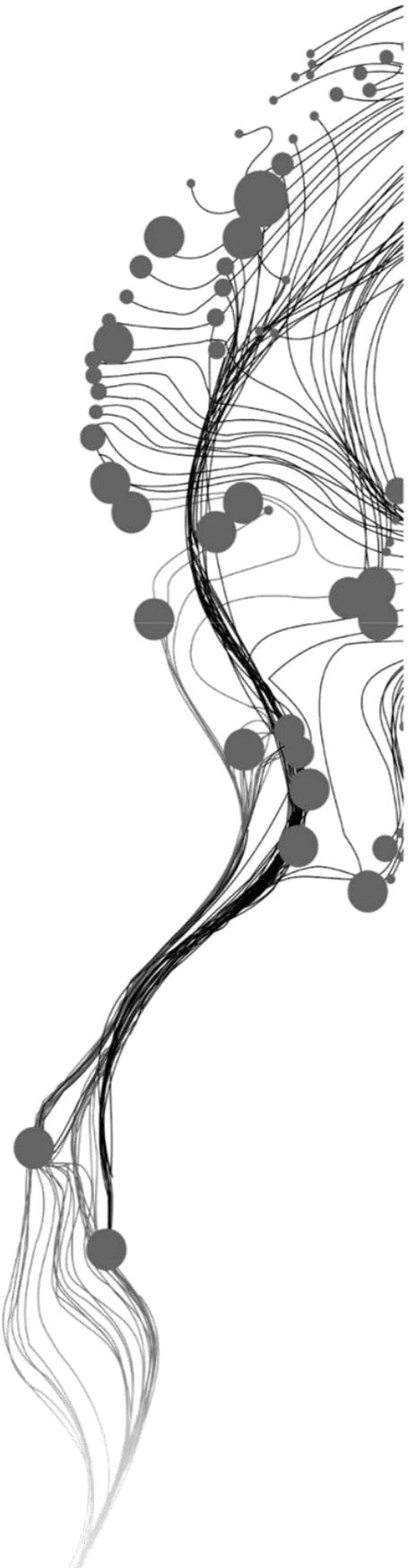# CROP TYPE DISCRIMINATION USING FIELD AND SATELLITE HYPERSPECTRAL MEASUREMENTS IN BUSIA, KENYA

Kelvin Aslen Mlawa
June 2023

SUPERVISORS:

Dr. Roshanak Darvishzadeh
Prof. Dr. Andy Nelson

CROP TYPE DISCRIMINATION USING FIELD AND SATELLITE HYPERSPECTRAL MEASUREMENTS IN BUSIA, KENYA

KELVIN ASLEN MLAWA
Enschede, The Netherlands,
June, 2023

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:

Dr. Roshanak Darvishzadeh
Prof. Dr. Andy Nelson

THESIS ASSESSMENT BOARD:
Dr.ir. Vrieling, Anton (Chair)
Prof. Abel Ramoelo (External Examiner)

ABSTRACT

Crop type discrimination is essential for effective agricultural monitoring and management, enabling better resource allocation and yield prediction. This study investigated the potential of hyperspectral remote sensing data to discriminate between maize and cassava crops in Busia County, Kenya, at both the leaf (using layer stack) and satellite levels. Through fieldwork, comprehensive analysis of spectral signatures, and the identification of significant bands, this research aimed to provide valuable insights into hyperspectral-based crop discrimination.

The study used field measurements and satellite data. The field measurements included field boundary coordinates and spectral measurements of cassava and maize taken at the leaf layer stack level using an ASD FieldSpec3 spectroradiometer during the field visit from 11 July to 4 August 2022. The PRISMA satellite data acquired on 1 July 2022 was used to extract the mean spectral reflectance of crop fields. The Mann-Whitney U test was used to determine whether there were significant differences between cassava and maize and in which bands/ regions these differences existed. Continuum removal and band depth analysis were then used to better understand the differences in the absorption features of cassava and maize spectral reflectance at the leaf and satellite levels. Finally, Random Forest Classification was used to discriminate between crop types at both leaf and satellite levels.

The results showed that there were significant differences between the spectral reflectance of cassava and maize at the leaf level that appeared to exist in several wavelength regions (385-701nm, 709-1346nm, 1455-1658nm, 2070-2115nm, and 2237-2254nm). Continuum removal and band depth analysis along the absorption regions helped further understand the differences between the two crops. At the field level, analysis of the spectral signatures showed clear differences between cassava and maize leaf spectral reflectance in the visible, NIR, and SWIR along the bands with the highest absorption peaks at 496, 677, 679, 1162, 1459, and 1460 nm. Similarly, clear differences between cassava and maize were observed at the satellite level in the bands with the highest absorption peaks at 470, 650, and 1163 nm. These differences were most likely due to biophysical and biochemical differences between cassava and maize. By considering only the bands with the highest absorption peak and using random forest classification, this study successfully discriminated between cassava and maize at the leaf level with an overall accuracy of 94% and a kappa score of 0.89. Additionally, at the satellite level, this study was able to discriminate maize from the non-maize class with an overall accuracy of 77% and a kappa score of 0.54. These results provide valuable insights into the hyperspectral bands that could effectively discriminate between cassava and maize and further demonstrate the usefulness of hyperspectral data in crop discrimination.

Further testing of crop discrimination methods using only the bands with the highest absorption peaks is recommended when considering more crop types. This may help to refine the methodology and improve crop type discrimination using hyperspectral data in future studies.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Introduction and background

Africa's present population is expected to quadruple by 2050 (United Nations, 2019). As a result, the demand for grains in 2050 is expected to be more than three times that of 2010 (Van Ittersum et al., 2016). Providing adequate food for the continent's population will be a challenge, as present food production rates are two to three times lower than what is required to sustain food security (FAO, 2021; Maja & Ayano, 2021). A stable and reliable farming management system is required to increase farming yield and achieve food security. Crop type discrimination is a very important component of the farming management system. The accurate quantification of Crop type in agricultural areas is crucial for regional to global food security measures since it allows for precise acreage and yield estimation (Azar et al., 2017; Inglada et al., 2015; Ozdarici-Ok et al., 2015).

Many countries worldwide are utilizing remote sensing techniques, particularly hyperspectral remote sensing, as a promising approach for accurate crop discrimination, mapping, and classification (Thenkabail et al., 2018). Hyperspectral remote sensing has demonstrated its efficacy in a range of applications within the farming industry, including crop yield estimation (Ferencz et al., 2004), pest and disease control (Moran et al., 1997), crop health monitoring (Lichtenthaler et al., 1998), and precise crop classification (Schotten et al., 1995). By leveraging the spectral information captured by hyperspectral sensors, it becomes possible to distinguish between different crop types with high precision and details.

Over the last three decades, remote sensing's achievements in precision agriculture and farm management have emphasized the capacity to collect massive amounts of data from various sensors and platforms. Agricultural remote sensing has traditionally relied on data from multispectral broadband sensors (Thenkabail et al., 2018), and to some extent, active sensors; however, the introduction of hyperspectral remote sensing for agriculture monitoring has offered greater opportunity for detailed information extraction on Crop types, their status and characteristics.

In contrast to multispectral sensors that capture information in discrete broad spectral bands, hyperspectral sensors offer the advantage of capturing data in continuous narrow spectral bands. This enables the assessment of essential vegetation ecophysiological information in crops (Inoue et al., 2019). Spectral data measured using hyperspectral sensors provide valuable insights into physiological parameters such as chlorophyll content for crop growth monitoring, nitrogen content for effective fertilizer management. This rich and comprehensive information contained in hyperspectral data makes it the optimal choice for crop type discrimination (Blackburn & Ferwerda, 2008; Chen et al., 2008; Rao et al., 2007; Thenkabail et al., 2000, 2004).

Several studies have demonstrated the effectiveness of using hyperspectral data for crop type discrimination. Xue et al. (2017) utilised data obtained from Compact Airborne Spectrographic Imager (CASI) and Shortwave Infrared Airborne Spectrographic Imager (SASI) to discriminate various crops, including Corn, Fragrant-flowered Garlic, Cauliflower, Bell Pepper, Potato, Endive Sprout, and Watermelon, employing a sparse graph regularization (SGR) method for crop mapping. Thenkabail et al. (2004) employed a 1-nm-wide narrowband FieldSpec Pro Full Range spectroradiometer to discriminate six crops, namely corn, groundnut, rice, soybean, cowpea, and cassava, using Wilk's lambda, Pillai trace, and average squared canonical correlation methods. Nidamanuri and Zbell (2011) focused on the discrimination and mapping of alfalfa, winter barley, winter rape, winter rye (Secale cereale), and winter wheat (Triticum spp.) using field data from FieldSpec JR spectrometer and hyperspectral images captured by the HyMAP imaging system mounted on an airborne platform, employing three methods, namely mixture tuned matched filtering (MTMF), spectral feature fitting (SFF), and spectral angle mapper (SAM) methods. Another study by Dave et al. (2022a) suggested the use of a new band selection algorithm utilizing the statistical parameter, spectral information divergence (SID), comparing its performance with commonly used classifiers such as support vector machine (SVM), K-nearest neighbours, and artificial neural network (ANN), using three hyperspectral datasets, including AVIRIS-NG, Indian Pines, and Salinas. Furthermore, Sulaiman et al. (2022) reviewed the usability of

hyperspectral images for weed analysis in rice fields, discussing various supervised and unsupervised classification methods, such as support vector machines, artificial neural networks, decision trees, maximum likelihood classification, K-means classification, iterative self-organizing method (ISODATA), Laplacian support vector machine (LapSVM), and self-training. They emphasized the potential of hyperspectral images captured from UAV platforms when used in conjunction with machine learning algorithms. Moreover, Dave et al. (2022b) proposed the use of similarity measure and fuzziness (SS-SMFZ) for hyperspectral data sample selection, employing Airborne Visible near InfraRed Imaging Spectrometer-Next Generation (AVIRIS-NG) dataset and benchmark hyperspectral datasets, namely Indian Pines and Salinas, and using supervised machine learning classifiers, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF), to classify land cover classes, including reference crops such as tobacco peak vegetative, tobacco vegetative, and wheat soft dough. In a study by Spiller et al. (2021), crop type mapping of tomato and corn was performed using a one-dimensional convolutional neural network (CNN), comparing the outputs between PRecursore IperSpettrale della Missione Applicativa (PRISMA) images and Synthetic Aperture Snow Radar (SASI) and Compact Airborne Spectrographic Imager (CASI) airborne images. Additionally, Su (2020) reviewed the use of color imaging, hyperspectral imaging, and spectroscopy in the discrimination of weeds and crops, employing supervised and unsupervised machine learning algorithms. It is worth noting that most previous studies on crop discrimination have predominantly utilised airborne data, with relatively few studies focusing on field spectrometry.

Selecting the optimum spectral region in the measured crop spectrum is challenging but crucial in differentiating one crop from another (Arafat et al., 2013; Dhumal et al., 2015). It involves identifying the specific spectral bands that are most informative for distinguishing between different crop types, considering factors such as the absorption and reflectance characteristics of crops and the spectral resolution and noise levels of the hyperspectral sensor. Selecting the optimal bands for crop type discrimination in hyperspectral data presents a significant challenge due to the high-dimensional nature of the data. It requires striking a balance between capturing relevant spectral information and minimizing data redundancy.

Different crops exhibit distinct biochemical and biophysical characteristics, resulting in variations in their spectral signatures. These differences in spectral signatures enable the discrimination of different crop types when the appropriate approach is employed to capture and analyse these distinctions. Identifying and employing the most effective approach for crop discrimination, whether it involves band selection (Zhang et al., 2016), vegetation index calculation (Thenkabail et al., 2013), or other techniques such as machine learning algorithms (Singh et al., 2022), pattern recognition methods (Galvão et al., 2018), spectral feature extraction (Zhang et al., 2019), spectral angle mapping (Martin et al., 2011), spectral unmixing (Chi & Crawford, 2014), and hyperspectral image classification techniques, is crucial to ensure accurate crop discrimination (Aneece & Thenkabail, 2021; Janse & Deshmukh, 2021). By selecting the optimal approach for crop discrimination, researchers can leverage the unique spectral characteristics of crops, effectively distinguishing between them and enhancing the accuracy of crop type discrimination.

This study aimed to address a significant research gap in the field of crop type discrimination by capitalizing on the potential of new and advanced hyperspectral sensors, specifically the PRISMA satellite, in conjunction with field data. While previous studies had predominantly relied on the use of multispectral sensors for crop-related research, including crop type discrimination, there had been limited exploration of the capabilities offered by the latest hyperspectral sensors. Furthermore, although some studies had made use of existing hyperspectral sensors, the potential of the PRISMA sensor for crop type discrimination had remained unexplored. By filling this gap in the literature, this research sought to demonstrate the effectiveness of combining PRISMA's high-resolution hyperspectral data with field data for accurate and detailed discrimination of different crop types. The findings of this study contributed to advancements in crop mapping and management strategies, ultimately aiding in addressing the challenges posed by the increasing food demand resulting from Africa's population growth.

The advances in hyperspectral remote sensing and the launch of recent satellites such as Deutsches Zentrum für Luft- und Raumfahrt Earth Sensing Imaging Spectrometer (DESIS), PRISMA, and Environmental Mapping and Analysis Program (EnMAP) have created new opportunities to utilise hyperspectral satellite imagery to characterise and monitor agricultural areas. The above literature review shows that out of the small number of studies that used hyperspectral satellite data in an agricultural context, most have considered crop traits retrieval (Marshall et al., 2022; Mohammadi et al., 2023; Tagliabue et al., 2022; Verrelst et al., 2021) and none have explored its capacity for crop type discrimination.

This study will focus on two levels of crop type discrimination in Busia County, Kenya. First, at the leaf level, exploring the differences between the spectral signatures of cassava and maize at the leaf level using leaf layer stacks and then at the canopy level using recent hyperspectral satellite data (PRISMA) for discriminating maize from other crops.

Discrimination of the two crops will help in accurate crop mapping which in turn will help for better crop management.

1.2. Objectives, research questions and hypotheses

Main objective and Specific objectives
The aim of this study is to discriminate major crop types (cassava and maize) in Kenya using hyperspectral remote sensing.

Specific objectives

Objective 1: To study the spectral signatures of cassava and maize leaves and understand their variations in the field in Busia County, Kenya.

Objective 2: To explore whether the hyperspectral wavebands/regions suitable for discrimination of cassava and maize at the field level are relevant at the satellite level.

Objective 3: To discriminate maize from casava and maize from other crops in Busia County, Kenya using hyperspectral data at the leaf and satellite levels, respectively.

1.3. Research questions and hypothesis

Question 1: How the spectral signatures of cassava and maize leaves vary at the leaf level when measured from a stack layer?

Hypothesis: The spectral reflectance of cassava in NIR and SWIR regions is higher compared to that of maize difference in biophysical characteristics.

Question 2: What are the differences between hyperspectral bands suitable for discriminating cassava and maize at the field and satellite levels?

Hypothesis: Optimum hyperspectral wavebands suitable for discrimination of cassava and at the field and satellite levels are similar except those located around the water absorption peaks.

Question 3: How to discriminate maize from casava and maize fields from other crops in Busia County, Kenya, using hyperspectral measurements?

Hypothesis: Supervised machine learning classifiers like Random Forest (RF) can be used to discriminate maize from casava and maize fields from other crop fields.

2. METHODOLOGY

2.1. Study area

Located in Kenya's extreme west, Busia County has a total area of 1,694.5 km². It has boundaries with Kakamega to the east, Bungoma to the north, Lake Victoria to the south-west, Siaya to the south-east, and the Republic of Uganda to the west. It is located between 0° and 0.45° North latitude and 34.25° East longitude. The Lake Victoria Basin encompasses the majority of Busia County. The altitude is undulating and ranges from approximately 1,500m in the Samia and North Teso Hills to 1,130m above sea level on the shores of Lake Victoria. Based on census statistics from 2019, Busia had a population of about 893,681.

The yearly rainfall in Busia County ranges from 760 to 2000 mm. The long rain season, which lasts from late March to late May, accounts for 50% of the annual rainfall, while the short rain season, which lasts from August to October, accounts for 25%. From December to February, there is a dry season with occasional showers. The temperatures across the county are quite consistent. The yearly mean maximum temperature ranges from 26°C to 30°C, while the mean minimum temperature ranges from 14°C to 22°C.

The hills of Samia and Budalang'i in Busia County are covered in a natural forest, although other areas of the county feature on-farm woodlots integrated with agricultural production. Beans, sorghum, finger millet, rice, cowpea, groundnuts, sweet potato, banana, oil palm, green gram, cotton, tobacco, sugarcane, and pepper are some of the other significant crops farmed in Busia County, in addition to major crops cassava and maize. Horticultural crops are also grown, including, among others, pineapples, tomatoes, kale, cabbage, watermelons, native vegetables, mangoes, amaranth, jack fruit, onions, and papaya.
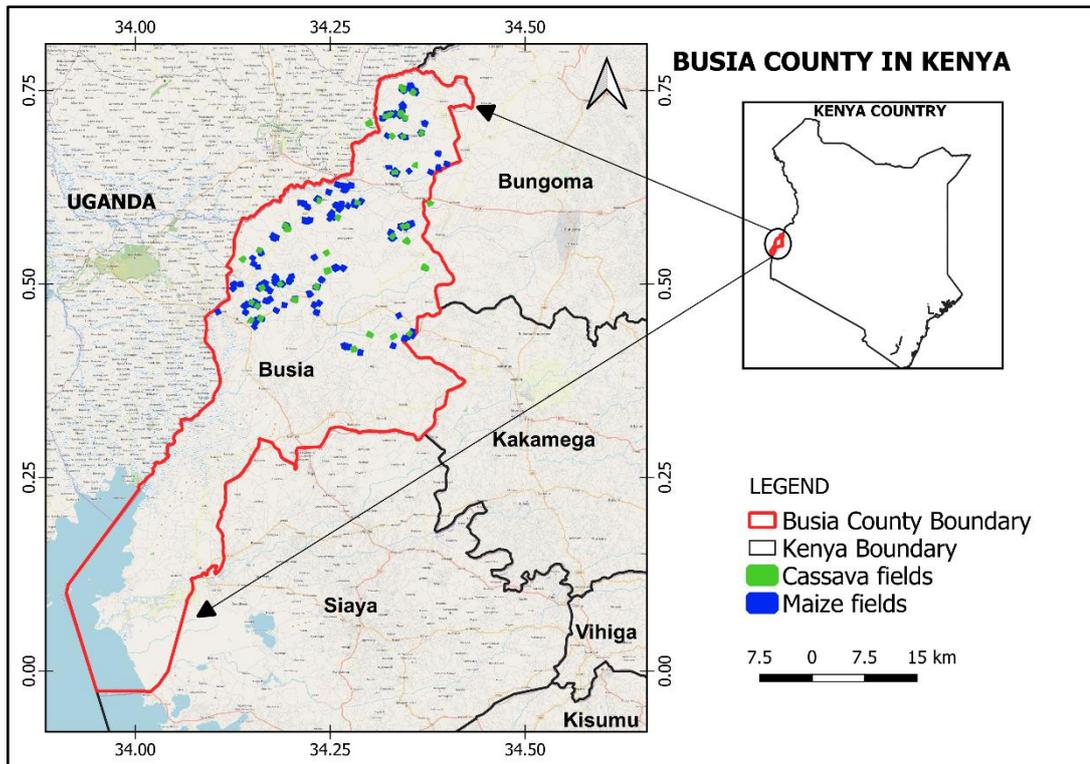
Figure 1 Study area (Busia County in Kenya) and distribution of cassava and maize fields.

2.2. Methodological flow chart

The crop type discrimination performed in this study is outlined in the flow chart (Figure 2). The major steps include field measurements, data pre-processing, and data analysis at the field and satellite levels. The flow chart also shows at which steps different objectives of this study were achieved.
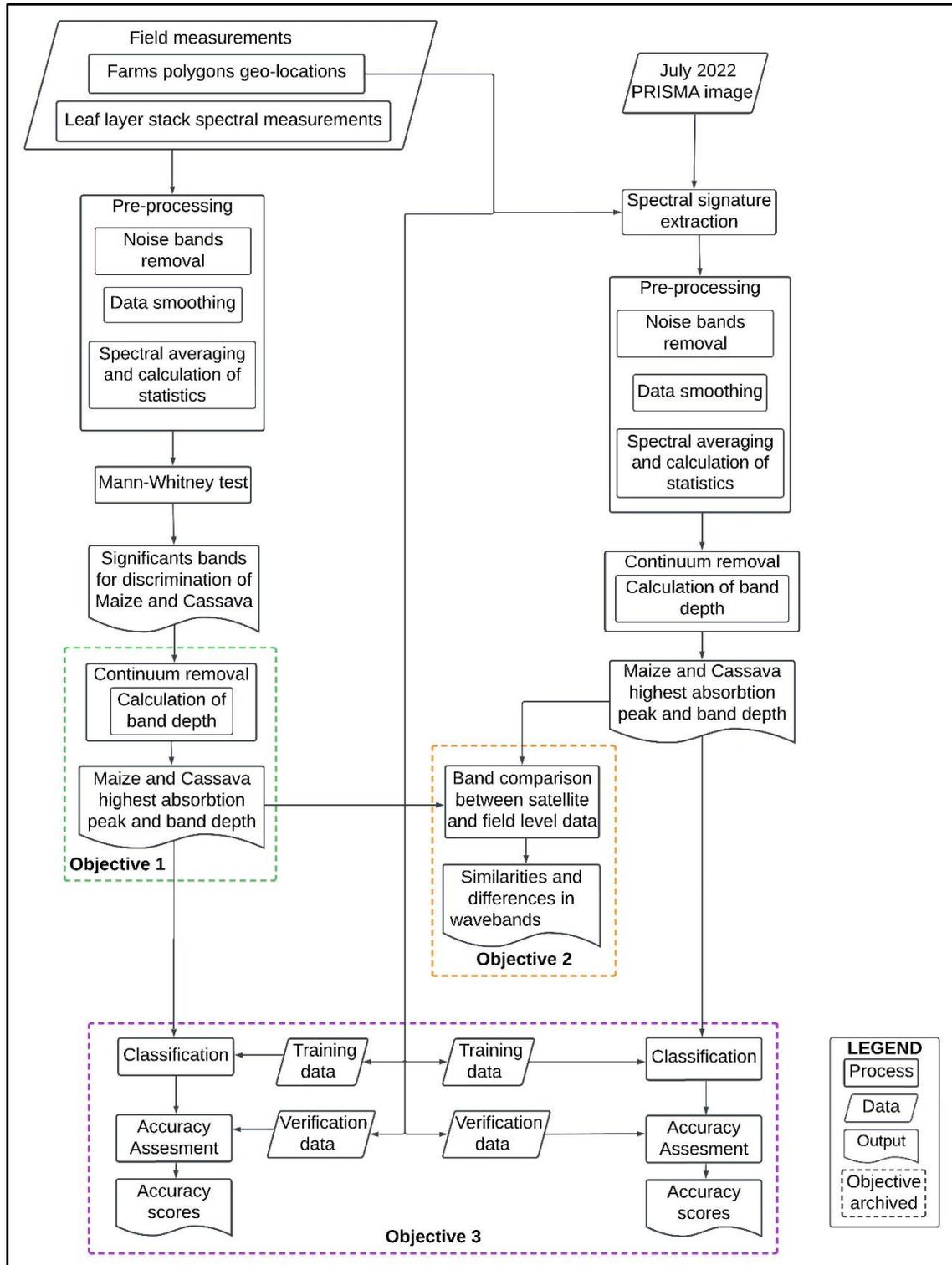
Figure 2 Methodological flowchart of the study .

2.3. Data acquisition and pre-processing
2.3.1 Field data measurement and pre-processing

Leaf samples, specifically spectral measurements at the leaf stack layer level using a spectroradiometer, were collected from a total of 30 maize fields and 31 cassava fields (Figure 3) between 11 July and 4 August, 2022. Reflectance spectra of maize and casava layer stacks were then measured in the field around noon (seen in Figure 3), with the sun being almost at the nadir position. In each layer stack, five sample measurements were taken, with ten reflectance measurements recorded for each measurement, this resulted in a total of 50 reflectance measurements saved for each layer stack, using an ASD FieldSpec3 spectroradiometer. The wavelength range of the field spectroradiometer is between 350 nm to 2500 nm, with a spectral sampling of 1.4 nm in the 350 nm to 1000 nm range and 2 nm in the 1000 nm to 2500 nm range (Huang et al., 2014). The fiber optic, with a field view of approximately 13.3 cm, was placed in a pistol and handheld approximately 30 cm above the 30 cm * 30 cm layer stack sample at the nadir position avoiding the surroundings' influence on the spectral measurements. The radiance of a standard white panel covered with $BaSO_4$ of known reflectivity was collected prior to reflectance measurements for normalisation of the target data.

Noisy bands below 400 nm, after 2475 nm and those affected by water absorption regions were removed. This resulted in the removal of 202 spectral bands, and the remaining 1949 wavebands were used for further analysis. The details of the removed bands can be found in Appendix I.

The fifty reflectance measurements taken from each layer stack sample were averaged to represent one sample. By averaging the fifty saved reflectance measurements made from each layer stack sample, it was possible to reduce the measurement noise. To further smooth the spectra, a moving Savitzky-Golay filter (Savitzky and Golay,1964) with a frame size of eleven data points (1st-degree polynomial) was used to smooth the reflectance spectra using MATLAB R2022a (MathWorks, Inc.), the smoothed spectra were used for further analysis in this study. The, reflectance spectra of cassava and maize samples are shown in Figure 4, and the plots for field data variability are shown in Appendix II.

Figure 3 Spectral measurements of Casava leaf layer stack during the field campaign in Busia County, Kenya

(a)

(b)
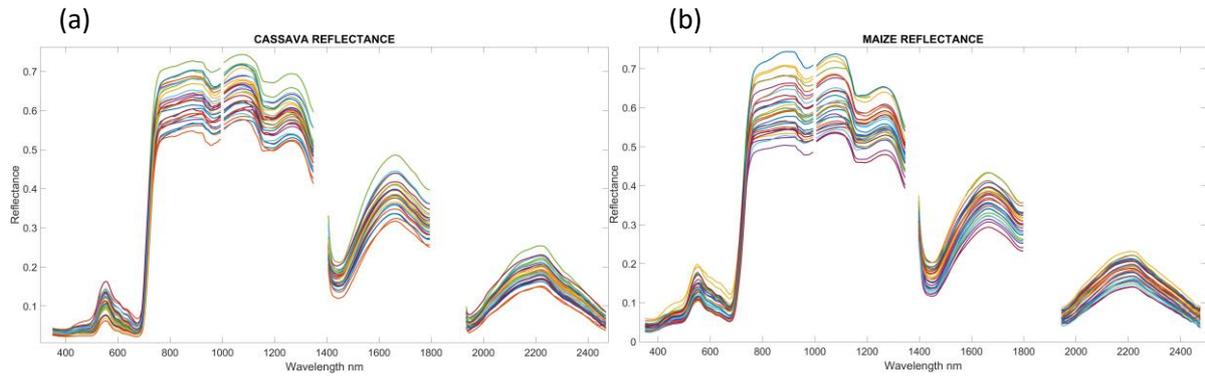
Figure 4(a) Measured leaf reflectance of Cassava (n=31), (b) Measured leaf reflectance of Maize (n=30) after smoothing using Savitzky Golay filter.

2.3.2. Satellite data extraction and pre-processing

In this study, we utilised data from the PRISMA (Precursore Iperspettrale della Missione Applicativa) satellite to perform our analysis. PRISMA is a hyperspectral satellite developed and operated by the Italian Space Agency (ASI) and was launched on 22 March 2019. It is designed to acquire high-resolution imagery with an extensive spectral range, enabling detailed and accurate characterisation of Earth's surface. The PRISMA sensor captures images in hundreds of narrow contiguous spectral bands about 237 bands with a 30 meter spatial resolution, providing valuable information about the reflectance properties of different materials and objects on the Earth's surface (Guarini et al., 2018). The hyperspectral data acquired by PRISMA allows for enhanced discrimination and identification of various land cover types, making it particularly well-suited for agricultural applications such as crop type discrimination.

The PRISMA image was acquired for this study on 1st July 2022. The PRISMA image partially covered the study area and was significantly affected by clouds, with nearly 80% of it obscured, as shown in Figure 5. We extracted the average spectral reflectance of 38 maize fields and 9 cassava fields in the study area from the PRISMA image that fall in clear areas, using ENVI Classic 5.6.3. Additionally, a non-maize class was established, consisting of fields with mixed crops, including maize planted together with cassava, beans, groundnuts, or wheat, from which we were able to extract 35 field samples.

In order to prepare the data for analysis, noisy bands affected by water absorption regions were removed, eliminating approximately 42 spectral bands (see Appendix III for further details); hence the remaining 188 bands were used for further analysis. A moving Savitzky-Golay filter with a frame size of 9 data points (1st-degree polynomial) was applied to the averaged reflectance spectra to further remove the unwanted noise and refine the data. The Savitzky-Golay filter reduces noise and has been widely used in previous studies for spectral smoothing (Ruffin & King, 2003). MATLAB R2022a (MathWorks, Inc.) was used for the processing and analysis of the data; the plots for satellite data variability is shown are Appendix IV.
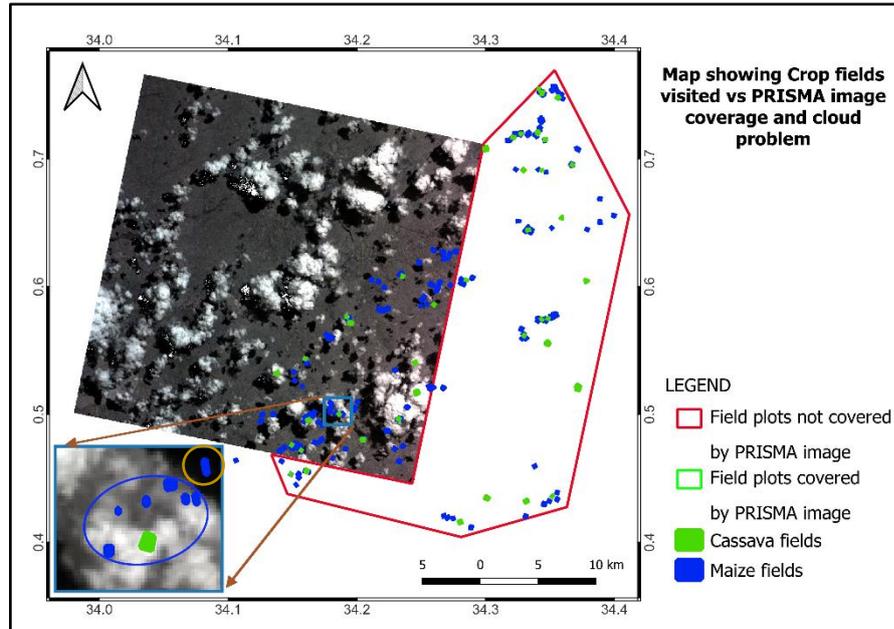
Figure 5 Map showing crop fields visited vs PRISMA image coverage and cloud problem
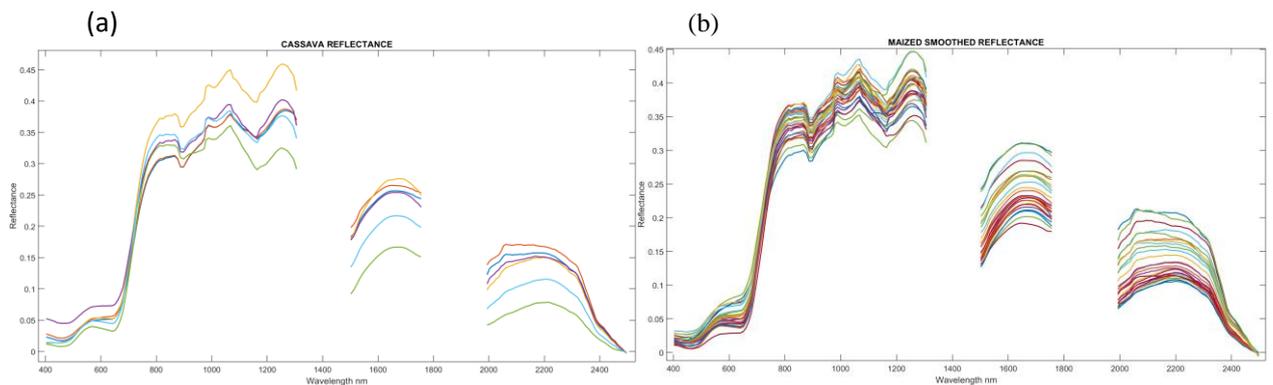
Figure 6 Spectral signature of a) Cassava fields and b) Maize fields obtained from PRISMA image data, in Busia, Kenya.

2.4. Data analysis
2.4.1. Mann-Whitney *U* test on field data

The Mann-Whitney U test compares the median spectral reflectance from maize and cassava to determine whether spectral differences exist. The test was set at 0.05 confidence level (Kokaly & Clark, 1999; Mitchell & Glenn, 2009; Schmidt & Skidmore, 2003). The main aim of the test was to check if significant differences in spectral bands exist between cassava and maize, and if so, which bands or regions show a significant difference between the two crops under an established level of confidence. This help to narrow down to the bands/regions suitable for discriminating the two crops (maize and cassava).

2.4.2. Absorption features

This study focused on analysing the spectral reflectance of cassava and maize by examining three known absorption features due to leaf biochemical content, as detailed in Table 1. Specifically, the chlorophyll absorption features in

the visible region ($R_{400-555}$ and $R_{555-771}$) were selected due to their relevance in estimating chlorophyll content, nitrogen concentration and other biochemicals in fresh standing canopies, as reported in previous studies (Mutanga et al., 2004; Sulaiman et al., 2022; Thenkabail et al., 2004). Additionally, an absorption feature found in the near infrared and short wavelength region ($R_{820-1262}$) was also considered, as this range has been shown to be influenced by biochemical components (water, cellulose, carbohydrate and lignin) in several studies (Curran et al., 2001; Kokaly & Clark, 1999; Thenkabail et al., 2000, 2004). These three absorption regions were considered for both cassava and maize at both, leaf and satellite levels. However, an additional absorption region in the short-wavelength region ($R_{1406-1677}$) was only considered for field-level data analysis, as noisy bands were removed from this region at satellite-level data (Verrelst et al., 2021).

2.4.3. Continuum-removal and feature extraction

Continuum removal was performed in regions with absorption features. The process of continuum removal involves quantifying the deviation of absorption bands of different samples (maize and cassava) from a common baseline, which is defined as the convex hull surrounding the data points of a reflectance spectrum (Mutanga et al., 2004). This common baseline, or "continuum," consists of continuous lines connecting local maxima points of the spectral reflectance (Mutanga et al., 2004). To isolate a specific absorption feature for analysis, the reflectance spectra is divided by the continuum at each wavelength: RCR = R/C where RCR is the continuum-removed spectra, R is the reflectance spectra and C is the continuum as shown in Figure 8(a), resulting in a continuum-removed spectra with values ranging from 0 to 1. The first and last points in the reflectance spectra are always local maxima and thus become 1 in the continuum-removed spectra. This technique allows for the isolated analysis of particular absorption features (band depth and wavelength position) within a spectrum as shown in Figure 7(b), Band depth (BD) was calculated by subtracting the continuum removed reflectance (R') by 1 (Kokaly & Clark, 1999).

Formally stated:

$$BD = 1 - R'$$

Where BD is band depth and R' is the continuum-removed reflectance value.
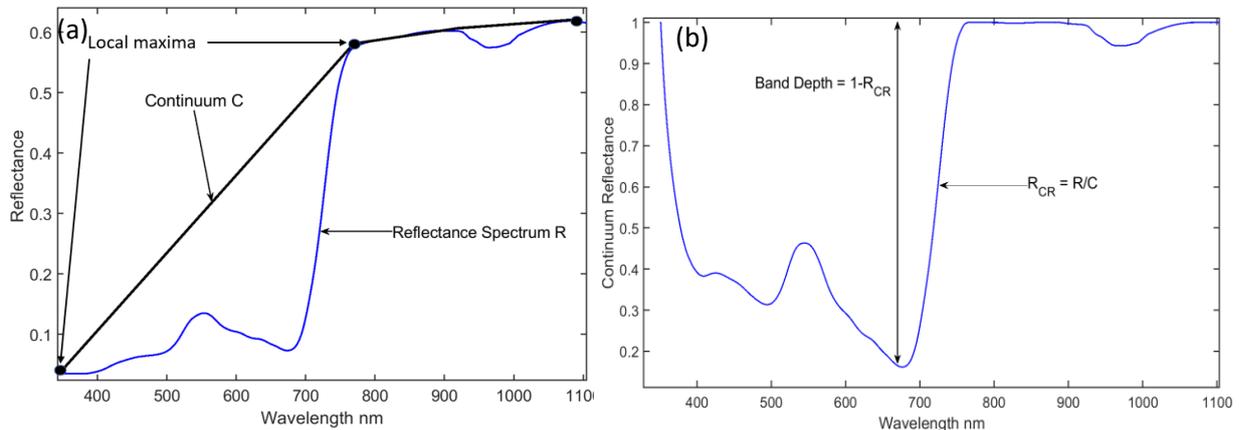
Figure 7 (a) Reflectance spectra of maize and continuum line, (b) band depth in maize continuum removed reflectance.

2.4.4. Random forest classification

A random forest (RF) classifier was used to classify the leaf level reflectance data measured at the field and satellite level data. RF has been shown to be effective in various classification tasks (Liaw & Wiener, 2002) and can handle complex patterns and nonlinearities within a dataset, in the context of our study, nonlinearity may arise due to the intricate spectral characteristics exhibited by the crops in the hyperspectral data collected at both field and satellite levels.. RF is an ensemble learning method that constructs multiple decision trees and then combines their outputs

to produce the final classification result (Breiman, 2001). During the construction of each decision tree, a bootstrap sample of the original data is used, and at each node of the tree, a random subset of features is considered for splitting. This approach helps to reduce overfitting and improve the model's generalisation performance.

We used 70% of the data to train the RF model, and the remaining 30% was used for validation. The performance of the RF model was evaluated using various metrics including producer and user accuracy, Kappa score, F1-score, p-value and overall accuracy. We evaluated the validation set to estimate the generalisation performance of the model. The results of the classification will be more elaborated in the result chapter and further analyzed under the discussion chapter of this work.

Classification was performed separately using random forest for hyperspectral data at leaf and satellite levels. To avoid overfitting the classification model, due to the high number of bands present in hyperspectral data, we only considered bands that showed very high absorption peaks after band depth analysis. At the leaf level classification was performed between cassava and maize classes but due to cloud cover only few cassava field samples (6 samples) were available at the satellite level. Hence the classification at the satellite level was performed between maize and non-maize class, which included 38 samples for maize class and 34 samples for the non-maize class. In this way, we ensured that a reasonable and comparable sample size was available.

3. RESULTS

3.1. Mean spectral reflectance of maize and cassava

The average spectral signatures of cassava and maize at the leaf and satellite levels are shown in Figure 8. Clear differences could be observed between these reflectance data in the visible and NIR regions both at the leaf (400nm to 700nm and 750nm to 1200nm) and satellite (400nm to 550 and 750nm to 1200nm) levels.
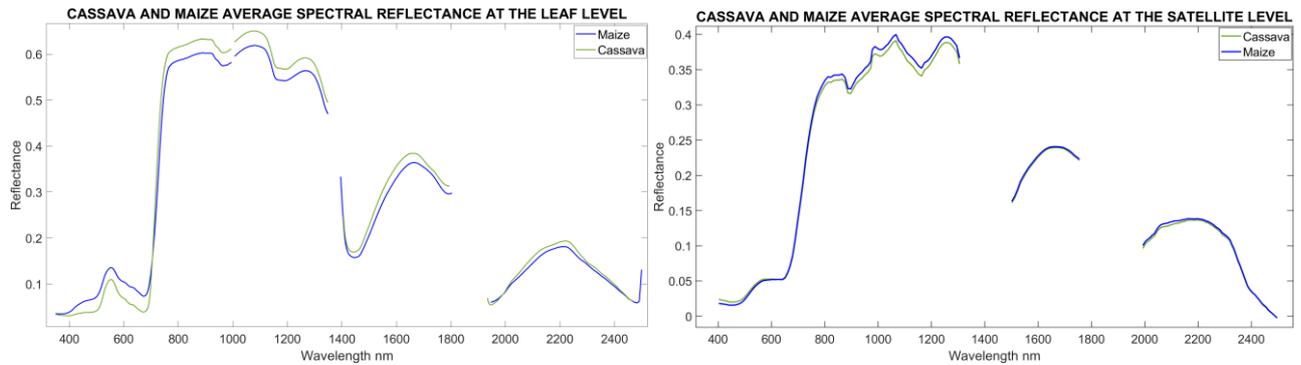
Figure 8 Average spectral signature of cassava and maize : (left) leaf level measured using ASD field spectroradiometer (cassava n=31 and maize n=30), (right) field level obtained from PRISMA image data (cassava n=6 and maize n=38), in Busia, Kenya.

3.2. Mann-Whitney *U* test results

After conducting a statistical test at a 0.05 level of confidence on the leaf level data across the entire spectral range, several spectral regions were identified in which the spectral bands from these regions showed significant differences between cassava and maize. These significant regions are presented in Table 1. From these significant spectral regions, we were able to identify absorption regions that were used for further investigation in this study.

Table 1 Selected significant spectral regions based on a Mann-Whitney *U* test.

Significant spectral regions at leaf Level (p-value < 0.05)
385nm -701nm
709nm - 1346nm
1455nm - 1658nm
2070nm - 2115nm
2237nm - 2254nm

3.3. Continuum removal and band depth analysis

After applying continuum removal to the selected absorption regions described in subsections 2.4.2 and 2.4.3, we analyzed the band depth on continuum reflectance between cassava and maize at the leaf and satellite levels. Tables 2 and 3 show the band depth in these absorption regions. As can be seen from these tables, the band depth varied between cassava and maize at the leaf and satellite levels. Nevertheless, there are similarities in bands with the highest absorption peaks between these two data levels.

Table 2 Absorption regions and band depth at the leaf level.

Wavelength range (nm)	Band with the highest absorption peak (nm)	Band depth	
		Cassava	Maize
400-555	496	0.502	0.309
555-771	677 & 679	0.901	0.809
820-1262	1162	0.089	0.085
1406-1677	1459 & 1460	0.362	0.374

Table 3 Absorption regions and band depth at satellite level.

Wavelength range (nm)	Band with the highest absorption peak (nm)	Band depth	
		Cassava	Maize
400-555	470	0.389	0.443
555-771	650	0.659	0.658
820-1262	1163	0.125	0.115
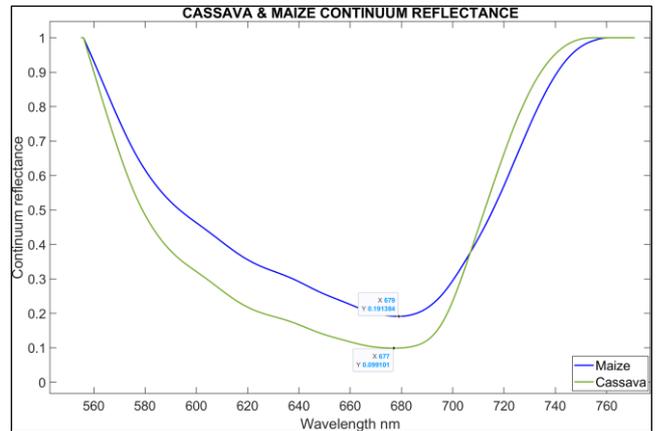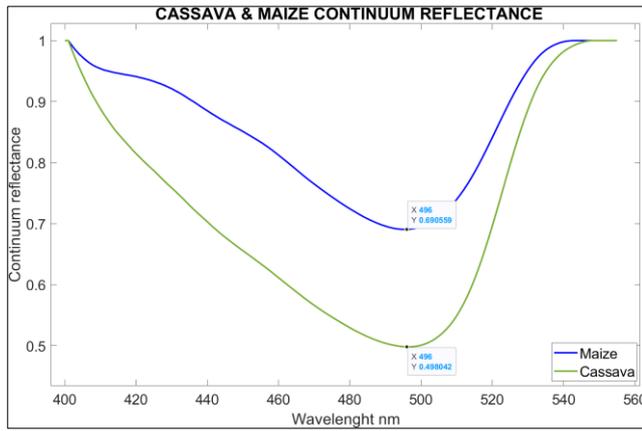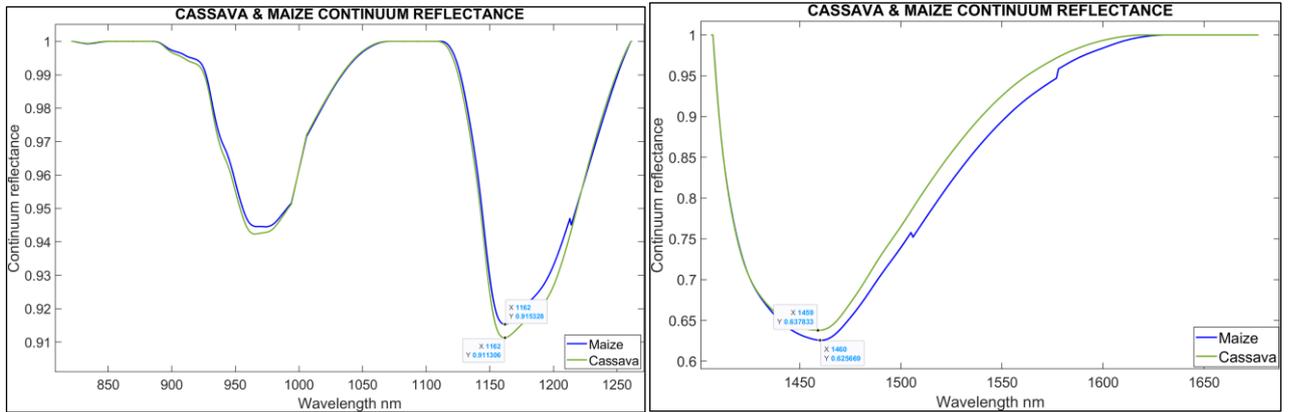
Figure 9 Continuum removed reflectance of smoothed field data

3.4. Random forest classification

3.4.1. Classification using leaf measurement collected in the filed

Significant bands with the highest absorption peaks (496, 677, 679, 1162, 1459, and 1460) were used to assess the classification of cassava and maize samples collected at the field level. To create the model, we used 30 maize samples (21 samples for training and 9 samples for testing) and 31 cassava samples (22 samples for training and 9 samples for testing). The classification of the two crops using leaf reflectance data resulted in an overall accuracy of 94%, as measured by the kappa score of 0.89 and F1 score of 0.94. Table 4 summarizes the classification outputs for the cassava and maize classes.

Table 4 Cassava and maize classification output at field level data based on testing data.

Crop	Producer accuracy	User accuracy	F1 score	Kappa score	*P*-value	Overall accuracy
Cassava	0.8	0.89	0.94	0.89	5.6329e-05	94%
Maize	0.9	1				

3.4.2. Classification at satellite level on PRISMA data

At the satellite level, due to the small number of cassava field samples classification was performed to discriminate maize from other crops. For this, the significant bands with the highest absorption peaks (470, 650 and 1163) were used to classify, maize and non-maize classes. We used 38 maize field samples (27 samples for training and 11 samples for testing) and 35 non-maize samples (25 samples for training and 10 samples for testing) to create the model. The classification output had an overall accuracy of 77%, a kappa of 0.54, and an F1 score of 0.78. Table 5 summarized the classification outputs for the maize and non- maize classes.

Table 5 Confusion matrix - results of maize and non-maize classification using PRISMA image data based on testing data.

Crop	Producer accuracy	User accuracy	F1 score	Kappa score	*P*-value	Overall accuracy
Maize	0.72	0.8	0.78	0.55	1.5068e-04	77%
Non-maize	0.82	0.75				

4. DISCUSSION

Discriminating major crop types (cassava and maize) in Kenya using hyperspectral remote sensing is of great significance in agriculture and crop monitoring. This study aimed to discriminate cassava and maize crops using their spectral signatures obtained through field spectroscopy of leaves layer stacks and PRISMA satellite in Busia County, Kenya. To study cassava and maize discrimination important bands were identified using Mann-whitney U test, continuum removal and band depth analysis. Random forest classification models were developed based on the bands from the absorption peaks, and the accuracy of crop discrimination was evaluated at both leaf and satellite levels. The main finding were, there was significant difference between cassava and maize spectral signatures, the regions suitable for cassava and maize discrimination at leaf level are suitable at satellite level also and by considering bands with highest absorption peak is possible to discriminate maize from the other crops.   In this section, we will discuss the observed differences in spectral signatures reflectance of the two crops, the suitable hyperspectral bands for discrimination, and classification outputs after only considering the bands with the highest absorption peak, also assumptions and possibilities for future studies at each section will be discussed.

4.1. Variations in spectral reflectance of cassava and maize

The difference observed in the spectral reflectance signatures of cassava and maize at the leaf level confirms the findings of earlier studies by Thenkabail et al. (2004), who reported variations in the spectral response of corn, rice, cowpea, groundnut, soybean and cassava crops. The significant differences in the visible and near-infrared (NIR) regions of the mean spectra of the cassava and maize, which are primarily influenced by the leaf physiology and biochemical composition of the species, suggest that these regions are particularly useful for discriminating between cassava and maize.

At the leaf level, maize exhibited higher reflectance than cassava in the visible range and cassava exhibited higher reflectance than maize in the NIR and SWIR range (Figure 9). This difference can be attributed to the existing variations in cassava and maize leaf properties such as chlorophyll content, water content in a leaf, leaf thickness and roughness. Several studies have shown that water content affects leaf reflectivity in the NIR range and chlorophyll content in the visible range, with lower water content leading to higher reflectivity and high chlorophyll content reducing leaf reflectivity (Daniel et al., 2003; García et al., 2009; Hegarty-Craver et al., 2020). These findings support the argument that the observed differences in spectral signatures between cassava and maize at the leaf stack level are mainly influenced by the biochemical makeup of the crops rather than differences in the growth stage (phenology). In this study, both maize and cassava were at the same growth stage (maturity stage); hence there was no difference in phenology to cause the difference in the reflectance between the two crops.

The Mann-Whitney U test results conducted on the leaf level data (Table 1) further support the observed differences between cassava and maize. The test revealed significant differences at several wavelength regions, particularly in the visible and NIR regions (385-701nm, 709-1346nm, 1455-1658nm, 2070-2115nm and 2237-2254nm), this statistical test was performed at leaf level only but it was not possible to perform this statistical test to satellite level data due to unbalanced number of samples between the two crops (maize, n=38 while cassava, n=6). These significant wavelengths, which exhibit distinct reflectance patterns for cassava and maize, can be attributed to the variations in leaf physiology and biochemical composition difference between the two crops. The significant wavelength regions align with several other studies that have identified similar wavelength regions to be significant in discriminating crop types because with hyperspectral data these wavelength regions contain information about crop biophysical and biochemical characteristics which can be used in distinguishing one crop from another (Curran et al., 2001; Odindi & Kakembo, 2009; Schmidt & Skidmore, 2003; Thenkabail et al., 2004).

One limitation of this study was the inability to measure canopy-level reflectance due to the height of the maize plants in the field, which exceeded two meter. Consequently, measurements were taken at the leaf level, providing data only on leaf reflectance. This limitation hinders a comprehensive understanding of overall canopy reflectance and introduces potential biases in the results and interpretations. To overcome this limitation in future studies, it is recommended to explore the use of field measuring platforms or alternative measurement techniques that enable the collection of canopy-level reflectance data. By incorporating such approaches, researchers can gain a more comprehensive understanding of spectral signatures and variations at the canopy level, facilitating a more precise assessment of crop discrimination and characterisation.

No biochemical laboratory analysis was conducted in this study, and the identification of biochemical components associated with differences at various wavelengths relies on existing studies that have performed laboratory experiments. To enhance future investigations, it is suggested that biochemical analyses be performed on cassava and maize samples. This would enable more accurate identification of the biochemical components responsible for spectral reflectance differences across different regions of the electromagnetic spectrum. These analyses, would enhance understanding of the underlying biochemical processes influencing spectral signatures and further can be used to refine crop discrimination techniques.

4.2. Hyperspectral bands suitable for discrimination at the leaf stack layer and satellite level

In our leaf level data, we identified six bands with the highest absorption peaks at wavelengths of 496 nm, 677 nm, 679 nm, 1162 nm, 1459 nm, and 1460 nm (Table 2). These bands were found to be associated with specific pigments and biochemical components, as supported by the existing literature on plant biochemistry. For example, chlorophyll "a" and "b" were identified as the primary pigments responsible for light absorption at 496 nm, 677 nm, and 679 nm (Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014), playing essential roles in photosynthesis. At 1162 nm, water and cellulose were found to be the main biochemical components responsible for light absorption (Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014), which has implications for remote sensing of plant water status and soil moisture. Furthermore, carbohydrates and lignin were found to be the primary biochemical components responsible for light absorption at 1459 nm and 1460 nm (Curran, 1989; Kokaly & Clark, 1999; Kumar et al., 2001), responsible for plant cell wall structure and function (Curran et al., 2001). At the satellite level, three bands were identified as having the highest absorption peak 470 nm, 650 nm, and 1163 nm (Table 3), we observed that chlorophyll "a" and "b" were again the main pigments responsible for light absorption at 470 nm and 650 nm, while cellulose was the primary biochemical component associated with light absorption at 1163 nm (Curran, 1989; Kokaly & Clark, 1999; Kumar et al., 2001), Cellulose, as the primary structural component of plant cell wall, plays a crucial role in providing mechanical support and rigidity. These relationships between wavelengths and responsible biophysical parameter have been summarised in Appendix V.

Remarkably, we noted that the $R_{820-1262}$ absorption region exhibited similar patterns at both the leaf stack level and satellite level, with 1162 nm and 1163 nm as the highest absorption peaks, indicating the involvement of water and cellulose as the responsible biochemical components. This suggests that these bands may be useful for discriminating between cassava and maize at both the leaf stack and satellite levels. Another significant finding was observed in the $R_{400-555}$ absorption region, where we observed the highest difference in band depth between maize and cassava for both leaf stack-level and satellite-level data. This finding holds considerable importance for crop management and monitoring, providing valuable information for resource allocation, yield prediction, and early detection of stress or disease. The results align with existing finding which finds blue region and NIR regions to be suitable regions for crop discrimination (Buchhorn et al., 2013; Hennessy et al., 2020).

In this study, we adopt a comprehensive approach by examining sensitive bands for crop discrimination at both the leaf stack layer level and the satellite level, unlike other studies that typically focus on investigating the sensitive bands for crop discrimination either at the satellite level or the field level individually. Using continuum removal output and band depth analysis, we established the correlation between specific wavelengths and the biochemical

components of the plants. However, it is crucial to acknowledge the potential variations in plant samples due to factors such as genotype and farming inputs. In the Busia region, where a significant number of farmers face economic constraints, different maize and cassava species were observed in fields based on seed availability, and farmers utilised various farming inputs depending on their financial capacity. Therefore, further research is recommended to investigate the impact of seed variations and farming inputs on maize and cassava in the Busia-Kenya region. This can be accomplished through laboratory experiments and additional analytical investigations focusing on specific wavelengths in relation to crop type. By doing so, we can enhance our understanding of crop discrimination and characterisation, leading to more precise assessments and practical applications in agriculture.

4.3. Classification

Classification of cassava and maize crops at the field level using only bands with the highest absorption peaks achieved an overall accuracy of 94% (kappa score: 0.89, F1 score: 0.94), indicating the high predictive power of the model. This finding aligns with previous studies that employed hyperspectral field data for crop classification (de Leeuw et al., 2007; Piiroinen et al., 2015). However, unlike most of these studies which considered bands estimated to be significant, our study focuses solely on bands with the highest absorption peaks because this approach has not yet been tested in previous studies so it was worthy considering in this study. Despite the use of this new approach, the results are convincing, suggesting the potential of these bands for future research. Moreover, the use of only significant bands with the highest absorption peaks for classification at the leaf level highlights the accuracy and efficiency provided by hyperspectral data in crop discrimination. This approach could be further developed to enable precise and cost-effective monitoring of crop health and yield at the field level.

At the satellite level, the classification results of maize and non-maize exhibited slightly lower accuracy compared to the field-level data (overall accuracy of 94%), with an overall accuracy of 77% (kappa score: 0.54, F1 score: 0.78). This discrepancy can be attributed to various factors, including the influence of additional variables which have been acknowledged by other studies to affect classification such as, background (soil) effect (Sabat-Tomala et al., 2020), atmosphere noise (Underwood et al., 2003), weed (Lu et al., 2020) and small farmland size relative to the spatial resolution of the PRISMA image (30m). Nevertheless, the successful classification of maize and non-maize classes at the satellite level demonstrates the potential of hyperspectral remote sensing for crop mapping and monitoring, other studies in the literature have also shown promising results in land cover classification of agricultural crops by using hyperspectral satellite images. For instance, Pepe et al. 2022 use similar hyperspectral remote sensing sensor (PRISMA) at Jolanda di Savoia estate Emilia-Romagna region Italy to perform classification of nine crop types, including maize, and achieved an overall accuracy of 90%. These findings further support the effectiveness of hyperspectral remote sensing for crop discrimination.

It is important to acknowledge that the classification models used in this study were based on a limited number of samples (30 maize samples and 31 cassava samples for leaf stack layer level classification, and 38 maize samples and 35 nonmaize samples for satellite-level classification) and focused on a specific maturity stage of the crops only. Enhancing the accuracy of the classification models could be achieved through future studies with larger sample sizes, considering different stages of crop growth, and encompassing more diverse crop types.

Overall, the findings of this study indicate that hyperspectral remote sensing, employing bands with the highest absorption peaks, can deliver accurate and efficient crop classification at both field and satellite levels. The presented approach has the potential to improve crop monitoring and management, contributing to more sustainable and productive agricultural practices.

5. CONCLUSION

This study successfully demonstrated the potential of hyperspectral remote sensing for crop type discrimination at the field level and satellite level for maize and cassava crops in Busia County, Kenya. The study had three specific objectives: studying the spectral signatures of maize and cassava, exploring whether the hyperspectral wavebands/regions suitable for discrimination of maize and cassava at the field level are relevant at the satellite level, and discriminating between maize and cassava fields using hyperspectral field data and satellite data (classification).

Analysis of the spectral signature of cassava and maize at the field level revealed clear differences between the two crops. These differences were identified in specific absorption regions. The main biochemical components associated with the difference in spectral characteristics between maize and cassava include chlorophyll a and b (at bands 496, 677, and 679 nm), water and cellulose content (at band 1162 nm), and carbohydrates and lignin (at bands 1459 and 1460 nm). These findings indicate that at the leaf level, distinct spectral characteristics can differentiate between cassava and maize, providing valuable insights into crop discrimination.

Considering the same absorption regions as used at the field level, the distinct differences between cassava and maize were identified at bands 470 and 650 at the satellite level. Chlorophyll a and b were the main responsible biochemical components at these bands, while band 1163 nm exhibited cellulose content as the main biochemical component responsible for the difference in spectral characteristics. Moreover, the absorption region R400-555 showed the highest difference in band depth between maize and cassava at both the field and satellite levels, with chlorophyll a and b being the main biochemical components found in this region. These findings suggest that discrimination between the two crops can be achieved at this region, regardless of the spatial resolution of the sensor used (field or satellite measurement).

By considering only bands with the highest absorption peaks, it is possible to discriminate between maize and other crop classes. The results of this study showed that at the field level, by only considering bands with the highest absorption peak, an accuracy of 93% was achieved during the classification of cassava and maize. At the satellite level, an accuracy of 77% was attained during the classification of maize and non-maize classes. These findings demonstrate the potential of using hyperspectral satellite data for discriminating maize fields from other crops in Busia County, Kenya.

This study contributes to the field of crop monitoring and management by providing valuable insights into crop type discrimination. With the provision of accurate and efficient discrimination of crop types at both the field and satellite levels.

Overall, the results of this study suggest that hyperspectral remote sensing, by using only bands with the highest absorption peaks can provide accurate and efficient discrimination of crop types at both the field and satellite levels. The approach presented in this study has the potential to improve crop monitoring and management, leading to more sustainable and productive agricultural practices. Future research can build on the findings of this study to develop more advanced methods for crop discrimination which may involve more biochemical components of the crops and involve more crop types using hyperspectral remote sensing, and to explore its application in other agricultural regions worldwide.

6. LIST REFERENCES

Aneece, I., & Thenkabail, P. S. (2021). Classifying crop types using two generations of hyperspectral sensors (Hyperion and DESIS) with machine learning on the cloud. *Remote Sensing, 13*(22). https://doi.org/10.3390/RS13224704

Arafat, S. M., Aboelghar, M. A., & Ahmed, E. F. (2013). Crop Discrimination Using Field Hyper Spectral Remotely Sensed Data . *Advances in Remote Sensing, 2*, 63–70. https://doi.org/10.4236/ars.2013.22009

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., & Brivio, P. A. (2017). Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. *Https://Doi.Org/10.5721/EuJRS20164920, 49*, 361–380. https://doi.org/10.5721/EUJRS20164920

Buchhorn, M., Walker, D. A., Heim, B., Raynolds, M. K., Epstein, H. E., & Schwieder, M. (2013). Ground-Based Hyperspectral Characterisation of Alaska Tundra Vegetation along Environmental Gradients. *Remote Sensing 2013, Vol. 5, Pages 3971-4005, 5*(8), 3971–4005. https://doi.org/10.3390/RS5083971

Chi, J., & Crawford, M. M. (2014). Spectral unmixing-based crop residue estimation using hyperspectral remote sensing data: A case study at Purdue University. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(6), 2531–2539. https://doi.org/10.1109/JSTARS.2014.2319585

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment, 30*(3), 271–278. https://doi.org/10.1016/0034-4257(89)90069-2

Curran, P. J., Dungan, J. L., & Peterson, D. L. (2001). Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies. *Remote Sensing of Environment, 76*(3), 349–359. https://doi.org/10.1016/S0034-4257(01)00182-1

Daniel, K. W., Tripathi, N. K., & Honda, K. (2003). Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Australian Journal of Soil Research, 41*(1), 47–60. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00049573&v=2.1&it=r&id=GALE%7CA98314097&sid=googleScholar&linkaccess=fulltext

de Leeuw, J., Jia, H., Yang, L., Liu, X., Schmidt, K., & Skidmore, A. K. (2007). Comparing accuracy assessments to infer superiority of image classification methods. *Http://Dx.Doi.Org/10.1080/01431160500275762, 27*(1), 223–232. https://doi.org/10.1080/01431160500275762

Dhumal, R. K., Vibhute, A. D., Nagne, A. D., Rajendra, Y. D., Kale, K. V., & Mehrotra, S. C. (2015). Advances in Classification of Crops using Remote Sensing Data. *International Journal of Advanced Remote Sensing and GIS, 4*(1), 1410–1418. https://doi.org/10.23953/CLOUD.IJARSG.130

FAO. (2021). *Africa Regional Overview of Food Security and Nutrition 2021: Statistics and trends.* https://doi.org/10.4060/cb7496en

Galvão, L. S., Epiphanio, J. C. N., Breunig, F. M., & Formaggio, A. R. (2018). Crop Type Discrimination Using Hyperspectral Data : Advances and Perspectives. In *Biophysical and Biochemical Characterisation and Plant Species Studies* (pp. 183–210). CRC Press. https://doi.org/10.1201/9780429431180-6

García, N. L., Famá, L., Dufresne, A., Aranguren, M., & Goyanes, S. (2009). A comparison between the physico-chemical properties of tuber and cereal starches. *Food Research International, 42*(8), 976–982. https://doi.org/10.1016/J.FOODRES.2009.05.004

Guarini, R., Loizzo, R., Facchinetti, C., Longo, F., Ponticelli, B., Faraci, M., Dami, M., Cosi, M., Amoruso, L., De Pasquale, V., Taggio, N., Santoro, F., Colandrea, P., Miotti, E., & Di Nicolantonio, W. (2018). PRISMA hyperspectral mission products. *International Geoscience and Remote Sensing Symposium (IGARSS), 2018-July*, 179–182. https://doi.org/10.1109/IGARSS.2018.8517785

Hegarty-Craver, M., Polly, J., O'Neil, M., Ujeneza, N., Rineer, J., Beach, R. H., Lapidus, D., & Temple, D. S. (2020). Remote Crop Mapping at Scale: Using Satellite Imagery and UAV-Acquired Data as Ground Truth. *Remote Sensing 2020, Vol. 12, Page 1984*, *12*(12), 1984. https://doi.org/10.3390/RS12121984

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. *Remote Sensing 2020, Vol. 12, Page 113*, *12*(1), 113. https://doi.org/10.3390/RS12010113

Huang, W., Guan, Q., Luo, J., Zhang, J., Zhao, J., Liang, D., Huang, L., & Zhang, D. (2014). New optimized spectral indices for identifying and monitoring winter wheat diseases. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2516–2524. https://doi.org/10.1109/JSTARS.2013.2294961

Inglada, J., Arias, M., Tardy, B., Morin, D., Valero, S., Hagolle, O., Dedieu, G., Sepulcre, G., Bontemps, S., & Defourny, P. (2015). Benchmarking of algorithms for crop type land-cover maps using Sentinel-2 image time series. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 3993–3996. https://doi.org/10.1109/IGARSS.2015.7326700

Janse, P. V., & Deshmukh, R. R. (2021). Crop Discrimination using Non-Imaging Hyperspectral Data. *International Journal of Engineering and Advanced Technology*, *10*(5), 269–273. https://doi.org/10.35940/IJEAT.E2802.0610521

Kokaly, R. F., & Clark, R. N. (1999). Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sensing of Environment*, *67*(3), 267–287. https://doi.org/10.1016/S0034-4257(98)00084-4

Kumar, L., Schmidt, K. S., Dury, S., & Skidmore, A. K. (2001). Imaging spectroscopy and vegetation science. *Remote Sensing and Digital Image Processing*, *4*, 111–155. https://research.wur.nl/en/publications/imaging-spectroscopy-and-vegetation-science

Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sensing 2020, Vol. 12, Page 2659*, *12*(16), 2659. https://doi.org/10.3390/RS12162659

Maja, M. M., & Ayano, S. F. (2021). The Impact of Population Growth on Natural Resources and Farmers' Capacity to Adapt to Climate Change in Low-Income Countries. *Earth Systems and Environment*, *5*(2), 271–283. https://doi.org/10.1007/S41748-021-00209-6/FIGURES/5

Manjunath, K. R., Kumar, A., Meenakshi, M., Renu, R., Uniyal, S. K., Singh, R. D., Ahuja, P. S., Ray, S. S., & Panigrahy, S. (2014). Developing Spectral Library of Major Plant Species of Western Himalayas Using Ground Observations. *Journal of the Indian Society of Remote Sensing*, *42*(1), 201–216. https://doi.org/10.1007/S12524-013-0305-0/TABLES/4

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 191–210. https://doi.org/10.1016/j.isprsjprs.2022.03.008

Martin, M. P., Barreto, L., RiañO, D., Fernandez-Quintanilla, C., & Vaughan, P. (2011). Assessing the potential of hyperspectral remote sensing for the discrimination of grassweeds in winter cereal crops. *International Journal of Remote Sensing*, *32*(1), 49–67. https://doi.org/10.1080/01431160903439874

Mitchell, J. J., & Glenn, N. F. (2009). Leafy Spurge (Euphorbia esula) Classification Performance Using Hyperspectral and Multispectral Sensors. *Rangeland Ecology & Management*, *62*(1), 16–27. https://doi.org/10.2111/08-100

Mohammadi, S., Belgiu, M., & Stein, A. (2023). Improvement in crop mapping from satellite image time series by effectively supervising deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *198*, 272–283. https://doi.org/10.1016/J.ISPRSJPRS.2023.03.007

Mutanga, O., Skidmore, A. K., & Prins, H. H. T. (2004). Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, *89*(3), 393–408. https://doi.org/10.1016/J.RSE.2003.11.001

Odindi, J. O., & Kakembo, V. (2009). The use of laboratory spectroscopy to establish pteronia incana spectral trends and separation from bare surfaces and green vegetation. *South African Geographical Journal*, *91*(1), 16–24. https://doi.org/10.1080/03736245.2009.9725326

Ozdarici-Ok, A., Ok, A. O., & Schindler, K. (2015). Mapping of Agricultural Crops from Single High-Resolution Multispectral Images—Data-Driven Smoothing vs. Parcel-Based Smoothing. *Remote Sensing* , *7*(5), 5611–5638. https://doi.org/10.3390/RS70505611

Piiroinen, R., Heiskanen, J., Mõttus, M., & Pellikka, P. (2015). Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 1–8. https://doi.org/10.1016/J.JAG.2015.02.005

Ruffin, C., & King, R. L. (2003). *The analysis of hyperspectral data using Savitzky-Golay filtering-theoretical basis. 1*. 756–758. https://doi.org/10.1109/IGARSS.1999.774430

Sabat-Tomala, A., Raczko, E., & Zagajewski, B. (2020). Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sensing 2020, Vol. 12, Page 516*, *12*(3), 516. https://doi.org/10.3390/RS12030516

Schmidt, K. S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, *85*(1), 92–108. https://doi.org/10.1016/S0034-4257(02)00196-7

Singh, P., Srivastava, P. K., Shah, D., Pandey, M. K., Anand, A., Prasad, R., Dave, R., Verrelst, J., Bhattacharya, B. K., & Raghubanshi, A. S. (2022). Crop type discrimination using Geo-Stat Endmember extraction and machine learning algorithms. *Advances in Space Research*. https://doi.org/10.1016/J.ASR.2022.08.031

Sulaiman, N., Che'ya, N. N., Mohd Roslim, M. H., Juraimi, A. S., Mohd Noor, N., & Fazlil Ilahi, W. F. (2022). The Application of Hyperspectral Remote Sensing Imagery (HRSI) for Weed Detection Analysis in Rice Fields: A Review. *Applied Sciences 2022, Vol. 12, Page 2570*, *12*(5), 2570. https://doi.org/10.3390/APP12052570

Tagliabue, A., Bowie, A. R., Holmes, T., Latour, P., van der Merwe, P., Gault-Ringold, M., Wuttig, K., & Resing, J. A. (2022). Constraining the Contribution of Hydrothermal Iron to Southern Ocean Export Production Using Deep Ocean Iron Observations. *Frontiers in Marine Science*, *9*, 98. https://doi.org/10.3389/FMARS.2022.754517/BIBTEX

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, *91*(3–4), 354–376. https://doi.org/10.1016/J.RSE.2004.03.013

Thenkabail, P. S., Lyon, J. G., & Alfredo, H. (2018). Advanced Applications in Remote Sensing of Agricultural Crops and Natural Vegetation. In P.S. Thenkabail, J. . Lyon, & A. (Eds. . Huete (Eds.), *Advanced Applications in Remote Sensing of Agricultural Crops and Natural Vegetation: Vol. IV* (2nd ed.). CRC Press. https://doi.org/10.1201/9780429431166

Thenkabail, Prasad S., Mariotto, I., Gumma, M. K., Middleton, E. M., Landis, D. R., & Huemmrich, K. F. (2013). Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterisation and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(2), 427–439. https://doi.org/10.1109/JSTARS.2013.2252601

Thenkabail, Prasad S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. *Remote Sensing of Environment*, *71*(2), 158–182. https://doi.org/10.1016/S0034-4257(99)00067-X

Underwood, E., Ustin, S., & DiPietro, D. (2003). Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment*, *86*(2), 150–161. https://doi.org/10.1016/S0034-4257(03)00096-8

United Nations. (2019). *World Population Prospects 2019 Highlights*.

Van Ittersum, M. K., Van Bussel, L. G. J., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., Claessens, L., De Groot, H., Wiebe, K., Mason-D'Croz, D., Yang, H., Boogaard, H., Van Oort, P. A. J., Van Loon, M. P., Saito, K., Adimo, O., Adjei-Nsiah, S., Agali, A., Bala, A., … Cassman, K. G. (2016). Can sub-Saharan Africa feed itself? *Proceedings of the National Academy of Sciences of the United States of America*, *113*(52), 14964–14969. https://doi.org/10.1073/PNAS.1610359113/SUPPL_FILE/PNAS.201610359SI.PDF

Verrelst, J., Rivera-Caicedo, J. P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., Panigada, C., Hank, T., & Berger, K. (2021). Mapping landscape canopy nitrogen content from space using PRISMA data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, 382–395. https://doi.org/10.1016/J.ISPRSJPRS.2021.06.017

Zhang, X., Sun, Y., Shang, K., Zhang, L., & Wang, S. (2016). Crop Classification Based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(9), 4117–4128. https://doi.org/10.1109/JSTARS.2016.2577339

Zhang, Y., Gao, J., Cen, H., Lu, Y., Yu, X., He, Y., & Pieters, J. G. (2019). Automated spectral feature extraction from hyperspectral images to differentiate weedy rice and barnyard grass from a rice crop. *Computers and Electronics in Agriculture*, *159*, 42–49. https://doi.org/10.1016/J.COMPAG.2019.02.018

APPENDIX SECTION

Appendix I: Noise band removed from field spectral measured data

|	Cassava(nm)	No. bands lost	Maize (nm)	No. bands lost
Noise band due to sensor changes in spectral sampling characteristics. (from 1.4 to 2nm)	1000	1	1000	1
water absorption bands	1347 up to 1405	59	1347 up to 1405	59
Atmospheric water vapor bands	1796 up to 1943	148	1796 up to 1943	148
Noise at the end of signature	2470 up to 2500	31	2481 up to 2500	31
The total number of bands lost		**239**		**239**

Appendix II: Field level mean reflectance standard deviation (on top) and first derivative (on bottom)
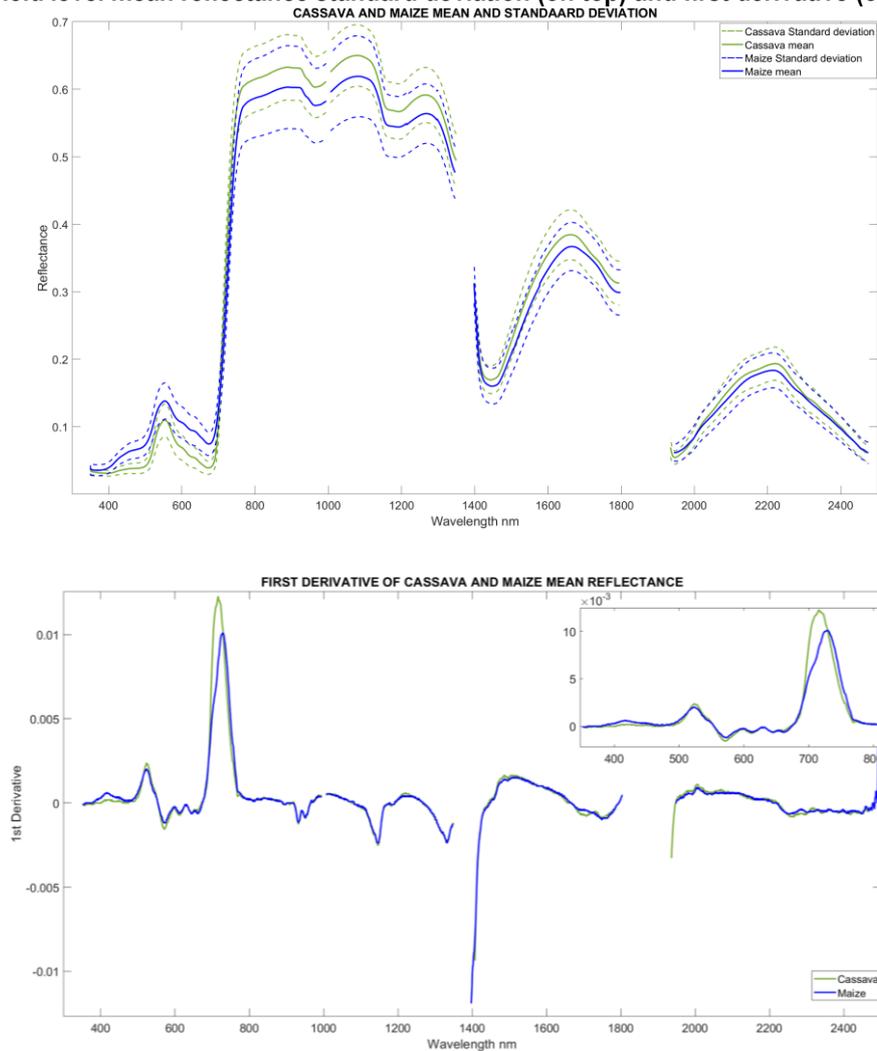
Appendix III: Noise band removed from satellite spectral measured data

	Cassava (nm)	No. bands lost	Maize (nm)	No. bands lost
water absorption bands	1370 up to 1491	17	1370 up to 1491	17
Atmospheric water vapor bands	1765 up to 1984	25	1765 up to 1984	25
The total number of bands lost	**42**			**42**

Appendix IV: Satellite level mean reflectance standard deviation (on top) and first derivative (on bottom)
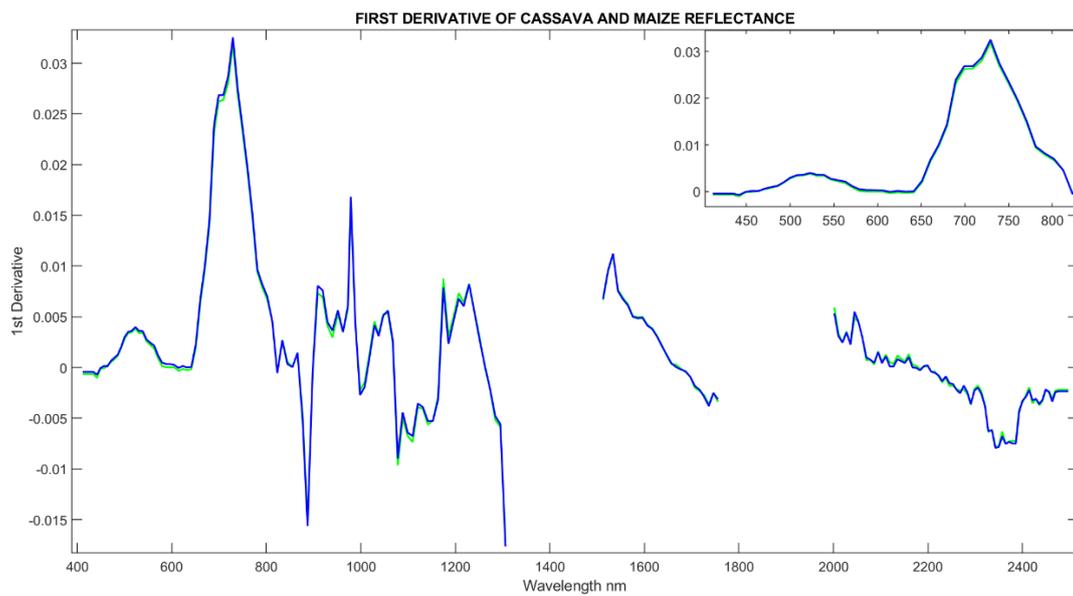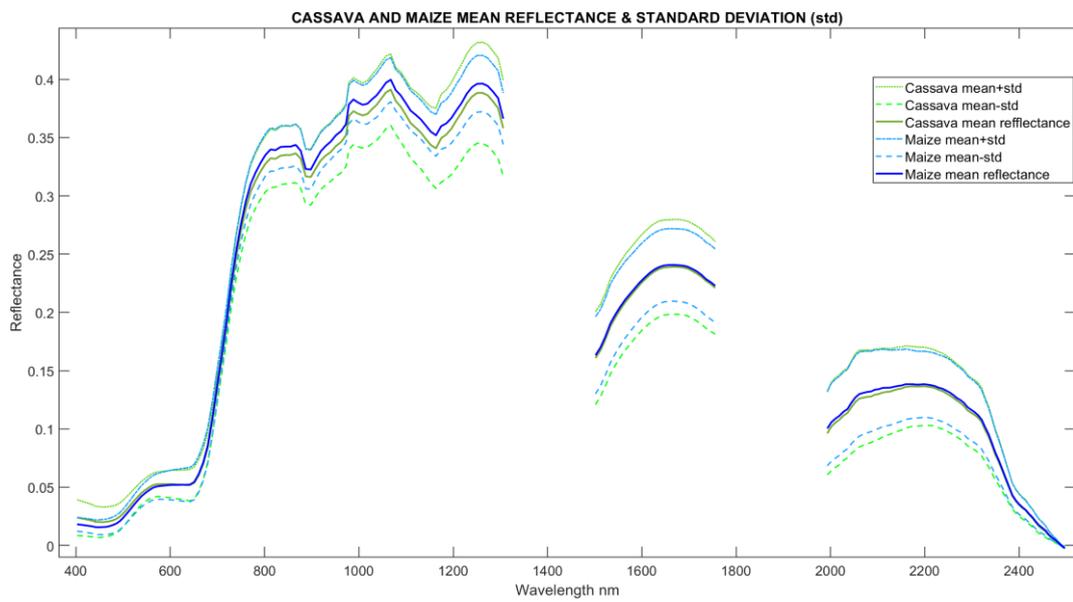
Appendix V: Bands with the highest absorption peak and related known biochemicals and supported citations used to establish the relationship

Wavelength Range (nm)	Band with the highest absorption peak (nm)	Known causal biochemical	Reference	Data Level
$R_{400-555}$	496	Chlorophyll "a" and "b"	(Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014)	
$R_{555-771}$	677 679	Chlorophyll "a" and "b"	(Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014)	**Field level/ Stack level**
$R_{820-1262}$	1162	Water and Cellulose	(Curran, 1989; Kokaly & Clark, 1999; Kumar et al., 2001)	
$R_{1406-1677}$	1459 1460	Carbohydrates and Lignin	(Curran, 1989; Kokaly & Clark, 1999; Kumar et al., 2001)	
$R_{400-555}$	470	Chlorophyll "a" and "b"	(Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014)	
$R_{555-771}$	650	Chlorophyll "a" and "b"	(Curran, 1989; Curran et al., 2001; Kokaly & Clark, 1999; Manjunath et al., 2014)	**Satellite level**
$R_{820-1262}$	1163	Cellulose	(Curran, 1989; Kokaly & Clark, 1999; Kumar et al., 2001)	

Appendix VI: Cassava field pictures

Appendix VII: Maize field pictures
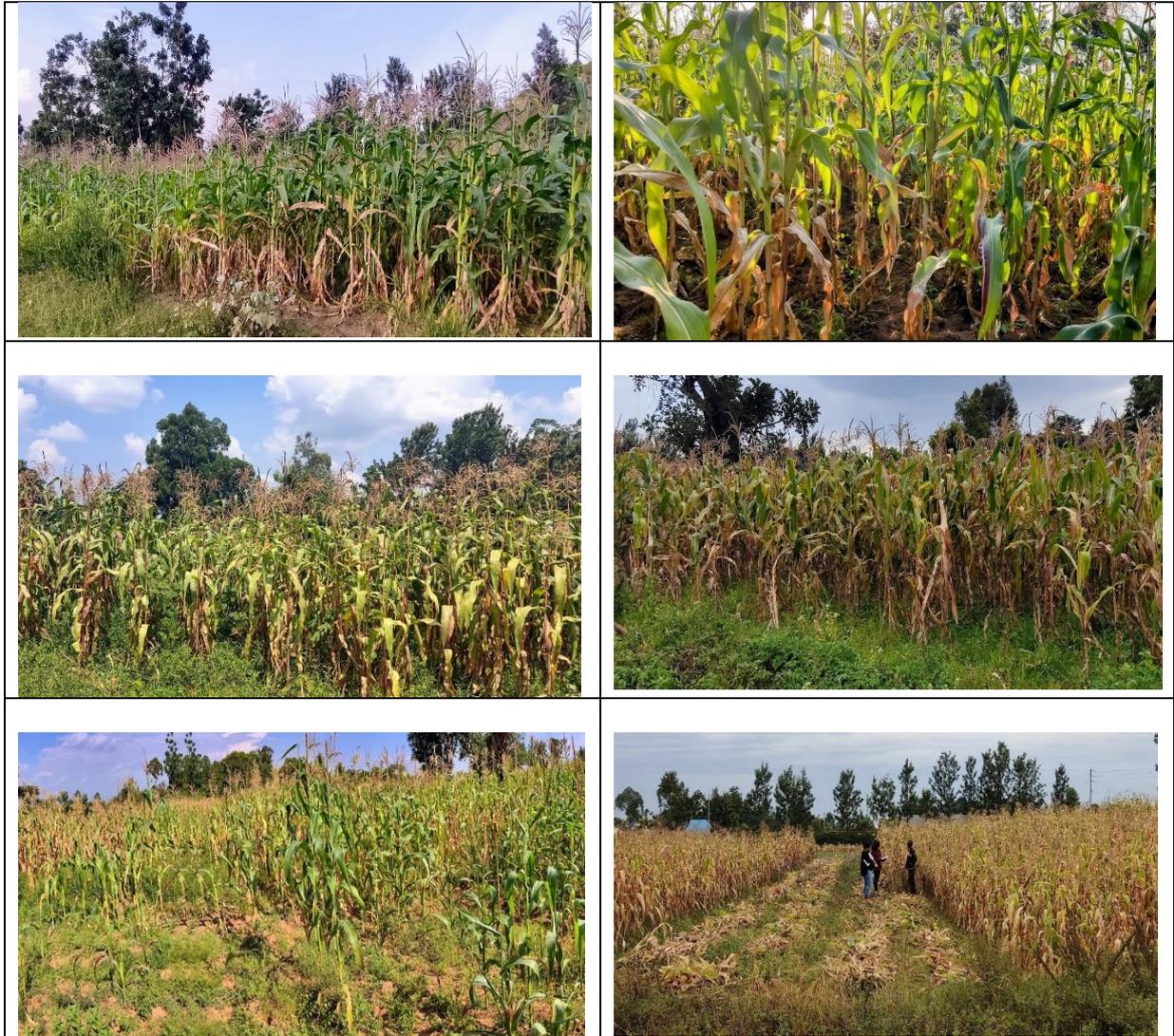
Appendix VIII: Field measurements