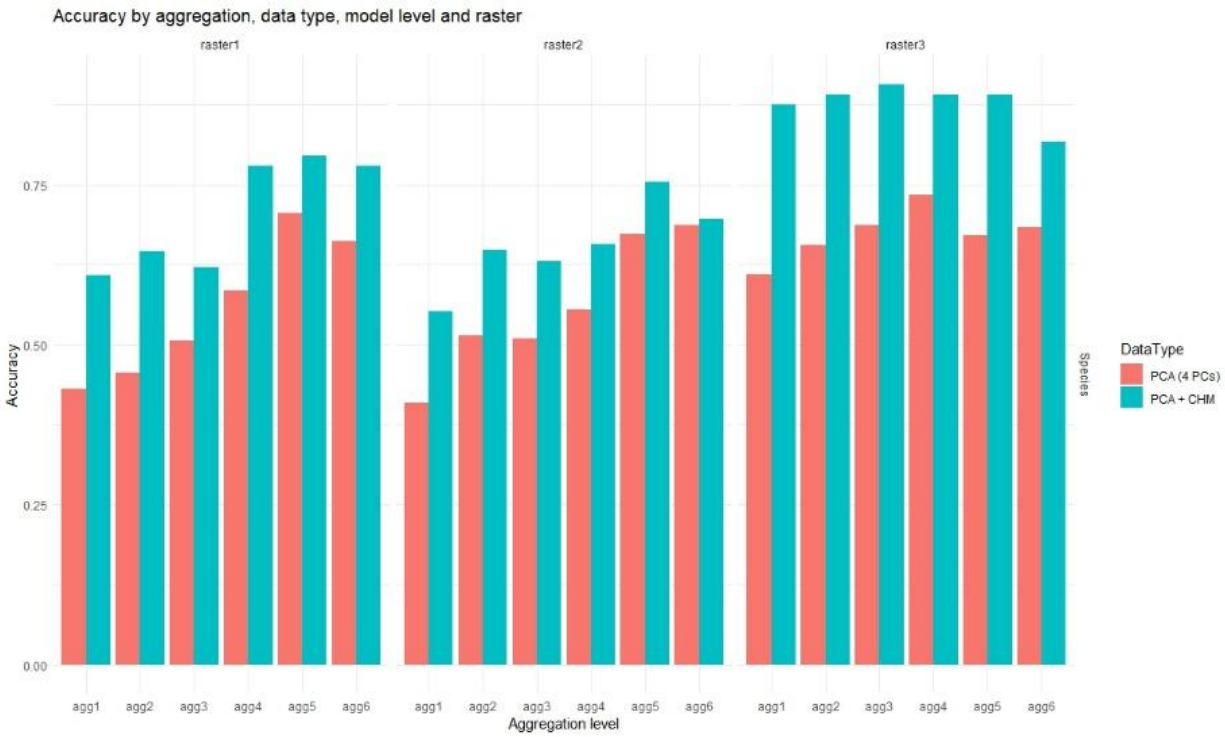


Figure 26: Visualisation of accuracy trends across all aggregation levels for all rasters. The figure shows improved performance when including CHM.