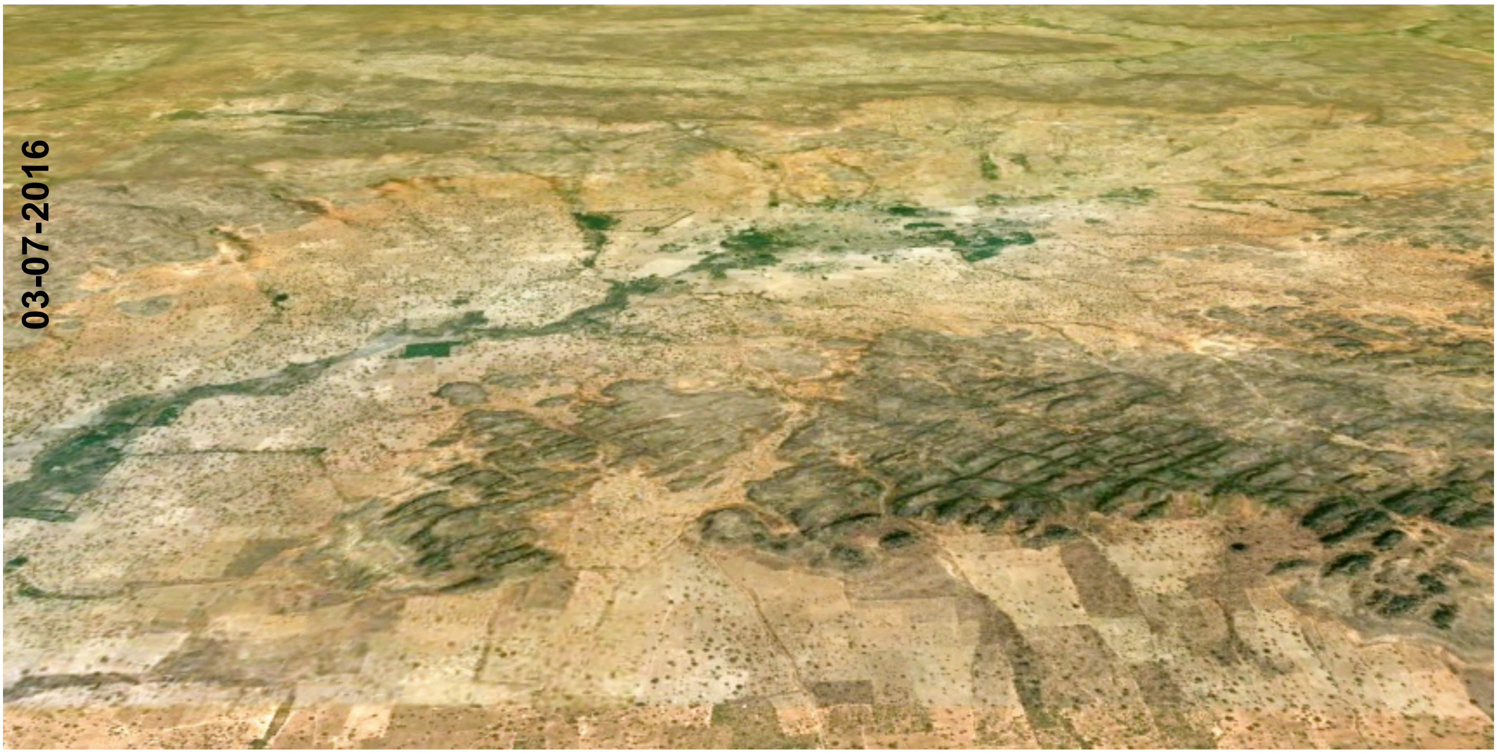
Centre for Geo-Information

Thesis Report GIRS-2016-25

Improving crop classification with landscape stratification based on MODIS-time series

Bart Driessen

920101199030



03-07-2016

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Improving crop classification with landscape stratification based on MODIS-time series

Bart C.H. Driessen

92 01 01 199 030

Supervised by:

dr. Lammert Kooistra

Laboratory of Geo-information Science and Remote Sensing,

Wageningen University, the Netherlands

dr. Antonius G.T. Schut

Plant Production Systems Group,

Wageningen University, the Netherlands

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

3rth of July, 2016
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2016-25
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgments

This work concludes my MSc in Geo-Information Science at Wageningen University.

I want to thank my supervisors dr. Lammert Kooistra and dr. Tom Schut for their continuing support, critical eye and constructive suggestions during the development and planning of this study. I would also like to express my appreciation to the STARS project for making the data used in this study available. In addition, I wish to acknowledge the help in the preprocessing of the satellite imagery employed in this study by the University of Twente.

The process of writing a thesis - on your own, without a practical focus - has not always been easy for me. Therefore, I wish to thank my parents and friends for their encouragement and support throughout my study and throughout my thesis work.

Abstract

This study tests whether stratification based on moderate resolution MODIS imagery can be used as an alternative to stratification based on detailed soil and elevation maps. To this end, the accuracies of methods for stratification and classification of WorldView 2 & 3 imagery were compared for a series of twelve images covering a case study area in Sougoumba, Mali.

Three stratification layers have been constructed: one based on the average, one on the amplitude, and one on the seasonality of a MODIS time series. Subsequently, classification has been performed using various algorithms (RF, SVM, ML, k-NN and multinomial logistic regression) on a training set of 3792 samples and validated using a training set of 1881 samples.

Though all stratification methods proposed have a positive influence on classification results, the methods based on the amplitude and seasonality of MODIS time series yielded the highest classification accuracies. The proposed stratification techniques have a lot of potential for upscaling but some modifications will be necessary to apply them to other areas, especially when moving towards zones with multiple growing seasons per year.

This study shows that MODIS time series can be a very useful tool for developing stratification layers, and may provide a viable alternative to stratification techniques previously tested.

Keywords: crop type classification, stratification, WorldView 2/3 imagery, time-series, MODIS EVI/NDVI

Table of Contents

1. Introduction

1.1. Background

1.1.1. Importance of crop classification in West Africa

In recent years, stratified spectroscopy has shown to be a very useful tool in crop cover mapping (Blaes et al., 2016). The development of stratification procedures is however still in its infancy, and few studies have yet been consecrated to the development of procedures that are suitable for upscaling.

Reliable and timely information on crop area and production is of great importance for decision making by all stakeholders in agriculture. These stakeholders encompass producers, processors, resource managers, marketers, investors and governmental agencies. Due to the globalization of market economies, the availability of this information is more essential than ever before (Dadhwal, 2002).

Policy makers require regional crop production information for socio-economic development planning and the control of macro cropping structure adjustment, but also to ensure food security and to perform land-use planning (Miao et al., 2012). In addition, crop identification on a parcel-scale is of great value when adhering to administration requirements (Garcia-Torres et al., 2015). In an age where food security is challenged by a multitude of factors including population growth, land use changes, water shortage and so on, this information is vital to prevent the oversupply of food - and subsequent disruption of local markets - as well as the undersupply of food that leads to the outbreak of famines (Miao et al., 2012).

Thanks to initiatives such as the establishment of the Land Parcel Identification System (LPIS) in the EU - in light of this project, the EU grants financial aid to producers of certain crops in return for the farmers' declaration of their parcels' area - land cover information is readily available for most Western countries (Oesterle & Wildmann, 2003). However, despite the high demand for such types of information in most developing countries, the data is usually missing, outdated or inconsistent.

1.1.2. Agricultural development of West Africa through time

West Africa is an area that is very complex in terms of geomorphology and demography. The African continent is very old and due to variations in the extent of erosion, different geological units have surfaced in different areas. Political instability and neighboring conflicts have resulted in highly multicultural societies with differences in terms of social status, availability of resources and cultural practices.

Naturally, the area is covered with tropical forests, with trees receding towards the woodlands, savanna, wooded steppe and desert zones (Figure 1). Boundaries of these land cover types have always changed, strongly driven by climate. In West Africa, they vary primarily in function of the rainfall gradient (Bainville & Dufumier, 2007).

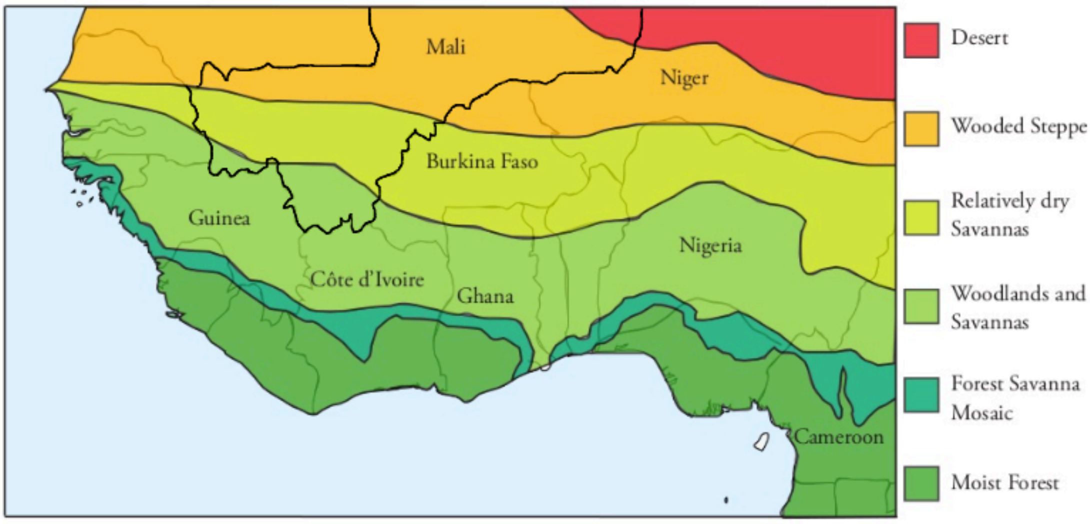


Figure 1. Vegetation map of West Africa (Janssen, 2015). These vegetation patterns are in close concurrence with rainfall gradients. The country in which the case study adopted in this study is located - Mali - is indicated in black.

A few decades ago, the Sudano-Sahelian vegetation zones have experienced a lasting drought, which started at the end of 1960s and culminated in the 1980s, as well as one of the world's highest demographic growths, which has resulted in a threefold increase population during the second half of the 20th century. This strong demographic growth has been supported by an increase in the surface area of rain-fed cultivated lands at the expense of woody savannas (Ruelland et al., 2009).

1.1.3. Current status of agricultural information in West Africa

In most regions in West Africa crop cover statistics are available at a municipality level, but they are not spatially explicit. This is because traditionally, crop areas are reported based on census data, which does not provide information on geographical distribution. In addition, this process is tedious, time consuming and costly (Garcia-Torres et al., 2015). Also, the process is only useful for hindsight evaluation, and cannot act as an early warning system.

Several efforts have been done to create spatially explicit *land cover* information for West Africa; such as the West Africa Land Use and Land Cover Trends Project (USGS, 2016). However, a more detailed assessment of the quantity and spatial distribution of *crops* at a *local scale* is for most regions at present not available.

STARS (Spurring a Transformation for Agriculture through Remote Sensing) aims to increase the availability of these information products using satellite imagery. Their general aim is to “increase the quality, volume and understanding of food production in emerging economies and to improve the farming activities and livelihoods of some of the world’s poorest people” (STARS Project, 2015).

One of the projects of STARS is called STARS-ISABELA *(Imagery for Smallholders: Activating Business Entry points and Leveraging Agriculture)* and is focused on two regions in West Africa: Nigeria (Kofia) and Mali (Sougoumba). This study will be focused on this second region (Sougoumba, Mali) and aims to contribute to the second proposition of this project: the development of digital libraries and algorithms for smallholder crop recognition (Traore, 2014).

1.2. Problem description

West Africa is characterized by a strong in-field heterogeneity due to intercropping with different varieties, crop types and fruit trees (Blaes et al. 2016). The Koutiala district in which the case study of this project (Sougoumba, Mali) is located is furthermore characterized by large geomorphological differences initiated by its location on the border between the Taoudeni Basin and the Dorsale de Léo, impacting crop growth and background reflectance due to the resulting differences in soil type and hence in the spectral signature of crops. On the plateaus, periods of droughts disturb the growth pattern of crops, causing a proportion of the plants to die. Farmers compensate for this by reseeding their fields, resulting in an even higher in-field heterogeneity. This in turn results in a highly mixed spectral signature that contains spectral information on a multitude of crops in various growing stages. Large socio-economic differences within the area and irregular property boundaries further complicate the agricultural pattern (Traore, 2014).

The combination of these factors result in a heterogeneous landscape - on field- as well as on regional scales - where spectral signatures of land surfaces are not only influenced by land use but also by other factors such as parent material and geomorphological structure, which makes it difficult to map the crop types present in the area.

One way of improving classification results is to remove one of these effects – the effect of variability in terms of productivity – by means of introducing a stratification of the area. Productivity is strongly related to geomorphological variability (Blaes et al., 2016), and is therefore an interesting base for stratification.

Stratifying a landscape to account for regional variability is not a new phenomenon. Several studies have previously been dedicated to studying the possible benefits of stratification on classification accuracy (Vintrou et al., 2012; Nelson & Hornig, 1993; Chomé, 2015).

However, previous studies were either characterized by a very coarse spatial resolution or were based only locally available data (Vintrou et al., 2012; Chomé, 2015). To stratify and classify larger areas, it is important to develop automated procedures that are time-efficient and repeatable, and therefore call for a more numerical approach. The fact that these procedures have not yet been developed leaves an opportunity for further research.

1.3. Research objective and research questions

This study proposes a stratification of an agricultural area based on MODIS timeseries in an attempt to improve crop classification accuracy. The underlying hypothesis for basing the strata on plant greenness is that the average landscape productivity over time reflects the underlying geomorphology, and explains the largest amount of regional variability in this area. The general pattern expected is one in which plant greenness increases from the plateau towards the valley.

The main objective of this study is two-fold: to find a way to make a non-arbitrary stratification layer based on plant greenness, and to use this stratification layer to optimize crop classification accuracy.

To reach these objectives, three steps will be taken: i) analyzing the results of conducting crop classification without stratification, which is used as reference, ii) developing a stratification methodology using remote sensing iii) analyzing the effect of stratification on the classification accuracy.

The research questions that follow from this objective are the following:

- What is the effect of stratification on crop classification accuracy?
  - What are the differences between different methods of delineating the strata?
  - What are the differences between the strata individually and which stratum is most easy to classify?
- Are these differences persistent over several years?

To answer these questions, a stratification procedure will be developed based on a MODIS time series, crop classification will be performed using WorldView 2 & 3 data, and the accuracy of the classification procedures will be assessed and compared for a case study area: the area around Sougoumba, Mali.

2. Literature review

2.1. Satellite sensors for crop mapping

In the past decades major steps have been made with respect to sensor development, leading to the availability of the high-resolution imagery needed for automated digital crop mapping. Some of the sensors developed - such as ResourceSat 2 - are even specifically designed to tackle crop classification (EO, 2016).

Most crop classification studies focus on the VNIR-SWIR portion of the electromagnetic spectrum (García-Torres et al., 2015; Jia et al., 2013; Kumar et al., 2015; Miao et al., 2012; Ok et al., 2012; Peña & Brenning, 2015; Tatsumi et al., 2015), though studies utilizing radar data have been conducted as well (Sonobe et al., 2015; Xie et al., 2015). These studies adopt a resolution ranging from 2 meters (García-Torres et al., 2015) to 500 meters (Hao et al., 2015). Though some differences in model performance are visible, all have shown to yield viable classification accuracies - varying from an overall accuracy of 74.99% to 94%.

A study by Dahdwal et al. (2002) has shown that classification accuracy decreases with a coarsening of spatial resolution (Figure 2). This is presumably caused by the fact that spectral mixing increases as the spatial resolution coarsens. However, their study did not take into account sensors with a spatial resolution below 23.5 meters. As the data available for the realization of this study (WorldView 2 & 3) has a resolution at nadir of 1.85 meters, it is precarious to derive conclusions about the kappa coefficient on the basis of this relationship.

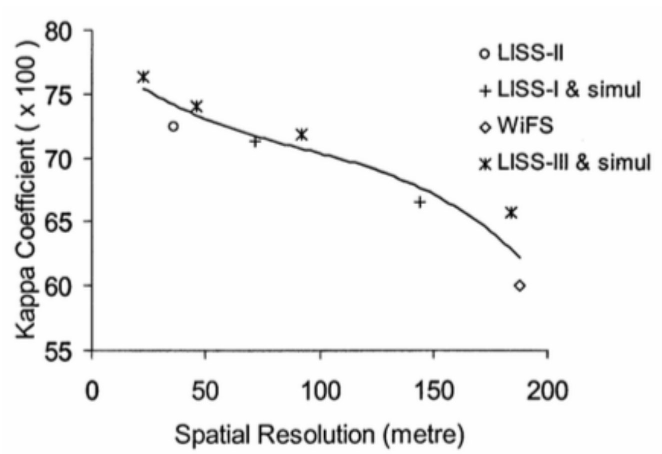


Figure 2. The effect of spatial resolution on classification accuracy (Dadhwal et al. 2002).

2.2. Method development for crop mapping

The parallel progression towards more and more sophisticated statistical- and machine learning algorithms has inspired the scientific community to progressively experiment with maximizing the crop classification accuracy.

The methods used for crop classification today therefore vary tremendously, and encompass differences in the part of the spectrum used (VIS-NIR, microwaves) as well as differences in the adopted classification technique - very generally, one could divide these into highly complex machine learning algorithms such as Applied Neural Networks versus statistical methods such as Maximum Likelihood. To make matters even more complex, these different methods are applied to study areas of various scales at various resolutions.

Statistical methods - such as Maximum Likelihood - are about fitting a model to data, and assume that a process one observes is fully driven by a hidden model that can be excavated. In Machine Learning, it is not assumed that models are equivalent to some hidden background models; both reality and the model are treated as black boxes and the model box is trained in such a way that its output will be similar to that of the reality box. This means that there is no parameter estimation in Machine Learning (Breman, 2001). The concept of not only likelihood but the whole model selection based on the training data is replaced by optimizing the accuracy (whatever defined; in principle the goodness in desired use) on the unseen data; this allows to optimize both precision and recall in a coupled manner. However, this comes with the downside of models often being highly complex and tailored to a single dataset.

A number of studies employing statistical- as well as machine learning algorithms consecrated to crop cover classification have been listed below (Table 1).

Table 1. An overview of a number of important studies on crop classification.

OA = Overall Accuracy, UA = Users Accuracy, PA = Producers Accuracy.

SVM = Support Vector Machines, RF = Random Forest, k-NN = k-Nearest Neighbor, ELM = Extreme Learning Machine - a single layer hidden Neural Network, DT = Decision Trees, OCIM = Object-based Crop Identification and Mapping (= OBIA + DT combined), LDA = Linear Discriminant Analysis, MLC = Maximum Likelihood Classification, SAM = Spectral Angle Mapper, ANN = Applied Neural Networks

Author	***Area***	***Crop types***	***Sensor***	***Method***	***Accuracy***
Garcia-Torres et al., 2015	Cordoba, southern Spain	Broad beans, chickpeas, citrus orchards, cotton, corn, Mediterranean forest, oat, olive orchards, poplars grove, potatoes, summer crops, sunflower, winter wheat, winter crops, young trees	GeoEye-1 imagery (2m)	DT	OA = 80.7%. UA ranged from 0% to 100%, PA ranged from 0% to 100% Oats and citrus orchards both had a UA and PA of 0%,

					which in case of oats was explained by its morphological similarities with winter wheat.
Hao et al., 2015	Kansas, U.S. Central Great Plains	Alfalfa, corn, sorghum, soybean, wheat	MODIS (500m)	RF	OA = 88.45% kappa = ? UA ranged from 80.2% to 95.4%, PA ranged from 81.8% to 99.6%
Jia et al., 2013	North China Plain, Yucheng County, China	Non-vegetated areas, trees, cotton, wheat	HJ-1 CCD (30m)	SVM	OA = 91.7% kappa = 0.87 UA ranged from 72.22% to 97.79%, PA ranged from 63.41% to 96.01%
Kumar et al., 2015	Varanasi district of Uttar Pradesh, India	Corn, linseed, lentil, mustard, barley, wheat, other crops, water, spare vegetation, dense vegetation, fallow land, built up, sand	LISS IV (from Resource-sat-2) (5.8m)	SVM	OA = 93.45% kappa = 0.9216 UA ranged from 58% to 97%, PA ranged from 47% to 98%
				ANN	OA = 92.32% kappa = 0.9124 UA ranged from 48% to 94%, PA ranged from 41% to 98%
				SAM	OA = 74.99% kappa = 0.7062 UA ranged from 14% to 75% , PA ranged from 41% to 82%
Miao et al., 2012	Taigu, in the middle of Shanxi province of China	Wheat, corn, soybean, vegetable, fruit tree, poplar, other-veg, non-veg	Landsat 7 ETM+ (30m)	MLC	OA = 89.61% kappa = 0.85 UA ranged from 71.52% to 98.79%, PA ranged from 74.75% to 95.07%

				SVM	OA = 81.55% kappa = 0.73
Ok et al., 2012	Karacabey Plain, Marmara region, northwest of Turkey	Corn, tomato/pepper, rice, wheat, sugar beet	SPOT-5 (10m)	RF	OA = 85.89% kappa = 0.7977 UA ranged from 66% to 96%, PA ranged from 64% to 94%
				MLC	OA = 77.96% kappa = 0.6782 UA ranged from 63% to 97%, PA ranged from 41% to 97%
Ozdari-Ok & Akyurek, 2015	Karacabey Plain, an agricultural area in Bursa, in northwest Turkey	Corn, tomato, rice, sugar beet, wheat, grassland	Kompsat-2 MS (4m)	SVM	OA = 91.71% kappa = 0.76 UA ranged from 83% to 99%, PA ranged from 85% to 99%
Peña & Brenning, 2015	Maipo River basin, Metropolitan Region, central Chile	Walnut, table grape, almond, European plum	LANDSAT-8 (30m)	LDA	OA = 94% kappa = ? UA ranged from 89% to 98%, PA ranged from 86% to 97%
Peña-Barragán et al., 2011	Agricultural area of Yolo County in California, USA	Alfalfa, almond, walnut, vineyard, corn, rice, safflower, sunflower, tomato, meadow, oat, rye, wheat	ASTER	OCIM	OA = 79% kappa = 0.75 UA ranged from 69% to 100%, PA ranged from 50% to 95%
Sonobe et al., 2015	Western Tokachi plain, Hokkaido, Japan	Beans, beet, grass, maize, potato, wheat	PALSAR	ELM	OA = 79.3% kappa = 0.737 UA ranged from 64.0% to 91.6%, PA ranged from 65.1% to 87.4%
				k-NN	OA = 77.0% kappa = 0.709 UA ranged from

					64.3% to 88.1%, PA ranged from 52.3% to 90.8%
Tatsumi et al., 2015	Ica region, Southern Peru	Alfalfa, asparagus, avocado, cotton, grape, maize, mango, tomato	Landsat 7 ETM+ (30m)	RF	OA = 81% kappa = 0.70 UA ranged from 76% to 98%, PA ranged from 39% to 98%

2.3. Comparison of studies outside of the study area

Some studies tested several classification algorithms in a single study, enabling the possibility to look solely at the effect of the classification algorithm employed.

Ok et al. (2012) were involved in a study in the Northwest of Turkey and classified five crops using two different classification algorithms. The two methods they compared were Random Forest and Maximum Likelihood. Random Forest outperformed Maximum Likelihood in this study. The authors indicate that the promising results of the Random Forest method can be explained by the well-built algorithm of RF, along with their use of parcel-based classification (= dividing the image into homogenous objects using the knowledge of agricultural field boundaries). They also mention that similar results were achieved using different parameter combinations for the RF algorithm, suggesting that the RF algorithm is stable and consistent.

Miao et al. (2012) performed a crop classification study in the Shanxi province of China. They compared MLC and SVM, and found that MLC outperformed SVM (OA MLC = 89.61% vs. OA SVM = 81.55%). The authors classified composites (multi-temporal), and mentioned that for smaller number of images, the SVM method performed best, but that at more than three images, MLC led to higher classification accuracies.

Sonobe et al. (2015) conducted a classification study on six crops in Hokkaido, Japan, comparing MLC vs. k-NN. They found that MLC slightly outperformed k-NN with an OA of 79.3% vs. an OA of 77.0%, but the classification results were very similar.

The studies shown in Table 1 show that the method used is not the only factor of importance, but that the number of ground control points, the study area, the use of multi-temporal imagery or not and the characteristics of the sensor used to classify the image have a strong influence on the prediction accuracy as well. As none of these studies were conducted in Africa and none of these studies employed WorldView 2 or 3, it is precarious to link their results or specific methods to this study. Most similar would be the study performed by García-Torres et al. (2015), who classified among others cotton and corn (two crop types that will be mapped in this study) at a 2m resolution, which is similar to the 1.85m resolution at nadir of WorldView 2 and 3. They reached an overall accuracy of 87% using Decision Trees.

To the author's knowledge, the only team employing WorldView-2 data for crop classification to date is Chellasamy et al. (2014). Chellasamy et al. used Multi-Layer Perception Neural Networks and developed a method to increase classification accuracy by iteratively removing outliers in the training sample dataset, reaching an overall accuracy of 82.3%.

2.4. Comparison of studies within the study area

The study area used in this study has - thanks to the unique dataset collected in this area by members of the STARS project - been subjected to previous research.

One of these studies has been executed by Chomé (2015), who used SVM on a GeoEye-1 image to classify the area of Sougoumba based on five crop types: peanut, millet, corn, sorghum and cotton. He classified the area without a stratification layer as well as with a stratification layer. He found that without stratification, the OA was 67%, the PA ranged from 0 to 87% and the UA from 0 to 80%. Grouping sorghum and millet together resulted in an 11% increase in OA. Introducing a stratification layer led to an increase in prediction accuracy for the valley, though it led to a decrease for the slope and plain. According to Chomé, this had to do with the fact that these areas have a higher variability and that therefore the outliers are concentrated in these two groups. This is amplified by the fact that a lower amount of training samples is located within these two strata. Peanut had a very low prediction performance (0% PA / UA for all methods tested by Chomé); this is partly explained by the fact that only 20 of the 409 training samples used in his study were used for the cultivation of peanuts.

A study by Ommeren (2016) was focused on maximizing prediction accuracy using a variety of classification methods. Using k-NN on a non-stratified single image, he achieved a classification accuracy of 54% in the same study area. Using a multi-temporal PVI composite (combining 10 images collected throughout the year of 2014), this accuracy increased to 65%.

He also assessed the effect of automated stratification procedures on the classification accuracies. The variables taken into account for delineating these strata were elevation, soil type and distance to buildup. Using stratification the single image accuracy increased to 53-56%, and the PVI composite accuracy to 80-81%.

2.5. In this study

Several decisions have to be made in advance when instigating a classification study. These would include the selection of the classification algorithm, whether to use an image composite or a single image and whether to stratify the area before classification.

This study is not focused on maximizing accuracy, but on creating a model that is suitable for upscaling. Therefore, it is important to use a low order polynomial for classification to prevent overfitting. A statistical method is preferred over a machine learning algorithm, as machine learning algorithms assume a black box and are prone to overtraining because of their high algorithmic complexity. The fact that the amount of training samples in this study is limited is yet another reason to choose for a low order polynomial function. This excludes the use of Neural Networks as a classification algorithm.

The vast amounts of classification techniques co-exist for a reason: certain techniques work better in certain regions under certain circumstances with certain datasets. As there is no such thing as determining the best classification method beforehand, several methods will be applied and compared in this study, among which Maximum Likelihood, k-Nearest Neighbours, Random Forest and Support Vector Machines.

The selection of the appropriate date is of vast importance for crop classification. Ozdarici-Ok (2014) showed that combining the images from June, July and August using an SVM classification resulted in an overall accuracy of 91.35% compared to an overall accuracy of 59.61%, 84.12% and 70.54% for the months of June, July and August separately. Combining only June and July had similar results with an overall accuracy of 91.71%.

Hao et al. (2015) found that combining imagery from several months using a RF classification progressively increased the overall classification accuracy up to a combination of images collected over a time period of 5 months; after that, a saturation point was reached at an overall accuracy of 88.45%.

However, for ensuring the success of constructing an image composite the number of available images is crucial. In this study, both the use of a composite as the use of single images will be assessed and compared.

The study conducted by Chomé (2015) shows that introducing a stratification layer certainly has potential for increasing prediction accuracy in this area, but that for strata on the plateau and the slopes more samples are needed and that the strata may need to be positioned slightly different to account for some of the variation within these strata. These recommendations will be incorporated in the design of this study.

To capture the regional variability that this area is characterized by, MODIS NDVI and EVI products from 2001 to 2015 will be used to function as an estimation of plant greenness of the area. These products are then morphed into three different strata using a variety of techniques. MODIS has been chosen for this purpose due to its high temporal resolution - reducing the (disturbing) effects of clouds -, its long, continuous image record and its suitability for upscaling - as it is freely available and covers large areas.

To assess the accuracy of the classification, Congalton (1991) recommends the calculation of a confusion matrix containing the overall accuracy, the user's accuracy, and the producer's accuracy per class. In addition, he recommends calculating the normalized accuracy and the kappa statistic. Congalton notes that a good rule of thumb is to have a minimum of 50 samples per vegetation category in the classification matrix. Also, it is important to increase the amount of samples in categories that show a lot of variation, such as uneven-aged forests or riparian areas. For this study, 250 would therefore be the absolute minimum amount of training samples for classifying five crop types.

3. Methodology

3.1. Conceptual framework

The conceptual model for this research is presented in Figure 3. Four input datasets are used: WorldView-2 and -3, the NDVI and EVI products from MODIS (2001-2015), a training dataset and a parcel database used for validation. For each of the two MODIS products (NVDI and EVI), all dates are grouped together to end up with a single figure per pixel using one of three techniques: (1) calculating the mean, (2) doing a BFAST time series analysis and extracting the average amplitude, (3) determining the average growing season. These measures are then used as base for stratification. Finally, classification is conducted using the WorldView 2 and -3 datasets for each of these strata, and the results are validated using various methods.

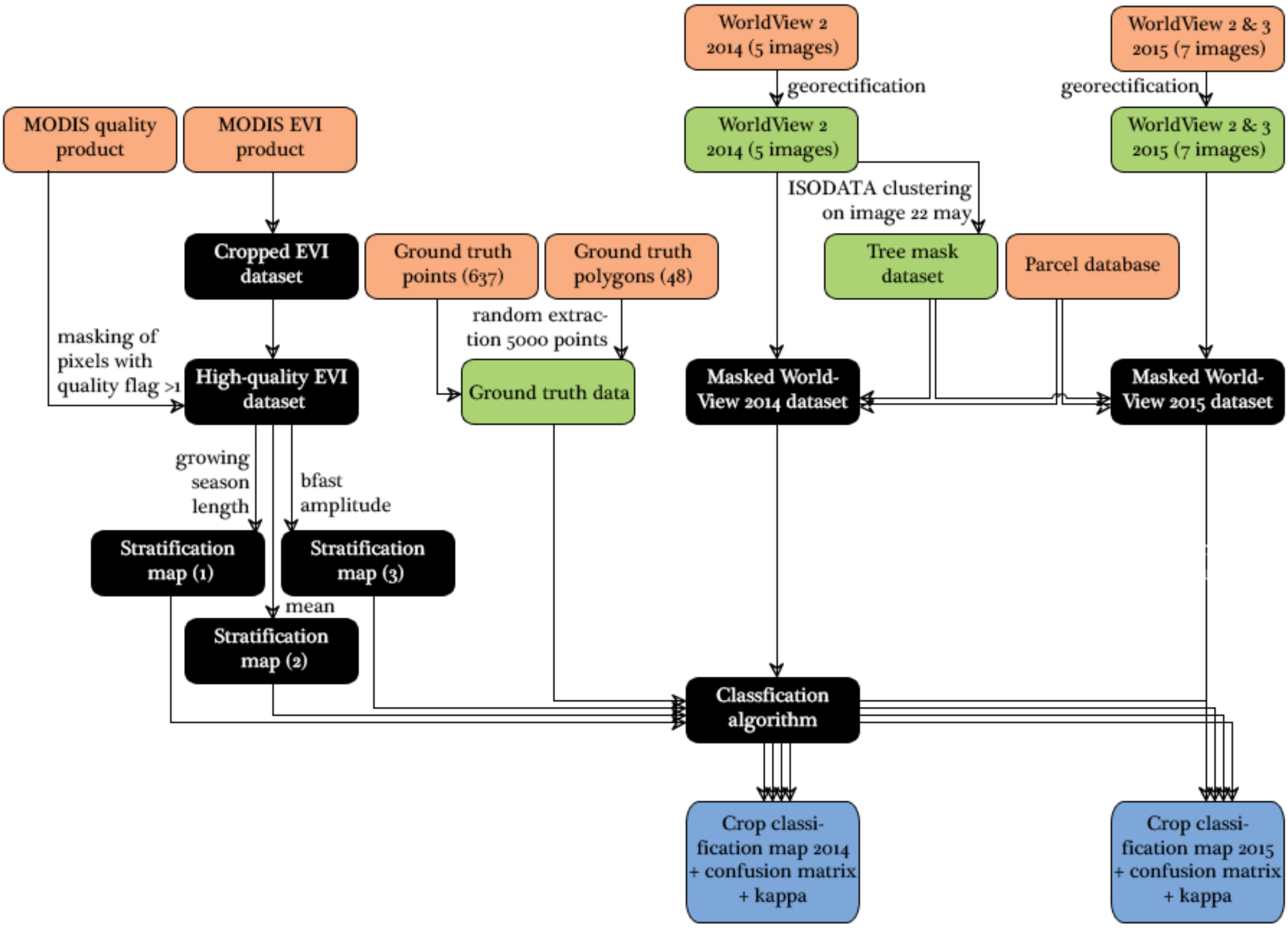


Figure 3. Conceptual model. The pink files are the input files; the green files are the files that serve as a basis for the rest of the procedure; the black files are the temporary intermediate files; and the blue files are the output files.

3.2. Study area

3.2.1. Location, legal status

The area of focus is a 100 square kilometer area around the village of Sougoumba. Sougoumba is part of the Koutiala district and is located in the south of Mali (Fig. 4). The village of Sougoumba has a surface area of about 1 km^2 and is centered around 12°10'20''N and 5°11'20''W. It is situated at an altitude ranging between 400-430 meters above sea level at a distance of about 40 km from the city of Koutiala (Davidse, 2015).

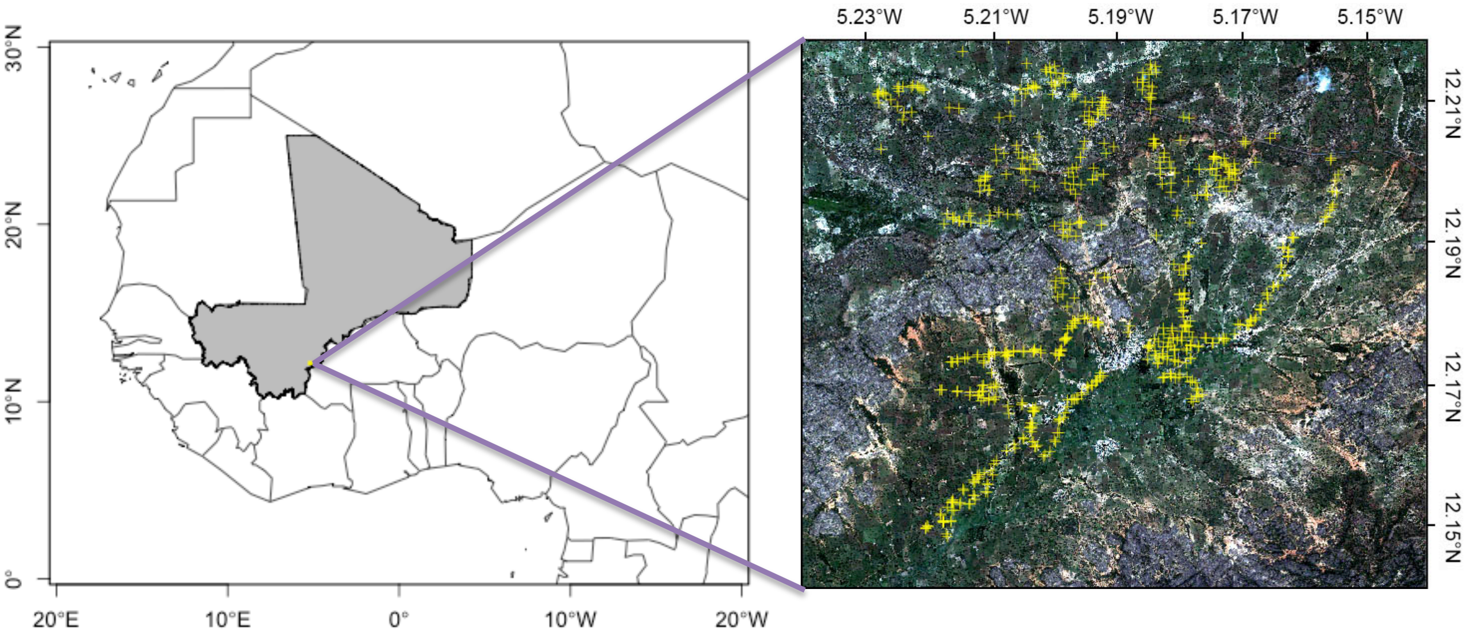


Figure 4. Sougoumba, Mali. On the right you can see the area around Sougoumba overlaid by the location of the training samples (displayed in yellow). Administrative boundaries have been downloaded from GADM (2016).

Mali is one of the fastest-growing countries in the world with a population growth of 3% per year (UN, 2012). The capital city of the province, Sikasso, is the country's second-largest city and is growing rapidly. In 2010, 43.6% of the population lived below the national poverty lines (World Bank Group, 2015). The growing number of people has among others a strong influence on the intensification of agricultural areas.

The area is strongly influenced by its political situation. The region currently hosts a large group of young individuals who returned to Mali from the civil war in Ivory Coast (Schut, 2015). Most of these citizens did not have the means to procure fertile grounds and mostly ended up obtaining land on the plateau or the slopes.

3.2.2. Climate

Sougoumba is characterized by a tropical monsoon climate, and is classified as *Aw* in the Köppen climate classification (Kottek et al., 2006). In this climate zone, the average temperature of the coolest month is 18°C or higher and the winter season is very dry, which is clearly visible from Figure 5 as the winter months of Sougoumba barely receive any rainfall at all. Most of the precipitation in these areas occurs due to convectional thunderstorm activity (Britannica, 2015).

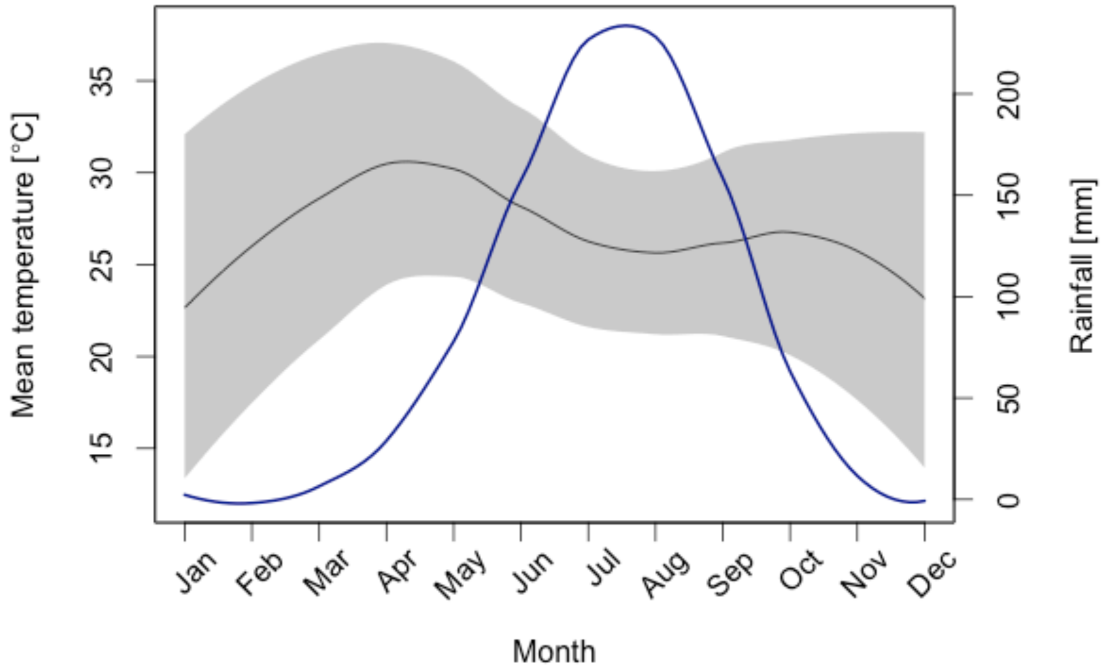


Figure 5. The blue line depicts the average rainfall Sougoumba 1950-2000 for each month. The black line delineates the average monthly temperature in Sougoumba from 1950-2000, and the gray bars indicate the range (lower margin represents the minimum temperature while the upper margin represents the maximum temperature). Source: WorldClim (2016).

The rainfall regime is largely controlled by the Intertropical Convergence Zone (ITCZ). During summer, the Intertropical Convergence Zone moves towards the north and brings convergent and ascending air to these locations, which generates convective rainfall (Britannica, 2015). During the winter season, the ITCZ moves towards the southern hemisphere and is replaced by the subtropical anticyclone, which brings about descending, cold air, resulting in a period of dry and clear weather.

Rainfall events are often intense and local. These very high amounts of water in a very small space lead to exceeding of the infiltration capacity of the soils, causing overland flow and erosion.

3.2.3. Geology and soils

The area is part of the Taoudeni Basin and consists mainly of a Precambrian granatized craton (2700 – 1600 m.y. old) covered by a thin sedimentary blanket; the oldest sedimentary rocks being as much as 1000 m.y. old (Dillon & Sougy, 1974). The Taoudeni Basin was once an epicontinental sea, and has been a depositional area since. Because these sediments are so old, we can expect that many of the clays have already been completely weathered and no longer contain a lot of soluble minerals. Sougoumba is located on the edge of this basin, close to the Dorsale de Léo, an uplifting area with a lot of erosion (Dillon & Sougy, 1974). Because of its location close to the border, we can expect a higher influx of fresh materials, but also a higher vulnerability of soils on slopes to erosion.

The soils of Sougoumba developed on these granite outcrops in a semi-arid tropical climate. There is some clay accumulation in the subsurface layer, and generally the soils are fairly acid. Most soils of this region are acrisols (Chomé, 2015).

The landscape can roughly be divided in plateau, breakaway and valley (Figure 6). The flat plateaus are mostly composed of iron plates on which shallow soils prevail that are vulnerable to water shortage due to the quick runoff as well as water stagnation due to the impenetrability of the parent material. The intermediate slopes are primarily covered by underdeveloped soils containing iron blocks and laterite gravel and are subjected to continuous erosion.

The lowlands and pediplains situated below are covered by a layer of sandy or silty colluvial material on which leached, ferruginous, tropical soils are formed. During the rainy season, water concentrates in the low-lying areas and thalwegs, which therefore feature hydromorphic soils (Bainville & Dufumier, 2007).

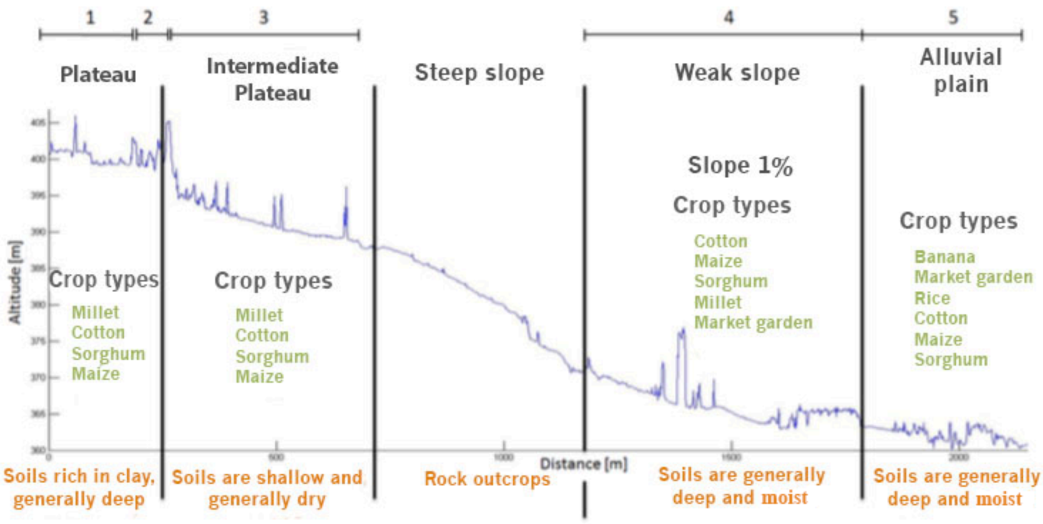


Figure 6. This figure is translated from Chomé (2015) and shows a transect of the area with the different topographical features identified in this area. Trees have not been removed from the transect; they are responsible for the peaks in the profile. The general tendency is that soils in- or adjacent to the valley are deep and humid, while the soils in the intermediate (breakaway) zone are shallow and dry due to the steep slopes and the rocky nature of the topsoil. The plateaus can also be dry due to the

limited water-holding capacity of gravel. However, some of the soils on the plateaus contain clay, which greatly enhances the water-holding capacity of the soils located here.

3.2.4. Site history and disturbances

Until the 1970's, the region of Koutiala was strongly subjected to slash- and burn agriculture. Only fields that were adjacent to the villages and received manure in the form of kitchen waste and small ruminant excretions were cultivated annually. From the 1970's onwards, farmers who engaged in the production of cotton gained access to farmer fundings in the form of seeds, mineral fertilizers and equipment by state companies (Bainville & Dufumier, 2007). This allowed farmers to expand the cultivated surface and to reduce labour time, and resulted in a rise of cotton that permitted farmers to obtain the income necessary to acquire cattle for draught power in lands where livestock farming was until then minor.

The main crops grown in the area today are cotton, maize, sorghum, millet and peanuts. These crops are grown primarily by smallholder farmers (Ruelland et al. 2009) in a rotational system (Blaes et al. 2015), and most of the work on the farms is executed manually. Cotton is generally fertilized, maize is sometimes fertilized and the other crops are generally not fertilized with artificial fertilizers. The fields are commonly intercropped with fruit trees.

3.3. Data use & data preprocessing

3.3.1. MODIS

The MODIS NDVI and EVI metrics were extracted from the NASA Land Process Distributed Active Archive Centre (LP DAAC) for the time period of 2001 to 2015. These products are level-3 products, meaning that they are radiometrically, geometrically and atmospherically corrected (LPDAAC, 2016). The metrics are 16-day composites and have a spatial resolution of 250 meters.

The coordinate system has been converted from sinusoidal to WGS84 using the MODIS reprojection tool. Subsequently, the MODIS products are cropped to extent of study area, which is from 12°8'5"N to 12°13'18"N and from 5°14'9"W to 5°8'36"W.

NDVI values with a quality flag ending with the binary numbers 10 (clouds) and 11 (pixel unreliable) have been removed.

3.3.2. WorldView 2 & 3

WorldView 2 & 3 images have been collected over a 100 square kilometer area around the village of Sougoumba throughout the growing seasons of 2014 and 2015 and have been made available via the STARS project. The extent of this area is 12°8'5"N - 12°13'18"N by 5°14'9"W - 5°8'36"W, which forms the boundaries of the area that will be classified. The spatial resolution of the imagery is 1.85m at nadir for WorldView 2, and 1.24m at nadir for WorldView 3 (DigitalGlobe, 2013; DigitalGlobe, 2014), and the spectral bands are shown in Table 2.

The WorldView images have been preprocessed by Tolpekin (2016). This preprocessing encompassed a standardized orthorectification, atmospheric correction and the removal of a geometrical shift.

The coordinate system has been converted from WGS 1984 UTM 30N (projected) to a World WGS 1984 (geographic). Subsequently, the images have been used as base for classification.

Table 2. The spectral bands of WorldView-2 & -3. The bands of WorldView 2 have a spatial resolution of 1.85m, while the bands of WorldView 3 have a spatial resolution of 1.24m (DigitalGlobe, 2013; DigitalGlobe, 2014).

Spectral band	Spectral region
Band 1 (Coastal)	400-450 nm
Band 2 (Blue)	450-510 nm
Band 3 (Green)	510-580 nm
Band 4 (Yellow)	585-625 nm
Band 5 (Red)	630-690 nm
Band 6 (Red Edge)	705-745 nm
Band 7 (Near-IR1)	770-895 nm
Band 8 (Near-IR2)	860-1040 nm

3.3.3. Parcel database and training data

In total, 637 fields have been identified in the field by Chomé (2015) using convenience sampling and have been stored as spatial point data. All 673 points have been used in the training set. In addition, 48 fields of smallholder farmers have been delineated as part of the STARS project activities in 2014 using quota sampling - in which the idea was that each croptype represented in the dataset was supported by 10 training samples; though two farmers later resigned, hence 48 fields remained - and were stored as spatial polygons. From these 48 polygons, 5000 points have been randomly extracted from a treeless image to strengthen the training set (Figure 7). The parcel database has been obtained from ICRISAT and has been drawn manually using aerial photograph interpretation in 2015.

By combining the polygon- and the point training data, a total number of 5673 training locations have been identified. This set is composed of 1067 millet, 510 peanut, 1160 sorghum, 1545 corn and 1355 cotton training points. The set is thus not completely equally distributed: there are few peanut sites compared to the other spectra. This set has been split into 3/4rth (classification) and 1/4rth (validation).

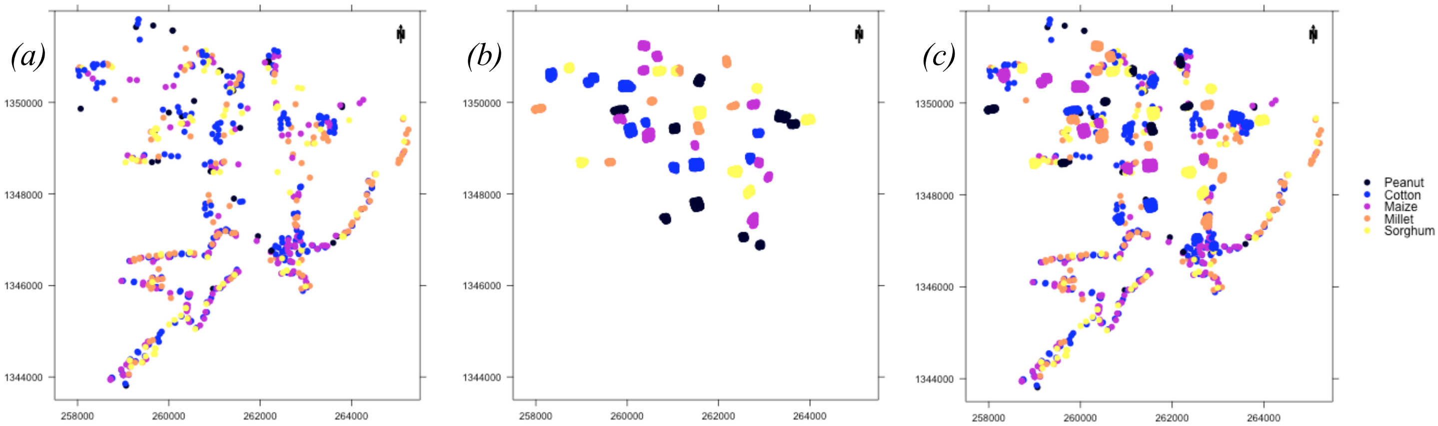


Figure 7. Strengthening the training set by combining the point data collected by Chomé (637 samples) (a) and polygons delineated as part of the STARS project (5000 samples) (b) resulting in a large point cloud (c).

3.4. Stratification procedures

3.4.1. Indices used as base for stratification

Two MODIS datasets have been used for stratification: the NDVI and the EVI.

The NDVI is the Normalized Difference Vegetation Index and is the most extensively used VI. It is - like any other vegetation index - based on the idea that due to chlorophyll absorption within actively photosynthetic leaves, the proportion of reflected red light *decreases* with plant development. On the other hand, reflected NIR light *increases* with plant development due to the scattering induced by reflection and transmission in healthy, turgid leaves (Huete et al. 1999). The NDVI is defined as follows:

$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})} \qquad \textbf{(1)}$$

in which ρ is measured in reflectance.

The NDVI is intuitive and is marked by its simplicity. However, it is shown to be overly sensitive to the brightness of the underlying background. Rain events, litterfall, roughness and the mineralogy and organic matter content of the soil substrate result in spatial and temporal variations in background reflection and hence influence the amount of reflected energy (Huete et al. 1999). Especially during dry seasons this causes problems, because of the increased presence of bare soil and the increased presence of dust.

To correct for the interactive canopy background and for atmospheric influences, Lui and Huete (1995) developed a feedback-based approach incorporating both background adjustment and atmospheric resistance concepts. The result of this was the development of the Enhanced Vegetation Index (EVI), which is calculated using the following formula:

$$EVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + C_1 * \rho_{RED} - C_2 * \rho_{BLUE} + L)} * (1 + L) \qquad (2)$$

in which L is the soil adjustment factor and C_1 and C_2 are two coefficients that control the influence of atmospheric aerosol scattering. The coefficients L, C_1 and C_2 are empirically determined as 1.0, 6.0 and 7.5 respectively (Huete, 1999).

These vegetation time series were compressed to a single figure using of the three methods described below (section 3.4.2, 3.4.3 and 3.4.4). These methods will only be tested on the VI that yields the most promising results.

3.4.2. (1) Mean NDVI / EVI

In this method, a per pixel mean of the NDVI and the EVI was calculated over the time period 2001-2015 using the MODIS EVI / NDVI datasets. These products allow for a temporal resolution of 16 days throughout this time period at a spatial resolution of 250 meters. Afterwards, the histogram of this measure was plotted and thresholds were set on this average greenness map to divide the image in different strata using the k-means algorithm.

This method says something about the *average* greenness over time.

3.4.3. (2) BFAST

BFAST - Breaks For Additive Season and Trend - decomposes a time series into trend, season, and a remainder and provides a tool for detecting and characterizing change within a time series (Verbesselt, 2016).

In this method, the amplitude of the seasonal trend is calculated from the MODIS products using BFAST (Figure 8). Doing this will give an estimate of the total greenness at the top of the growing season. This amplitude is then for each pixel averaged over all years, and this average amplitude is then used to stratify the area. Again, thresholds are set on this measure using the k-means algorithm to divide the image into three distinct strata.

This method says something about the *intensity* of the growing season.

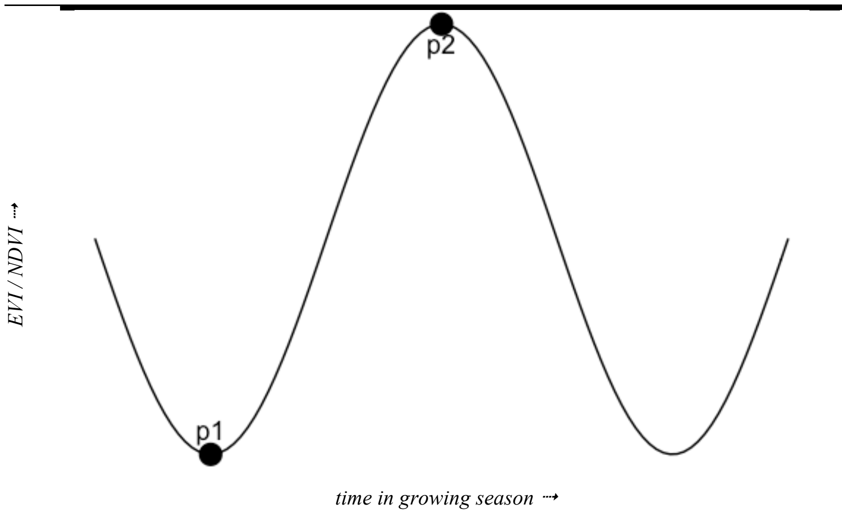
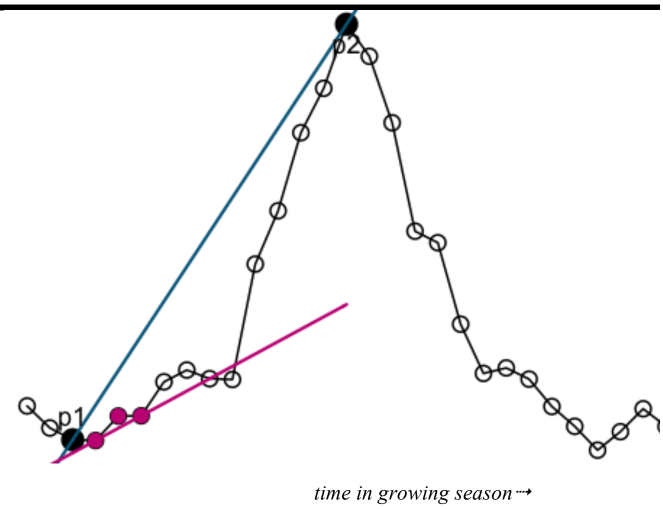
Amplitude from BFAST	*Estimation growing season*
	
$Amplitude = p2 - p1$	The comparison of the slope of a three- time ste window to the slope of point *p1* to point *p2*. As s slope of the three- time step window is steeper tha p1-p2, the window stops moving and the middle p as the start of the growing season.

Figure 8. On the left you can see the calculation of the amplitude from BFAST (section 3.4.3). On the right you can see the estimation of the length of the growing season (section 3.4.4).

3.4.4. (3) Length of growing season

In this method, the length of the growing season is estimated by analyzing a MODIS time series. The start and the end of this estimated growing season are determined by comparing the slope in greenness of a three- time step moving window to the slope in greenness from winter to summer. The first window that has a steeper slope than the winter- to summer slope is taken as the start of the growing season (Figure 8). The end of the growing season is calculated in the exact reverse way: it starts at the end of the next year and starts moving a three pixel moving window backwards, comparing the slope at each time step. Finally, the start of the growing season is subtracted from the end of the growing season to end up with an estimate of the length of the growing season.

This length is then for each pixel averaged over all years, and this average length is then used as a basis for stratifying the area. The threshold used to divide this continuous measure into three nominal classes has been determined using k-means.

This method says something about the *length* of the growing season.

3.4.5. Classification procedures

A variety of classification algorithms will be employed in this study, among which maximum likelihood, multinomial logistic regression, random forest and support vector machines. These methods have been implemented in R using the *rasclass* package, which uses other more specific packages in R (such as randomForest) to classify images using standard settings of classification algorithms. The choice for these methods has been based on what has been done before and on the feasibility of implementation in R by the author of this study.

A total of 3/4rth of the training set - built by combining the STARS data with the data collected by Chomé (section 3.3.3) has been used for the training of the classification models. This adds up to a total of 3792 samples.

WorldView 2 and -3 both have eight spectral bands, and for this study, a total of six images within the growing season of 2014 are available. The classification has both been conducted on single images (using the eight bands of each image) as on NDVI composites (using every image in the growing season in one go).

3.4.6. Validation procedures

Finally, to evaluate the accuracy of the produced crop classification products, a number of measures have been adopted to assess the performance of the algorithms, namely the calculation of a confusion matrix and kappa. In addition, a visual comparison has been conducted.

In total 1/4rth of the training set has been used for the validation of the classification models. This equals a total of 1881 samples.

Confusion matrix

To evaluate the thematic accuracy on a pixel-to-pixel basis, a confusion matrix containing the overall accuracy, the error of omission and the error of commission is calculated.

A confusion matrix gives an overview of the number of pixels that are correctly identified (Strahler et al. 2006). The error of commission - also called the user's accuracy - is the number of correct pixels in a category divided by the total number of pixels that were classified in that category. This is indicative of the probability that a pixel classified on the map actually represents that category on the ground. The error of omission - also called producer's accuracy - is the number of correct pixels in a category divided by the total number of pixels of that category as derived from the reference data (Congalton, 1991). This indicates how well a certain area can be classified.

Though this method gives an estimation of the overall accuracy as well as the accuracy per class, deriving definite conclusions from these statistics remains risky due to the fact that there may be non-thematic errors; for example, when the datasets are not perfectly geo-registered. This is because the confusion matrix tests the agreement with ground data, and not the agreement with the actual ground cover. Errors in training data will in this way protrude into your classification results.

Kappa coefficient

Kappa is a measure of agreement that compares the observed agreement to agreement expected by chance (Congalton 1991).

The result of performing a kappa analysis is the $\hat{k}$ statistic. This is a measure of agreement or accuracy and can be calculated as follows:

$$\hat{k} = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})} \qquad \mathbf{(3)}$$

in which,

r = number of rows in the matrix

x_{ii} = number of observations in row i and column i

x_{i+} = marginal totals of row i

x_{+i} = marginal totals of column i

N = total number of observations

The advantage of $\hat{k}$ over the overall accuracy is that $\hat{k}$ indirectly incorporates the off-diagonal elements as a product of the row- and column marginals (Congalton, 1991). The outcome of kappa determines whether the results presented in the error matrix are significantly better than a random result (i.e., the null hypothesis: $\hat{k} = 0$).

Kappa is a powerful technique due to its ability to provide information about a single matrix as well as to statistically compare matrices.

4. Results

4.1. Preprocessing

In this section, the results of each preprocessing step will be evaluated. These peprocessing steps are georectification, tree- and parcel removal and the combination of the ground truth points and the ground truth polygons.

4.1.1. Georectification

The first step in preprocessing is the georectification of the WorldView 2 and -3 imagery. All images used in this study featured a (small) shift in x and y direction, varying from -8 to +16 in x direction and -14 to +20 in y direction. This problem is illustrated in Figure 9 (a) and (b). The result of the georectification process is shown in Figure 9 (c) and (d). What you can see is that the images have been slightly adjusted in x- and y-direction (an easy way to see this is to look at the tree in the southwestern corner).

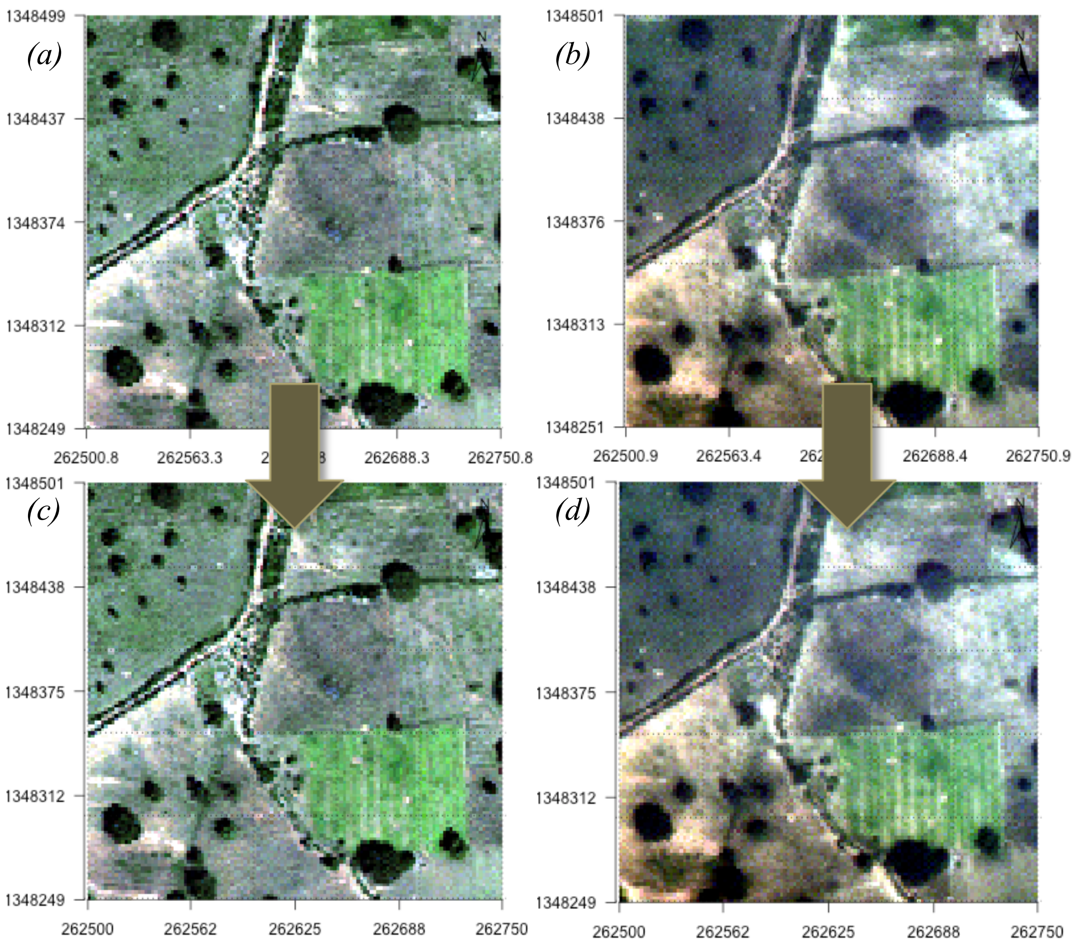


Figure 9. (a) and (b): When looking closely one can see that the WorldView-3 images preprocessed by the University of Twente are not all perfectly aligned. (c) and (d): The result of the manual georectification process performed during this study. For reference, look at the tree in the southwestern corner of this image. The images on the left were collected on 19-10-2015; on the right on 11-10-2015.

4.1.2. Tree masking and parcel mask

The second preprocessing step is the removal of non-cropland areas. This is done in two stages: first, the images have been masked using a parcel dataset; secondly, the trees have been taken out.

The removal of trees has been conducted using k-means clustering from a WorldView-2 image from May 22, 2014. This image was chosen because its point in time is ideal for distinguishing trees from other features, as it is still before the start of the agricultural season. After clustering the image into 20 classes, the classes representing trees have been visually determined and grouped. Subsequently, a buffer has been made around the grouped tree classes, to ensure that no trees or tree shadows would remain in the images.

A more general overview of this process is shown in Figure 10.

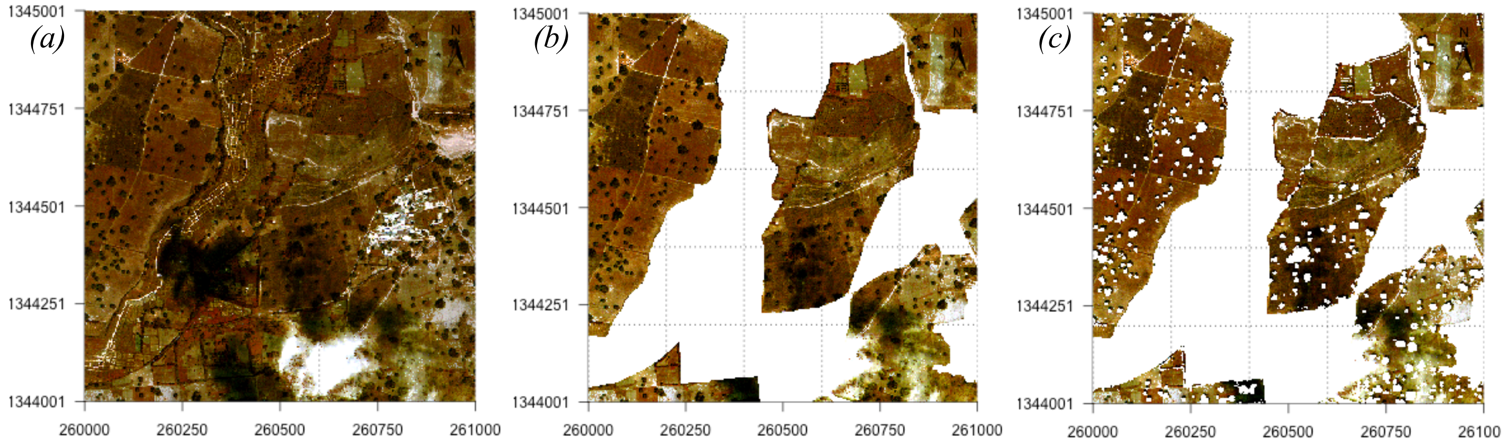


Figure 10. An RGB image of 29 July 2014 for part of the area where you can see the preprocessing step in which non-cropland features have been removed. (a) the input image; (b) the image after applying a field mask; (c) the image after tree removal.

Figure 10 also shows that some cloudy objects and cloud shadows are still visible in the images, which is bound to affect the classification results.

4.2. Stratification

In this section, the results of the stratification procedure are presented. As described in section 3.4, three distinct procedures are used for the derivation of a stratification map. These methods will be quantitatively compared and evaluated in section 4.6.

4.2.1. Mean

The results of the stratification procedure of the *mean* method are displayed in Figure 11 (a) and (b) (page 34).

The density plots show a rather Gaussian pattern, which is not really expected for a region with strong regional variability. You would expect to regional variability back in a density plot in the form of distinct peaks in the density curve, which would indicate natural breaks for a potential division of the area into different strata.

When comparing the images visually with the strata drawn by Chomé (2015) (Appendix B, Figure B1), it becomes clear that the valley and the plateau are somewhat visible but that the intermediate zones are not very clear from these images alone.

The maps indicate that the valley is the area with the highest NDVI and EVI and this is the area that is most intensively used for agriculture, which is in conformity with our expectations.

4.2.2. BFAST amplitude

The stratification maps of the BFAST *amplitude* method are displayed in Figure 11 (c) and (d) (page 34).

The density plots show some more distinctive peaks than previous method, which indicates that this method recognizes more regional variability than the previous one. The density curve mildly points towards the occurrence of three distinct levels, most likely corresponding to:

- rocky features, accommodating little to no green vegetation all year around;
- crops, which vary considerably in greenness;
- trees, whose reduction in greenness in the winter season is far less pronounced but still present

Here, the patterns - valley, breakaway and plateau - as described by Chomé (2015) are lot clearer than they were in the previous method. The areas with the highest amplitude are the valleys and the upper plateau. These are the areas that are most intensively used for agriculture and on which crops generally grow (reasonably) well.

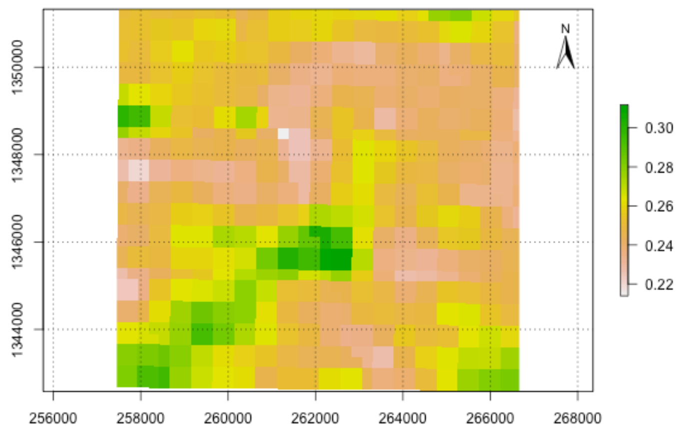
4.2.3. Length growing season

This method effectively aims to estimate the length of the growing season of whatever is growing on a specific point in the landscape. The stratification maps resulting from this procedure are displayed in Figure 11 (e) and (f) on page 34.

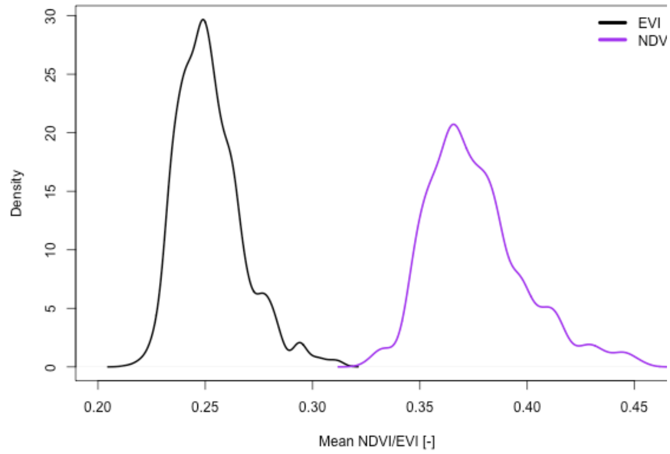
The density plot shows that the difference in growing season is about three to four months. The largest peak in the density plot is most likely caused by the presence of natural vegetation such as trees and shrubs, which have the longest growing season.

Figure 11. The stratification maps resulting from each of the algorithms.

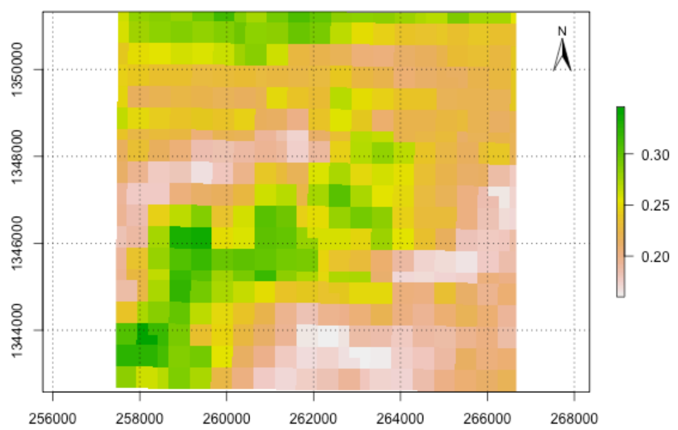
(a) Mean EVI



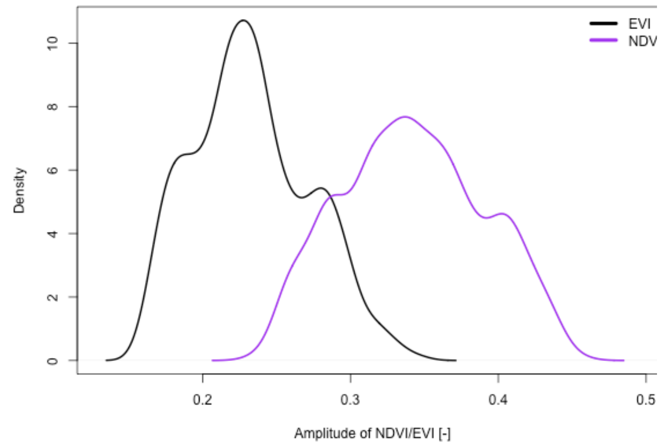
(a) Density plot mean EVI



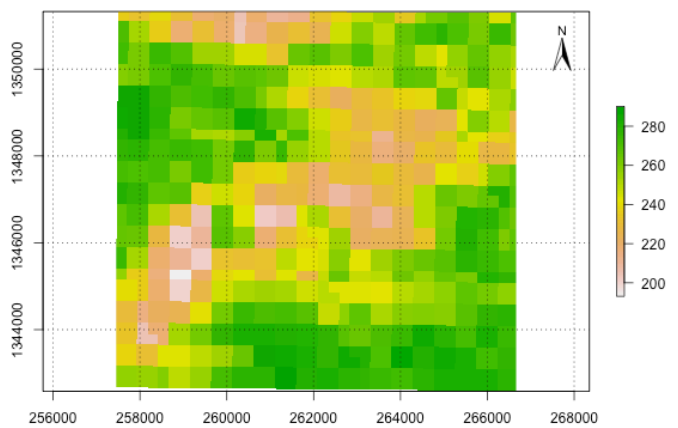
(b) Amplitude EVI



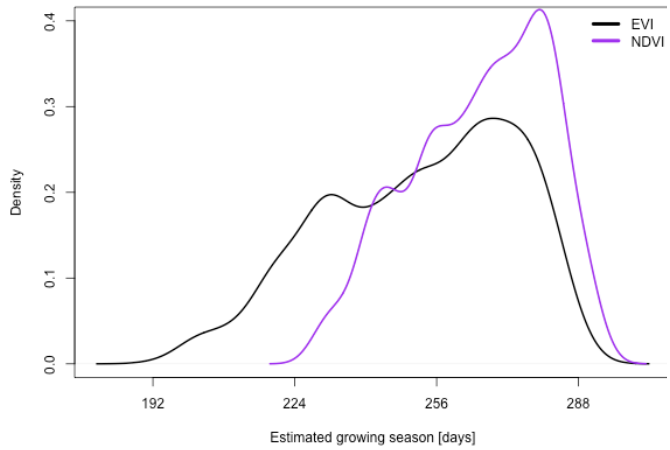
(b) Density plot amplitude EVI



(c) Seasonal EVI



(c) Density plot seasonal EVI



Spatially we can see that the areas with the shortest growing season are the areas in the valley and on the northern plateau. These are the areas that are most used for cultivation, and therefore have the least amount of natural vegetation.

The patterns described by Chomé (2015) are again clearly visible in the stratification maps produced from both the EVI as the NDVI.

4.3. General patterns

From a more theoretical point of view, crops have a short growing season with a high amplitude, meaning that during their growing season they are characterized by a steep red-edge index (which is really what vegetation indices such as the EVI and NDVI are mostly based on). On the other hand, the mean greenness of the agricultural areas would be quite average when comparing them to for example a forest, because forests compensate their lower "greenness peak" with a longer growing season. Natural vegetation on the other hand has a longer growing season with a rather moderate amplitude. The mean is really quite similar to that of agricultural vegetation, because it has a lower difference in red-edge but is growing over a longer time period.

Combining the results of each of these images we can see that the valley really follows these cropland patterns, with a short growing season, a high amplitude and an average mean. This corresponds to reality, as most crops are grown in the valley. The lowest points in the valley could be caused by increased soil moisture that crops cannot deal with, which leads to the planting of the presence of trees (mainly avocados and mangos), which resemble more strongly the pattern of natural vegetation.

The areas identified as breakaway by Chomé (2015) stand out because of their low NDVI and EVI, and are mostly areas where few crops are grown due to the unsuitability of the soils (very rocky; mostly leptosols).

Plateaus are more suitable for agriculture than these breakaways, and are mostly characterized by a somewhat higher amplitude with a shorter growing season. These soils are workable but often less deep than soils in the valley, and are more vulnerable to water shortage in summer and water stagnation during wet periods.

In all cases we can see some spatial differences between the use of the EVI and the NDVI for the construction of a stratification layer (Appendix C), but what this means for the classification accuracy is not clear from just studying the density plots and the maps. The EVI is known to deal better with the exclusion of background reflectance, and in an area such as this one where the background can vary a lot on short as well as on longer distances, we could expect this index to be slightly superior to the NDVI.

4.3.1. Division into strata

The stratification maps have subsequently been divided into three different strata using k-means classification. The end results are shown in Appendix D. Stratum one is somewhat comparable to the valley in most maps, while two and three could be identified as the breakaway and the plateau respectively. Here the differences between stratification methods are even more pronounced: method *mean* results in far less explicit patterns than the other three methods.

Because of the fact that the visual patterns between EVI and NDVI are very similar, the effect of using different strata as a base for stratification on crop classification accuracy has not been evaluated in this study, and for the rest of this study the strata based on the EVI have been used.

4.4. Distinguishing between crops

4.4.1. Temporal profile

Figure 12 shows the progression of the NDVI in 2014 for each crop type. What we can see is that the NDVI is very similar between crops and that the standard deviation is very high. Maize and millet are most impalpable; this is something that Chomé (2015) had already pointed out, which is why in his study he grouped those two together into a single category. It is also important to note that here the differences in discernibility of crops throughout the growing season are already very apparent; e.g. clearly, the best time slot to set sorghum apart from the other crops tested in this study would be the end of July / start of August.

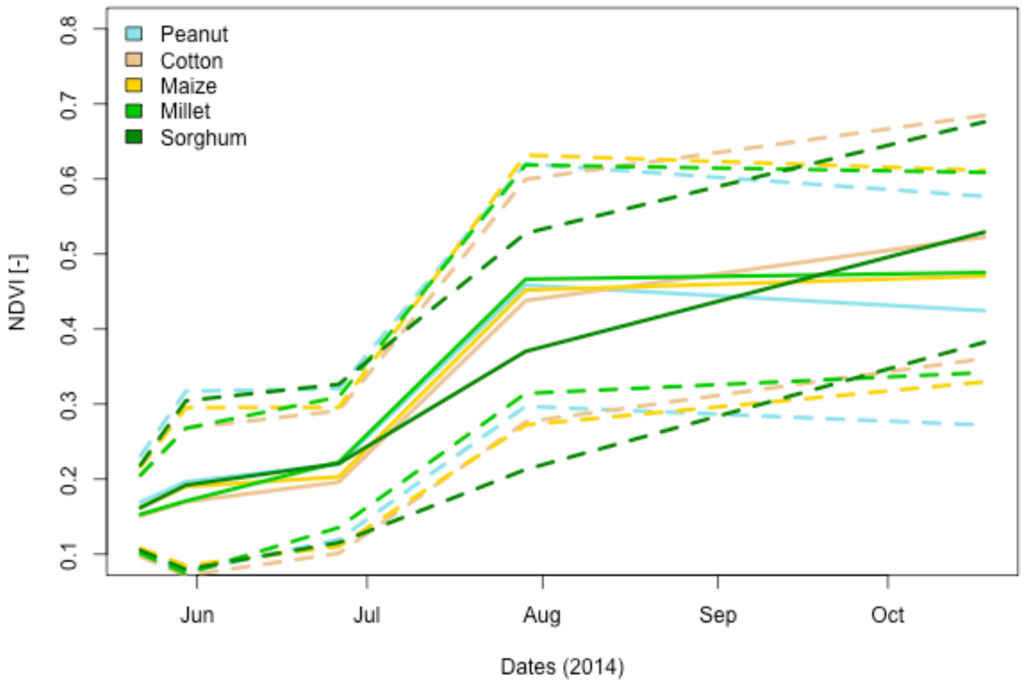


Figure 12. The temporal profile of 2014 for each crop. The dotted lines represent the standard deviation.

This study is based on the assumption that there is a difference between strata and that therefore stratifying the area could help in distinguishing between crop types. To acquire a preliminary idea of whether this assumption holds, the seasonal NDVI profile has also been made per stratum for each crop type separately. One of these profiles (cotton) is shown in Figure 13. Here one can see that the valley clearly stands out. This was only the case for cotton, maize and millet; not for peanut and sorghum. Appendix F gives an overview of the seasonal profiles of the other crops.

Though this clearly indicates the unique behaviour of the first stratum, the differences in vegetation greenness between stratum 2 and stratum 3 are marginal.

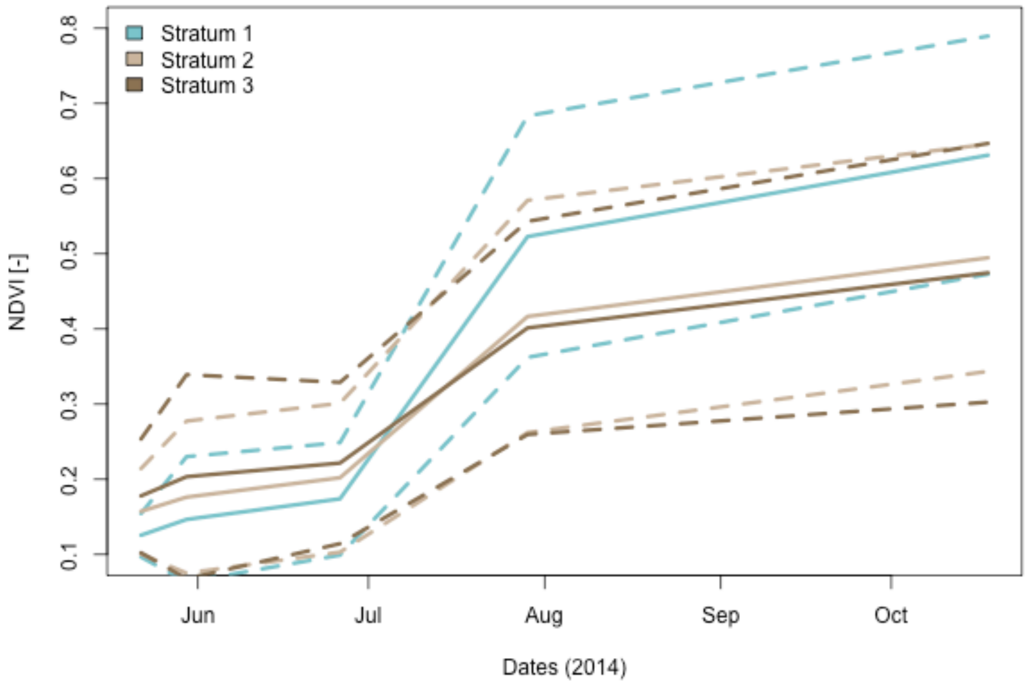


Figure 13. The temporal profile of 2014 for cotton per stratum. The strata are based on the *Amplitude* method with an EVI base.

4.4.2. Spectral profile

Figure 14 gives an overview of the spectral profile of each of the crops classified in this study. Again, the resemblance of the different curves is striking. In this case, this similarity is clearly caused by the influence of the soil. The profiles look more like the spectral profile of a soil than like the spectral profile of a crop: the red-edge is really not steep at all. This is caused by the fact that in this region (Sougoumba), the maximum soil coverage in cropland areas is generally between 70-75% and strongly varies per region. This really complicates the classification procedure.

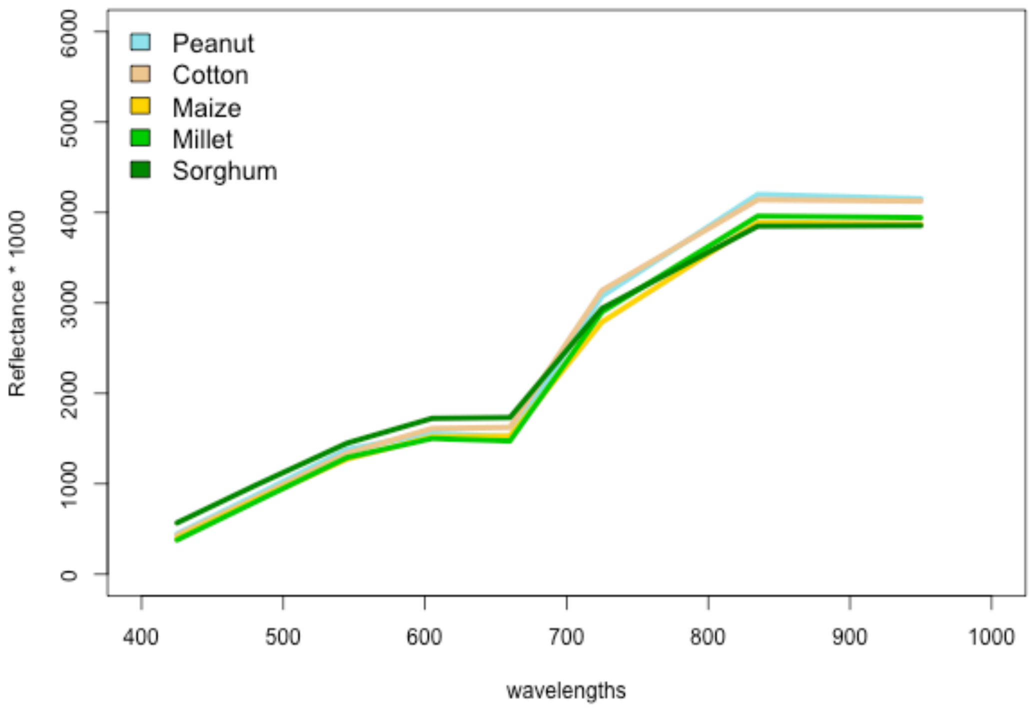


Figure 14. This image shows the mean spectral profile of the 5 different crops classified in this study for the image of the 29th of July, 2014.

As previously expected, the most important bands for distinguishing different crops would be wavelengths around the red-edge.

4.5. Non-stratified classification

4.5.1. The use of single images

A total of seven preprocessed images were used for 2015 and five for 2014.

Figure 15 shows an overview of the non-stratified single image classification accuracy for each of the images collected throughout the growing season. We can see that the images of 2014 have very similar overall accuracies, but that in 2015 there are some more differences between images.

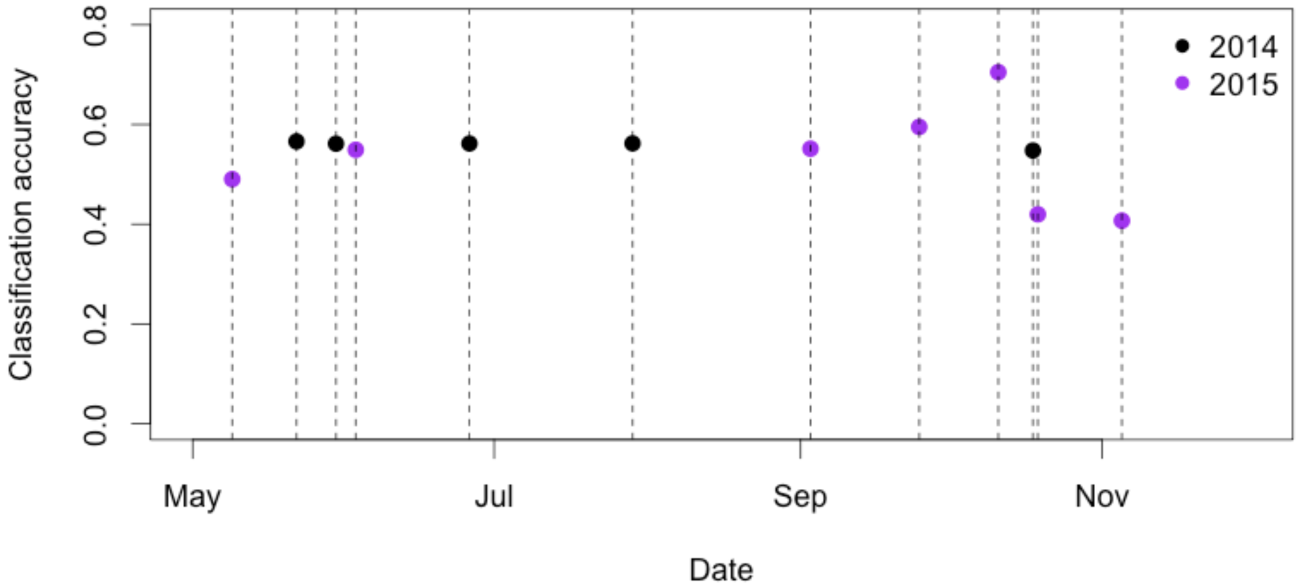


Figure 15. An overview of the non-stratified single image classification accuracy for each of the images available in this study using the Random Forest algorithm.

The fact that the images of November and the end of October 2015 are characterized by a low classification accuracy is not surprising, as at this point (sometime during October) most crops are harvested. In a paper by Blaes et al. (2016) it is noted that for this area, the best time to spectrally distinguish between most crops is in September - right before harvest, that is. This overview mildly supports that notion, though it is difficult to identify the exact period due to the image gap in September.

At the end of the growing season, the classification accuracy starts to vary a lot. This has probably something to do with harvesting; the fact that the image halfway October is performing so well is likely to be influenced by the fact that during this time, early crops have already been removed from the fields, while other crops are still in the field. You might thus be comparing one or two soil spectra to the other crops, which are obviously very spectrally different.

4.5.2. The use of composite images (multi-temporal)

When looking at the difference between using a single image and combining all images of a year of choice using the Normalized Difference Vegetation Index (NDVI), it turned out that using a single image resulted in a higher classification accuracy (Figure 16). For 2014, this difference in overall accuracy was 0.46 for the NDVI composite vs. 0.56 for a single image collected on the 29th of July, while for 2015 this difference was 0.56 for the NDVI composite vs. 0.70 for the image of 11th of October.

This indicates that spectral resolution seems to be of higher importance than spatial resolution. An improvement for further studies would be to study the effect of indices that capture more spectral variability than the NDVI (such as the EVI) on the overall classification accuracy. Ommeren (2016) showed that the NDVI performed worst compared to other indices; however, in his study, even the NDVI lead to an improvement of the classification results. This can be attributed to the fact that he used ten images for deriving a composite, while for this study only five images were used.

An extra observation one can make is that stratified- outperforms unstratified classification in all four cases.

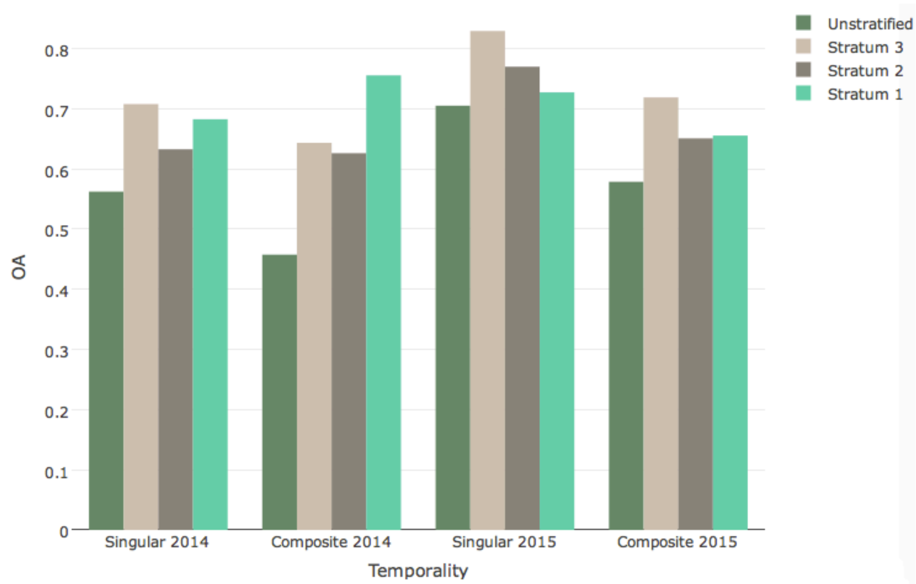


Figure 16. A comparison of the overall accuracy of single images (29-07-2014 and 11-10-2015) vs. the multi-temporal NDVI composites. The stratification method used for these calculations is *Amplitude*, and the classification algorithm is *Random Forest*.

4.5.3. Assessing the effect of different classification algorithms

When comparing differences between different algorithms, we can see that especially Random Forest and Support Vector Machines perform slightly better than the other techniques used (Figure 17).

However it also becomes clear that the most accurate method really depends on the image on which classification is performed.

To test the other factors discussed in this study, the Random Forest method is used. This is due to its higher suitability in terms of upscaling compared to SVM.

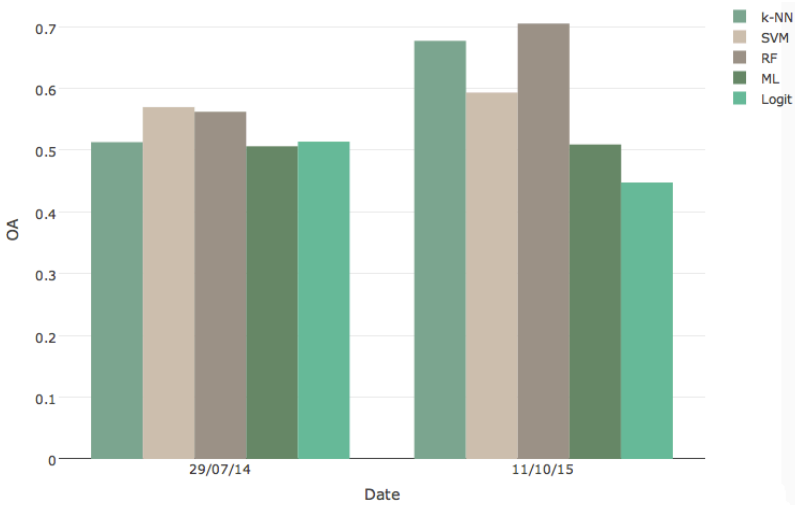


Figure 17. Differences between non-stratified classification accuracies of five algorithms for the dates 29-07-2014 (left) and 11-10-2015 (right).

4.6. Stratified classification

4.6.1. Spatial distribution of training data over the strata

Figure 18 displays the spatial distribution of the training data over one of the stratification layers (*Seasonal*). Some things to note about this are the fact that the valley (stratum 1) contains a lot of training data (thought mostly points), while stratum 2 has few training samples. Stratum 3 contains a moderate amount, but the training points are mainly derived from the training polygons in this area, and also the data is really concentrated on the western part of the study area. The southeastern part of the study area is covered by no training samples at all.

In addition, some of the fields classified in this study are located on the border between two strata. This is for example the case for the two maize fields in the northwestern most corner of the study area.

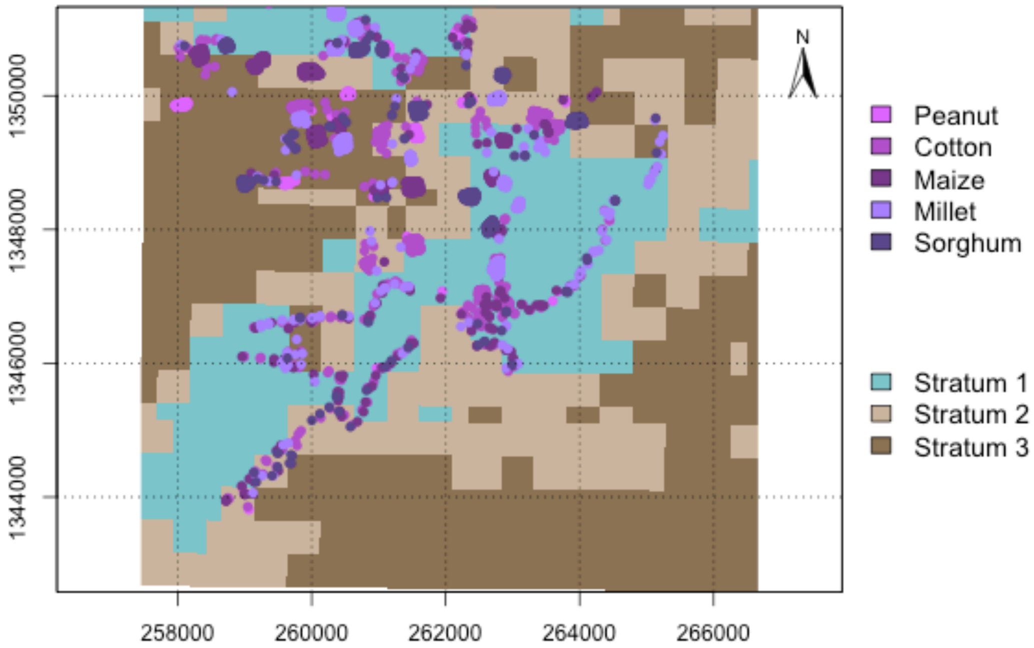


Figure 18. Distribution of samples over the different strata. The strata background layer is made using the *Seasonal* EVI method.

4.6.2. Difference between algorithms

Three different techniques for stratification are used in this study. The resulting stratification maps are shown in Appendix D and have previously been discussed in section 4.2.

Figure 19 shows the difference in overall accuracy between these different methods. What we can already see is that there are differences in strata for the method that is most suitable. Overall, *mean* performs slightly worse than the other methods. Stratum 1, corresponding to the valley delineated by Chomé (2015), performs especially worse for the *mean* method. The other two techniques perform quite well, especially when comparing them to the OA achieved on the same image without stratification (0.56).

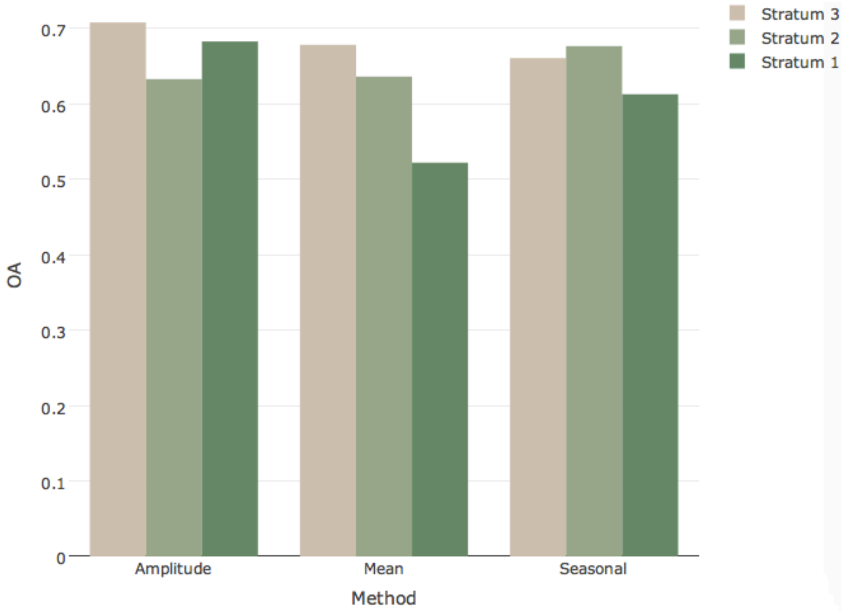
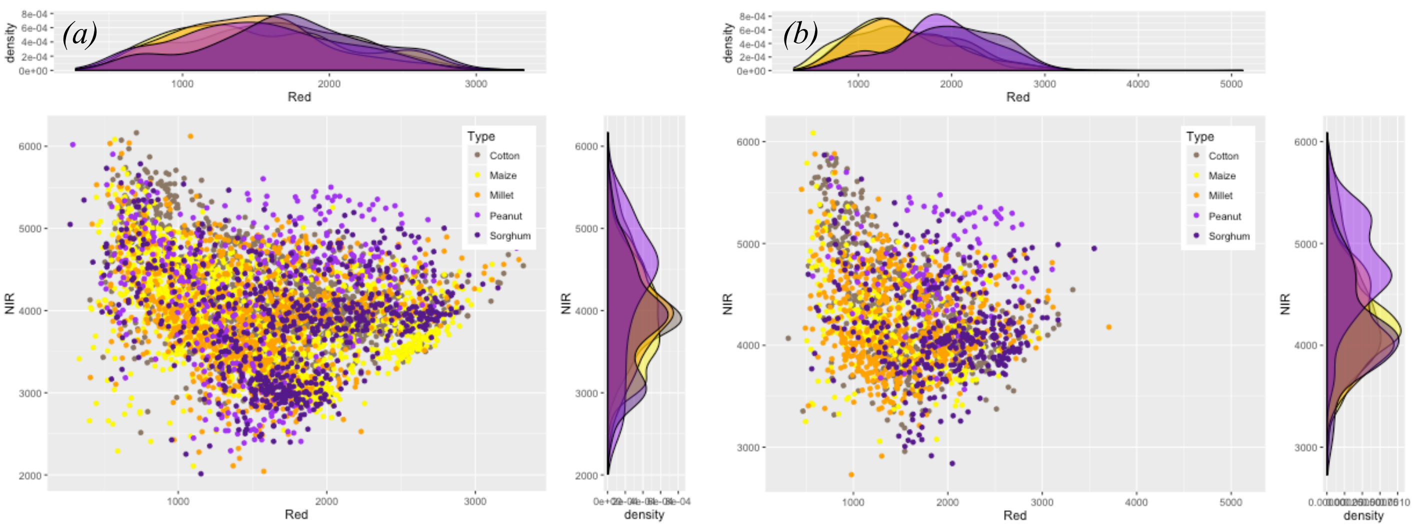


Figure 19. The different classification methods and their corresponding overall accuracy. This figure is the result of an EVI base on the image of the 29th of July, 2014.

4.6.3. Differences between classes

Figure 20a shows a non-stratified feature space plot for an image collected on 29-07-2015. It is very clear that there is a lot of overlap between the five different crops, making it difficult to differentiate between them.

Figure 20b, c and d show the feature space plots after stratification. Though some overlap is still there, it is visibly less than it was before stratifying the area.



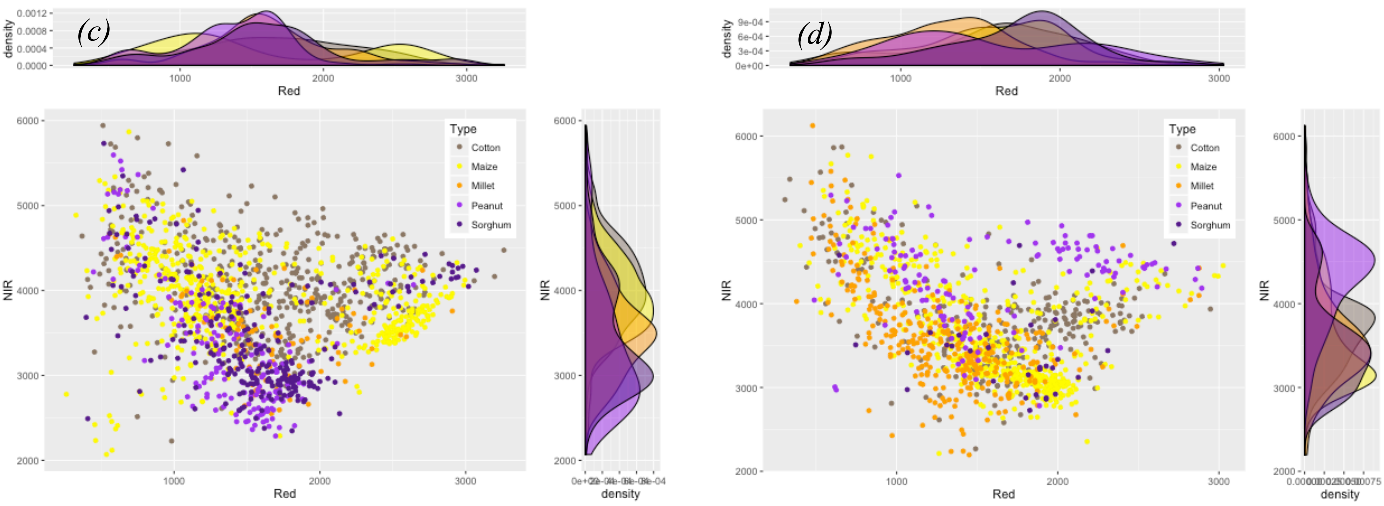


Figure 20. Feature space plots for the image taken on the 29th of July 2014. **(a)** non-stratified; **(b)** stratum 1 (valley); **(c)** stratum 2 (breakaway); **(d)** stratum 3 (plateau). The stratification method for these strata is *Seasonal*, which is based on the length of the growing season.

The problem with some of the strata is that the training set per stratum is a lot smaller than it is for the non-stratified classification procedure, and that the training samples are not evenly distributed over the strata (section 4.4.1). This means that for some strata, only a few points are available for training, which leads to inaccurate classification results.

From the feature space plots you can already see that cotton is even in the strata mostly hidden behind the other crops. Also, one can see the spectral similarity (in NIR and Red, at least) of sorghum and millet. Previously research by Chomé has shown that grouping these two classes together leads to a great improvement of classification accuracy.

Peanut is a difficult class to map in general. This is because it is characterized by a very short growing season and because peanut fields are often planted with other crops, such as pumpkin, afterwards. Also, peanut fields often contain a lot of weeds.

This primarily leads to issues when using images late in the season or using composites as classifier.

Table 3 shows the users- and producers accuracy per crop for different classification techniques (non-stratified vs. stratified using the *Seasonal* method). This table shows that the highest accuracies are found in the stratified areas, though individual strata sometimes perform less well for certain crops. The problem of an uneven distribution of the training samples is here once again shown: in stratum 3, less than twenty samples for validation were available for sorghum, which partially contributes to the particularly low accuracies for this crop type in this region.

Table 3. The confusion matrix resulting from the classification of a non-stratified single image collected on 29-07-2014. The highest users- and producers accuracies per crop are highlighted.

	Non-stratified		Seasonal, stratum 1		Seasonal, stratum 2		Seasonal, stratum 3	
	UA	PA	UA	PA	UA	PA	UA	PA
Cotton	0.55	0.38	0.48	0.37	0.68	0.5	0.84	0.53
Maize	0.61	0.68	0.72	0.75	0.64	0.74	0.56	0.5
Millet	0.62	0.71	0.64	0.39	0.71	0.81	0.70	0.83
Peanut	0.43	0.45	0.54	0.68	0.78	0.4	0.6	0.65
Sorghum	0.41	0.32	0.60	0.54	0.65	0.57	NA	0

4.7. Spatial patterns

Figure 21 displays two of the maps derived from this classification procedure. The general patterns seem similar, but the stratified maps are clearly different from the non-stratified maps, especially when it comes to the distinction between cotton, maize and millet.

Also, you can see the borders of the strata back in the image (e.g. in the upper left corner, at about *(258000, 1352000)*.

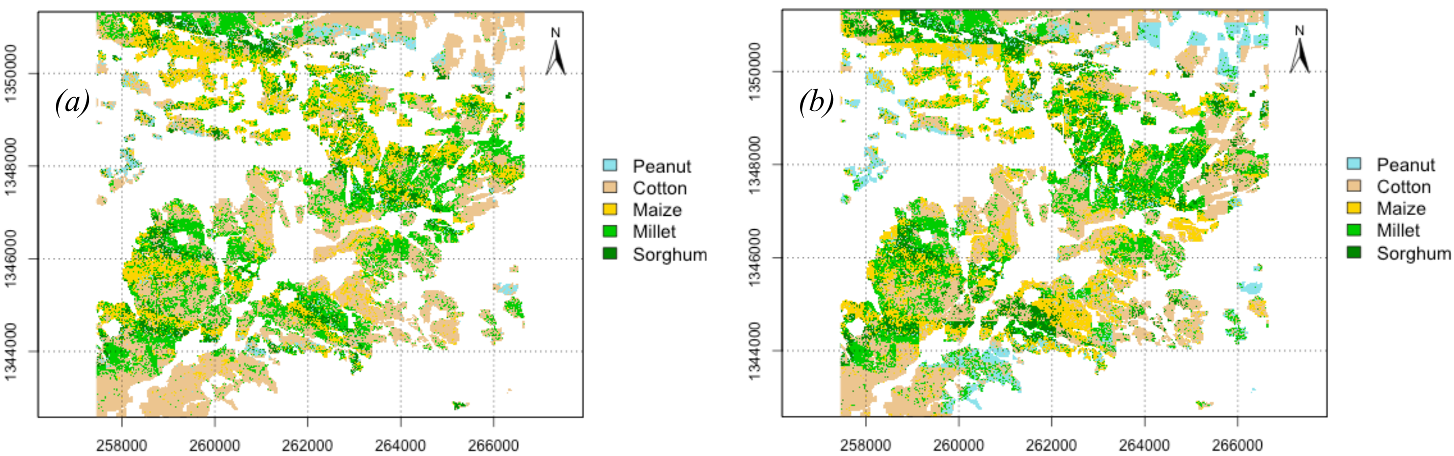


Figure 21. The maps resulting from a (non-stratified) Random Forest classification of the image of the 29th of July 2014 (a) and a stratified Random Forest classification of the image of the 29th of July 2014 using the Amplitude method described in section 3.4.3 (b).

4.8. Comparison to previous studies

Chomé (2015) has employed the exact same study area and a similar data source (GeoEye-1) to perform classification, and also implemented a stratification mechanism for classification based on manual identification of three different strata: valley, breakaway and plateau.

He also used three classes to classify the area and achieved accuracies of 0.67 (non-stratified) and 0.71 (stratum 1), 0.51 (stratum 2) and 0.59 (stratum 3) using SVM.

Ommeren (2016) also conducted a study regarding the use of stratification for improving classification accuracy in Sougoumba, Mali. In his study, stratification was based on a DEM and a soil map of the area.

Using k-NN on a non-stratified single image, he achieved a classification accuracy of 0.54 in the same study area. This is slightly higher than the k-NN classification accuracy obtained in this study (0.51), which can possibly be attributed by differences in training set used and differences in preprocessing of the imagery (orthorectification and tree removal).

Contrasting to the findings of this study, making an image composite increased classification results in his study. This can possibly be attributed to the fact that he used different composite measures (PVI and TSAVI) than the one adopted in this study (NDVI), and the fact that he had ten images available throughout the growing season instead of five.

5. Discussion

In this chapter, the research questions are revisited and the limitations of this study are discussed. In addition, the possibilities and obstacles for upscaling are discussed.

5.1. Revisiting research questions

What are the differences between different methods of delineating the strata?

Three methods were tested in this study. The first one, using the per pixel *mean* NDVI of the MODIS time series, focused on simply capturing the average greenness present in a certain area over a long time period. This stratification method improved classification accuracy from an overall accuracy (OA) of 0.56 (non-stratified) to an average of 0.61 (stratified) using the Random Forest (RF) classifier (Figure 16).

The second method, based on the *amplitude* of the NDVI of the MODIS time series, focused on capturing the intensity of the growing season in terms of the increase in NDVI from the annual lowest to the annual highest value. This method improved the classification accuracy from an OA of 0.56 (non-stratified) to an average of 0.67 (stratified) using a Random Forest (RF) classification for an image of the 29th of July, 2014 (Figure 16).

Finally, the third method of stratification focused on capturing the length of the growing season. This method improved the classification accuracy from an OA of 0.56 for non-stratified- to an average of 0.65 for stratified classification, using the RF classifier for an image of the 29th of July, 2014 (Figure 16).

Though the overall classification accuracy differed between images, the general pattern in which stratified classification outperformed non-stratified classification remained the same both for single images as for composites (Table E9, Appendix E). However, it is clear that some stratification techniques performed better than others. In particular, the latter two methods have higher overall accuracies and less fluctuation between the strata (Figure 16).

Compared to the literature listed in chapter 2 (with a classification accuracy varying between 0.75 to 0.94), the classification accuracies achieved in this study are somewhat lower. This is partly explained by the complexity of the study area and the fact that there has been no fine-tuning of classifier parameters due to the overall goal of increasing the upscalability of the methodology proposed. Previous studies in this study area have achieved similar classification accuracies (Ommeren, 2016; Chomé, 2015) and have also concluded that stratification techniques have a positive influence.

Ommeren (2016) achieved an increase in OA from 0.54 to 0.56 with a single image k-NN classifier using a stratification based on soil type. In this study, the single image overall accuracy increased from 0.56 to 0.67 using the RF classifier (amplitude method). Using a PVI composite in 2014, Ommeren (2016) achieved an OA of 0.65 (non-stratified), which increased to 0.81 when using soil strata. In this study, the NDVI composite of 2014 increased from 0.46 (no strata) to 0.67 (amplitude strata). This indicates that stratification based on MODIS provides a promising alternative for stratification based on ancillary data such as soil maps.

What are the differences between the strata individually and which stratum is easiest to classify?

Three strata were defined in this study and subdivided the full area into strata for valley, breakaway and plateau areas. The average OA classification accuracies for these three strata were 0.60, 0.65 and 0.68 respectively using Random Forest for an image recorded on the 29th of July 2014 (Figure 16). The first stratum - corresponding to the valley stratum - performed poorest due to the low classification accuracy in this stratum in the *mean* method (0.52).

Accuracy of classification per crop type was influenced by the number of training samples within the stratum, which differed between crop types. For example, cotton was best classified in stratum 3 (the plateau), while for maize stratum 1 (the valley) resulted in the highest classification accuracies (Table 3).

All stratification methods had a positive effect on the classification accuracy in this study area. The stratification method resulting in the highest classification accuracies was based on the *amplitude*, closely followed by the *seasonal* method. These results were expected as both are measures of productivity, which is strongly related to crop type.

Are these differences persistent over several years?

In this study, the overall classification accuracy of a non-stratified RF classification of a time-series of 12 images was assessed. The results indicated that the differences *between* years were small compared to the differences between images taken *within* one year. The problem with directly comparing images to truly evaluate small differences between different years is that the imagery was not collected at the exact same date under the exact same weather conditions, which makes direct comparison and the extraction of specific patterns between strata challenging. In addition to that, seasons are different each year, which again influences crop development and growth.

Of course differences in classification accuracy between years can be expected, because of differences in weather conditions and management practices. Years with a dry and hesitating start (such as 2015) may contain more mixed spectra than a wet year such as 2014, because plant emergence may be irregular, causing plants to die and resulting in farmers reseeding parts of their fields during the growing season. However, these differences were not significantly affecting the results in this study.

Nevertheless, stratification had a positive effect on the classification accuracy of the images tested for 2014 as well as for 2015. This is in agreement with the findings of Ommeren (2016), who also found that stratification based on auxiliary soil data improved classification results for 2014 as well as for 2015.

5.2. Bottlenecks in this study

5.2.1. Stratification procedure

Though the stratification procedures applied in this study were effective in increasing the accuracy of prediction, some factors could further improve the performance of the methods proposed.

The most prominent problem, especially for the *mean* stratification map, is the fact that pixels with a low quality - that were excluded from the analysis - may result in missing data for particular areas over a number of dates, and uneven distribution of information throughout the year, which leads to a poor representation of such areas. A possible solution would be to take a monthly average NDVI/EVI instead of a total average. This would alleviate the issue of having a skewed temporal distribution, but of course months that feature a lot of low-quality data would still be underrepresented in the dataset and might be less representative of an area.

Furthermore, in the *seasonal* algorithm, irregularities in the time series can strongly influence classification results: one anomaly can already cause a very steep slope to the next pixel (Figure 8), leading to the classification of this time step as the start or the end of the growing season, even though this might not be the case. In this study, pixels with a quality flag ending with *10* or *11* were removed, which is a first step in alleviating this problem. A step further in correcting for this issue is to set a threshold on the maximum slope; however, this is arbitrary and such a threshold is again not very suitable for upscaling, because it will need a different value for each area. Therefore it is difficult to solve, and it might just be something that should be taken into account when analyzing the results.

Vintrou (2012) previously applied MODIS for performing stratified classification using the NDVI and a textural profile as stratification inputs. Their classification aimed to distinguish between cropland and non-cropland at a 250x250m resolution for the whole of Mali. Vintrou (2012) did not mention the removal of pixels (apart from keeping the best of two images per month); if no pixels are removed then there is no such thing as a skewed distribution, but this might lead to other classification artifacts.

However, their methodology is not strongly influenced by the occurrence of multiple growing seasons and skewed distributions as they just used ISODATA clustering on the raw data to define the strata, instead of transforming the raw data using various algorithms before clustering them together. This might be an alternative stratification method to make upscaling even more feasible, but might have repercussions in terms of the resulting classification accuracies as the methods proposed in this study are more tailored towards capturing the variation between crops and their growth patterns.

5.2.2. Training data

Another problem faced during this study is the same problem that also Chomé (2015) encountered, which is the fact that very few ground control points were located on the plateau or the slopes. This makes it difficult to perform crop classification for these areas with a reasonable accuracy. This was especially influencing peanut. The fact that this crop type has a short growing season and is often followed by another crop in the same growing season complicates the classification procedure even further for this crop.

This study did not make use of a technique for removing outliers, while outliers are certainly present (Figure 20). The classification results could be improved by excluding pixels far from the centroid when training the classifier; they may be different crops that do not belong to any of the five classes specified in this study. One way to do this is to calculate the Mahalanobis Distance (MD). This technique was previously adopted in crop classification by Chellasamy et al. (2014) for the exact opposite reason. They used the Mahalanobis Distance to calculate the points *furthest* from the centroid rather than *closest* to the centroid, because Neural Networks require a lot of border samples to reach their potential. This, however, is only beneficial if the training dataset is free of errors, which is not the case in this study due to weeds, and even some inter- or relaycropping in the area, the background influence of the soil and the fact that some tree artifacts might still remain in the images even after tree removal.

A final issue regarding the use of training data is that in this study, the same training set was used for two consecutive years. Such a use of data will never result in a perfect representation of reality, since rotational agriculture is common in this region.

5.2.3. Classification

Regarding classification, an important discussion point is that not all clouds have been removed from the imagery, which strongly affects classification results negatively. As cloud removal techniques are still under development, this process is usually a time-consuming task as it has to be done manually, which is a limitation for upscaling.

Furthermore, one of the limitations of this study is that calibration and validation used the same dataset. This dataset was split randomly and 75% of the points were used for classification while 25% of the dataset was used for validation. This is not the best way to go about validation in a case such as this one where a majority of the training points have been extracted from strongly localized polygons, and are therefore strongly concentrated as soil characteristics and crop management are similar. A better approach would be to validate on a fully independent dataset.

In this study, classification of NDVI composites resulted in lower classification accuracies than classification of single images. This is in contrast with the outcomes of Ommeren (2016), and is ostensibly caused by the low number of images throughout the growing season used in this study (five for 2014, compared to ten used by Ommeren in 2014). Increasing the number of images available within the growing season would unequivocally increase the accuracy of the composite classifications tested in this study. Testing other indices such as PVI, which worked really well for Ommeren (2016), has potential for improving composite classification accuracies as well.

Finally, though the Random Forest algorithm cannot be strictly classified as a machine learning algorithm, it may be less suitable for upscaling than simpler techniques such as k-Nearest Neighbours because of its complexity and its amount of modifiable parameters. Ok et al. (2012) showed that Random Forest is robust along a line of parameters for crop classification purposes, but this has not been tested in this study.

5.3. Upscaling of methodology

Stratification based on MODIS signals has enormous advantages over expert-based interpretations of soil and/or DEM maps. The main advantage is situated in the fact that it is possible to cover large areas as MODIS is freely available on a global scale.

However, still some limitations remain. An important limitation for scaling up the methodology used in this study is the fact that only 1000 kilometers to the south of this region - in regions with access to irrigation - there are multiple growing seasons within one year. The *seasonal* method as it is right now cannot deal with this, and will result in misleading outcomes for these regions. With a view to the scaling up of proposed methodologies, a good solution to this would be the development of different versions of the techniques proposed for different climate zones. When looking at the study conducted by Vintrou (2012), we can see that MODIS time series also have potential for stratification on a larger scale. Their study focused on the whole of Mali, and given the fact that Mali is a country with a large variation in terms of crop growth due to its sheer size and its location in the Sahelian zone, we can conclude that even on a larger scale these methods might be viable. Still, the differences in scale and methodology of their study make that strong conclusions with respect to the upscaling of this study on cannot be drawn from their results. Stratification methods based on soil type - as previously used by Ommeren (2016) - do not have this problem of having to develop different algorithms for different climate regions. Of course, the downside of using this technique is that it is not as suitable for upscaling, because soil maps are not widely - and freely - available at the required spatial resolution. Also, these soil maps are based on expert opinion and have inaccuracies of their own.

For future research, which stratification technique would be optimal cannot be simply inferred from these results. The per pixel *mean* NDVI technique performs rather poorly compared to the other two methods, and because of the fact that it does not really say anything about the *intensity* of plant greenness but rather something about the *total amount* of plant greenness it is least suitable for distinguishing between crops and natural vegetation and even more so between crops. However, whether to use the amplitude or seasonal length signals for clustering and identification of strata depends the local- and regional circumstances. Plants grown in especially poor fields (with low moisture retention capacity and/or rooting depth) are sensitive to dry and to wet conditions, most notably in the early growth stages. Both very dry and very wet (locally saturated- or even flooded) soil conditions may therefore result in poor plant growth or even plant death. Farmers will in these circumstances re-seed or even transplant plants to fill gaps. However, re-seeded plants show a delayed greening up and will not reach the same LAI. Most crops are photoperiod sensitive and will therefore synchronize flowering time and can be harvested at about the same time. So reseeding or transplanting will result in different temporal growth patterns within one field, affecting spatial variability especially early in the growing season. In terms of the choice of stratification algorithm, this could indicate that the amplitude would be affected by this increased spatial variability while the length of the growing season would not. This means that as well as a spatial component, the choice of algorithm has a strong temporal component.

Though there are some limitations, the methodologies proposed have a lot of potential for upscaling because of the global coverage and free availability of MODIS time series. Still a limitation would be the availability of very high-resolution satellite data, which is at the moment mostly only available commercially. Previous crop cover studies such as Miao et al. (2012) and Peña & Brenning (2015) have shown the potential of using satellite data with a coarser spectral- and spatial resolution such as Landsat, but this is only possible when monitoring areas with a rather homogeneous crop cover pattern.

5.4. Recommendations

Recommendations for improving the classification accuracy lie mostly in the selection of the training polygons. More specifically, first and foremost, studying the map and determining which areas are mostly affected by clouds and removing those areas from the training set is essential. In addition, one could consider the adoption of additional (more complex) techniques such as a technique developed by Chellasamy et al., who increased classification accuracy by iteratively removing outliers in the training sample dataset which led to an increase in classification accuracy from 74.9% (using randomly selected training samples) to 82.3%.

Further validation should be done to evaluate whether this method is really robust, in particular in other climate zones. There is still some potential in the use of data collected in this area, as a few other crop cover datasets have been collected here that were not used in this study.

A possibility for further research would be to combine the stratification methods developed by Ommeren (2016) and the stratification methods developed in this study, and to assess the possibility to include different composition methods in this study.

Including of MODIS time series for stratification based on landscape productivity is recommended as the data is freely available and processing is straightforward. It is recommended to ensure that the number of strata are chosen wisely and evaluated by a local expert with knowledge of the area.

6. Conclusions

Several studies have proven the potential of spectroscopy for crop classification.

As of today, procedures to automate the process of crop classification remain unavailable. Though the benefits of introducing stratification layers have now been widely recognized, few stratification methods with potential for upscaling have yet been developed.

In this study, a few of such methods based on a MODIS time series have been developed and tested. In the first method, the strata have been derived from a per pixel mean of the MODIS series. In the second experiment, BFAST was conducted to derive the amplitude. In the third and last experiment, the growing season was determined by applying a new methodology on a BFAST time series.

It became clear that all three methods led to an improvement in classification accuracy, but that *amplitude* and *seasonal* (both based on a BFAST analysis) were most suitable for stratification. Furthermore, the methods proposed are certainly in the same range of classification accuracies as previously used stratification methods, and even perform slightly better compared to methods based on locally available data.

As this study focused solely on MODIS time series, an interesting topic for further research would be to combine these stratification methods with methods that have previously proven to increase the classification accuracy in the study area, such as a stratification model based on soil type. Though this study aimed to develop a methodology that is suitable for upscaling, it is recommended to act with caution when applying the stratification methods developed to different areas, especially in areas with more than one growing season per year.

References

Bainville, S., & Dufumier, M. (2007). Transformations de l'agriculture et reconfiguration des terroirs au Sud-Mali: une "pression démographique" à relativiser. *Belgeo. Revue Belge de Géographie*, (4), 403-414.

Blaes, X., Chomé, G., Lambert, M.J., Traoré, P.S., Schut, A.G., & Defourny, P. (2016). Quantifying Fertilizer Application Response Variability with VHR Satellite NDVI Time Series in a Rainfed Smallholder Cropping System of Mali. *Remote Sensing*, 8(6), 531.

Blaes, X., Traore, P., Schut, T., Ajeigbe, H., Chome, G., Boekelo, B., Diancoumba, M., Goita, K., Inuwa, A., Zurita-Milla, R. & Stratoulias, D. (2015). *STARS-ISABELA 2014-2015 Field data collection protocol.*

Breman, H., & De Wit, C.T. (1983). Rangeland productivity and exploitation in the Sahel. American association for the advancement of science. *Science,* 221, 4618.

Britannica (2015). Tropical wet-dry climate. [Accessed on October 26, 2015]. Retrieved from: *http://www.britannica.com/science/tropical-wet-dry-climate*

Chellasamy, M., Ferre, P.T., & Greve, M.H. (2014). Automatic training sample selection for a multi-evidence based crop classification approach. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(7), 63.

Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment,* 37(1), 35-46.

Dadhwal, V.K., Singh, R.P., Dutta, S., & Parihar, J.S. (2002). Remote sensing based crop inventory: A review of Indian experience. *Tropical Ecology*, 43(1), 107-122.

Davidse, J. (2015). Semi-automatic detection of field boundaries from high-resolution satellite imagery. *Master's thesis,* Wageningen University.

DigitalGlobe (2013). DigitalGlobe data products: WorldView 2. [Accessed on 3rd of June, 2016]. Retrieved from: *http://www.spaceimagingme.com/downloads/sensors/ datasheets/WorldView2-DS-WV2-Web.pdf*

DigitalGlobe (2014). DigitalGlobe data products: WorldView 3. [Accessed on June 3, 2016]. Retrieved from: *http://www.spaceimagingme.com/downloads/sensors/data sheets/DG_WorldView3_DS_2014.pdf*

Dillon, W.P. & Sougy, J.M. (1974). Geology of West Africa and Canary and Cape Verde Islands: The ocean basins and margins, *Springer*, 315-390.

EO (2016). ResourceSat-2. Retrieved from: *https://directory.eoportal.org/web/eo portal/satellite-missions/r/resourcesat-2*

García-Torres, L., Caballero-Novella, J.J., Gómez-Candón, D., & Peña, J.M. (2015). Census parcels cropping system classification from multitemporal remote imagery: A proposed universal methodology. *PloS one*, 10(2), 1371.

Hao, P., Zhan, Y., Wang, L., Niu, Z., & Shakir, M. (2015). Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. *Remote Sensing*, 7(5), 5347-5369.

Huete, A., Justice, C., & Van Leeuwen, W. (1999). MODIS vegetation index (MOD13). *Algorithm theoretical basis document.* Version 3.1. University of Arizona, April 1999.

Jia, K., Wu, B., & Li, Q. (2013). Crop classification using HJ satellite multispectral data in the North China Plain. *Journal of Applied Remote Sensing*, 7(1), 073576-1-073576-12.

Janssen, T. (2015). Deforestation and woody encroachment in the forest savanna transition zone - a remote sensing study on the changes of woody vegetation cover in the Kogyae Strict Nature Reserve, Ghana. *Master's thesis*, Wageningen University.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15, 259-263.

Kumar, P., Gupta, D.K., Mishra, V.N., & Prasad, R. (2015). Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data. *International Journal of Remote Sensing*, 36(6), 1604-1617.

Miao, Z., Qiangzi, L., & Bingfang, W. (2012, August). Investigating the capability of multi-temporal Landsat images for crop identification in high farmland fragmentation regions. *First International Conference on Agro-Geoinformatics* (pp. 1-4). IEEE.

Nelson, R., & Horning, N. (1993). AVHRR-LAC estimates of forest area in Madagascar, 1990. *International Journal of Remote Sensing*, *14*(8), 1463-1475.

Oesterle, M., & Wildmann, R. (2004). Land parcel identification as a part of the integrated administration and control system (IACS). *Wide Angle*, 153, 6-120.

Ok, A.O., Akar, O., & Gungor, O. (2012). Evaluation of random forest method for agricultural crop classification. *European Journal of Remote Sensing*, 45(3), 421-432.

Ommeren, W.L. (2016). Remote Sensing And Crop Recognition: Improving The Information System Supporting Smallholders In Mali. *Master's thesis,* Wageningen University.

Ozdarici-Ok, A., & Akyurek, Z. (2014). Object-based classification of multi-temporal images for agricultural crop mapping in Karacabey Plain, Turkey. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(7), 127.

Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6), 1301-1316.

Peña, M. A., & Brenning, A. (2015). Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. *Remote Sensing of Environment*, 171, 234-244.

Ruelland, D., Levavasseur, F., & Tribotté, A. (2010). Patterns and dynamics of land-cover changes since the 1960s over three experimental areas in Mali. *International Journal of Applied Earth Observation and Geoinformation*, 12, S11-S17.

Schut (2015). Personal communication.

Sonobe, R., Tani, H., Wang, X., Kojima, Y., & Kobayashi, N. (2015). Extreme Learning Machine-based Crop Classification using ALOS/PALSAR Images. *Japan Agricultural Research Quarterly: JARQ*, 49(4), 377-381.

STARS Project (2015). Agricultural improvement in emerging economies. [Accessed on 8 September 2015]. Retrieved from: *http://www.stars-project.org/en/about-us/benefits/*

Strahler, A.H. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, 10(2), 135-163.

Tatsumi, K., Yamashiki, Y., Torres, M.A.C., & Taipe, C.L.R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171-179.

Traore, P.C.S. (2014). Imagery for Smallholders: Activating Business Entry points and Leveraging Agriculture (ISABELA). *International Crops Research Institute For the Semi-Arid Tropics (ICRISAT)*

UN (2012). Demographic and Social Statistics Table 1c - Population growth and distribution. [Accessed on 26 October 2015]. Retrieved from: *http://unstats.un.org/unsd/demographic/products/socind/*

Verbesselt, J., Zeileis, A. & Herold, M. (2012). Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment,* 123, 98-108.

Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., Baron, C., & Seen, D.L. (2012). Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*, 14(1), 83-93.

World Bank Group (2015). Data by country: Mali. [Accessed on 26 October 2015]. Retrieved from: *http://data.worldbank.org/country/mali*

Xie, L., Zhang, H., Li, H., & Wang, C. (2015). A unified framework for crop classification in southern China using fully polarimetric, dual polarimetric, and compact polarimetric SAR data. *International Journal of Remote Sensing*, 36(14), 3798-3818.

Appendix A: Historical background of the area

As in every region on this planet, Ice Ages have had a major impact on this region; in particular, during these periods West Africa was very dry. However, climate changes in Africa have not only been brought about by the Ice Ages. High levels of rainfall around 11,000 BCE put an end to a particularly dry phase in the African climate. This 'wet phase', with more rainfall than today, reached a peak between 9000 and 6000 BCE. Thereafter, these high rainfall levels began to tail off, until rainfall levels had reached something similar to those prevailing today by 3500 BCE.

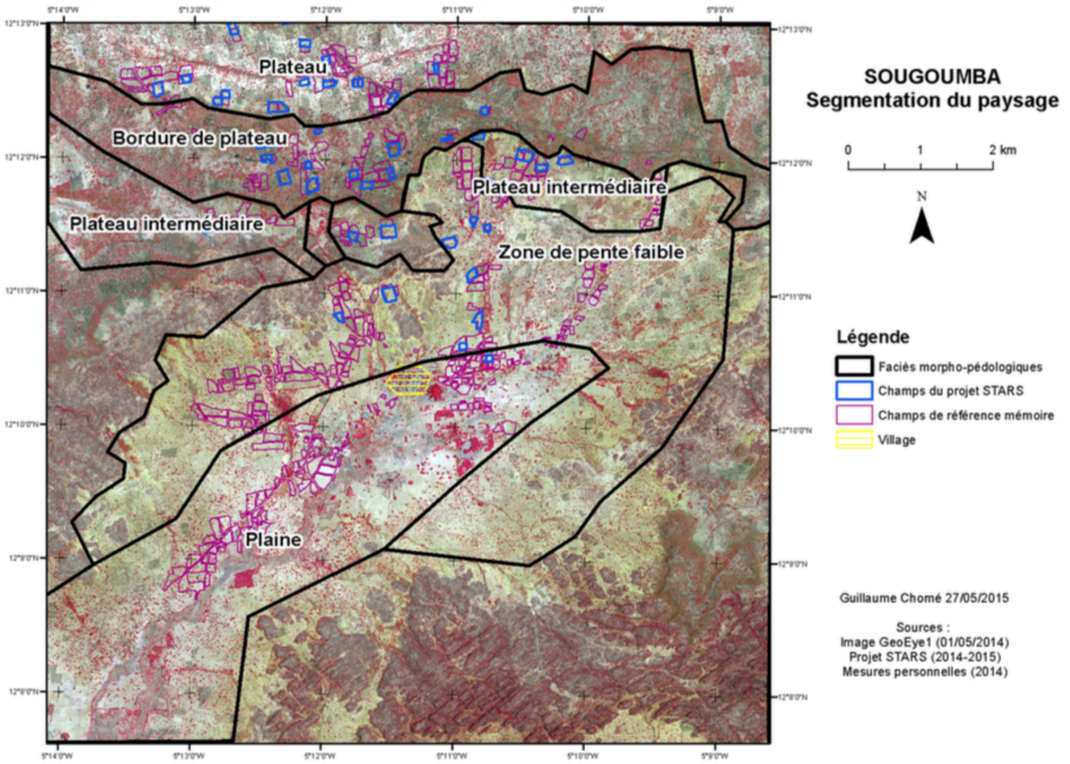
These climate changes had a strong impact on land cover. During the wet phase the deserts contracted, rainforests expanded and the Sahara became an open grassland savannah. Around 7000 BCE, the desert completely disappeared and the tropical forest zone extended far into the region that is known as Mali today (Shillington, 2012).

Around 200,000 yr. BP, humans started influencing land cover by using fire to facilitate hunting and shape their environment (Janssen, 2015). Though this influence was still rather minor, this rapidly changed with the onset of agriculture. The Niger-Congo peoples of West Africa took advantage of the wet phase between 9000 and 5500 BCE to develop planting agriculture in the expanded woodlands of the region. They planted yams, oil palms, peas, groundnuts and kola nuts. By 3000 BCE, they had domesticated guineafowls and were growing raffia palm to make raffia cloth. Between 3000 and 1000 BCE, a West African rice was domesticated in the wetlands of the inland delta of the upper Niger, which was later spread to the high-rainfall forest margins of Guinea, Liberia and Ivory Coast (Shillington, 2012).

These developments had a strong effect on land cover. Natural vegetation was replaced by croplands, people started to live in larger, permanent settlements, logging was done more locally, slash- and burn farming emerged, lands were divided, erosion increased and the population increased tremendously which magnified all the effects mentioned before.

Appendix B: Chomé (2015) stratification map

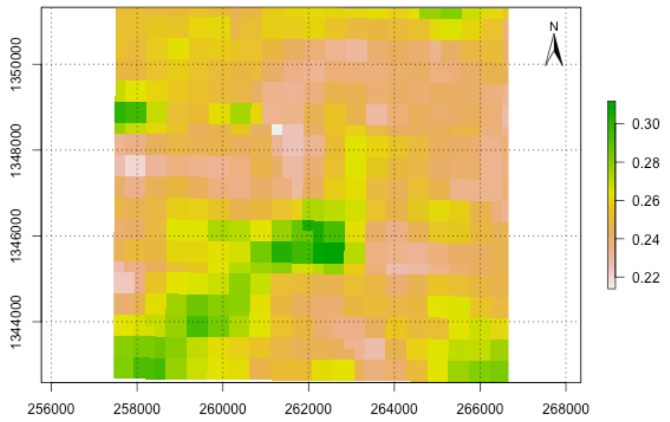
Figure B1. The strata delineated by Chomé (2015) based on expert knowledge of the area. Five distinct zones were identified, delineated in black: plaine (valley), zone de pente faible (areas with a small slope), plateau intermédiaire (intermediate plateau), bordure de plateau (plateau edge) and plateau.



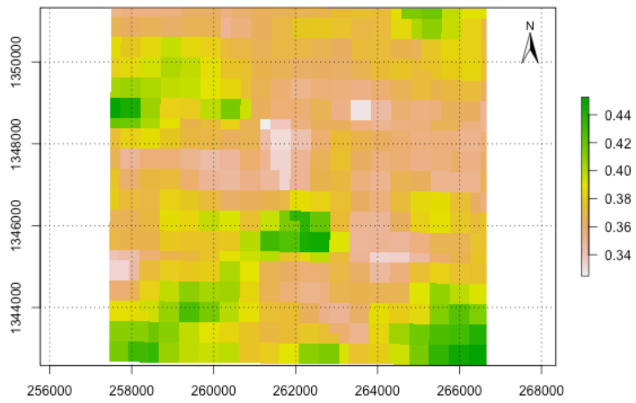
Appendix C. Stratification maps

This section shows the three stratification maps before classifying them into three distinctive categories, both for the EVI as for the NDVI MODIS series.

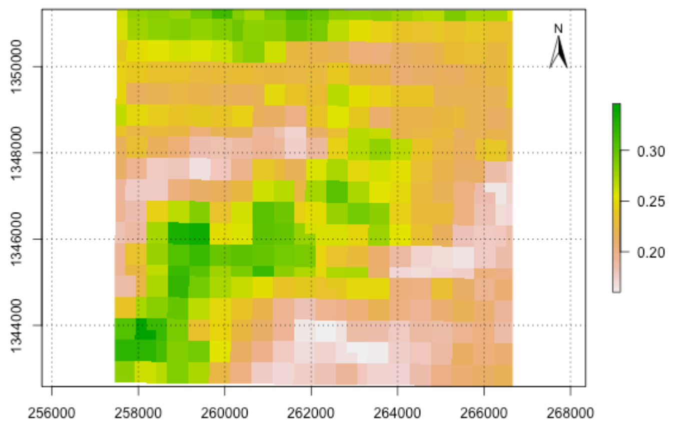
Mean EVI



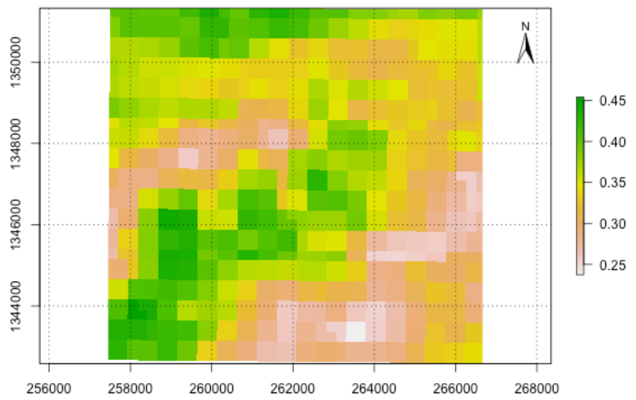
Mean NDVI



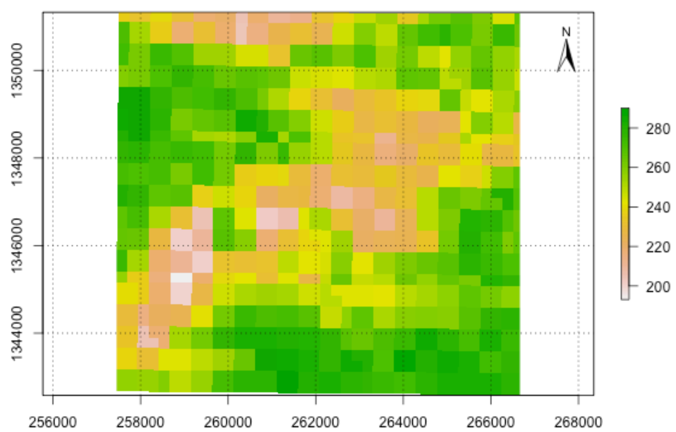
Amplitude EVI



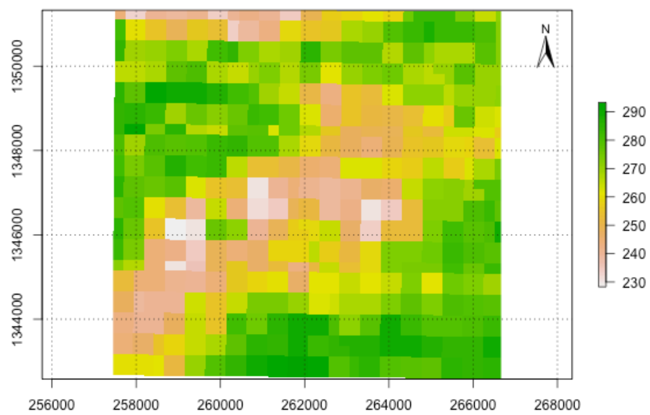
Amplitude NDVI



Seasonal EVI



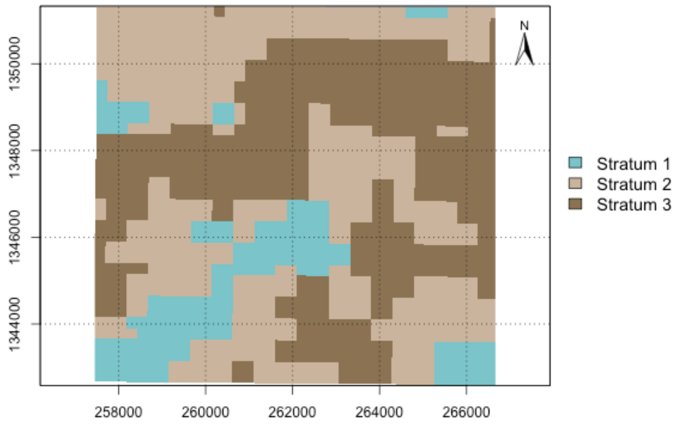
Seasonal NDVI



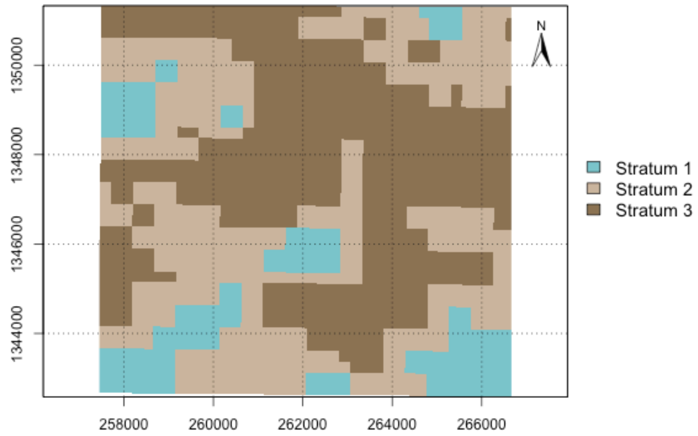
Appendix D: Classified stratification maps

This section shows the three stratification maps after classifying them into three distinctive categories, both for the EVI as for the NDVI MODIS series.

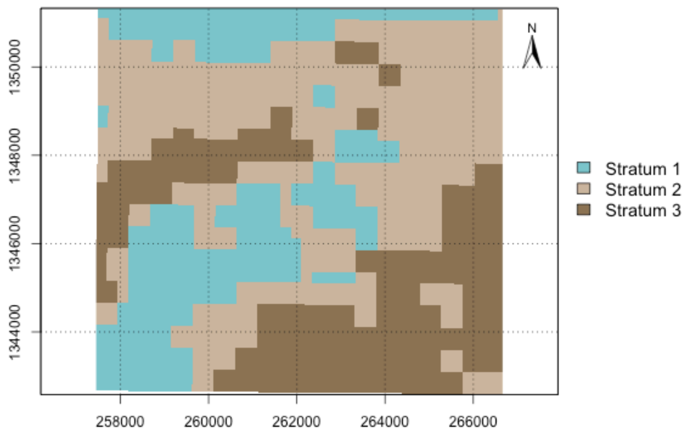
Mean strata EVI



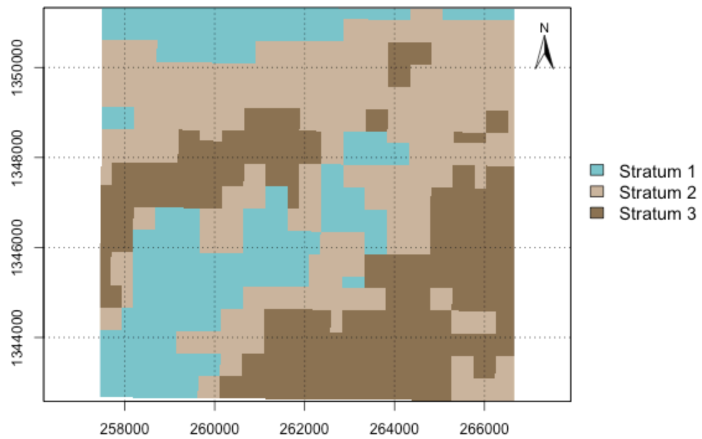
Mean strata NDVI



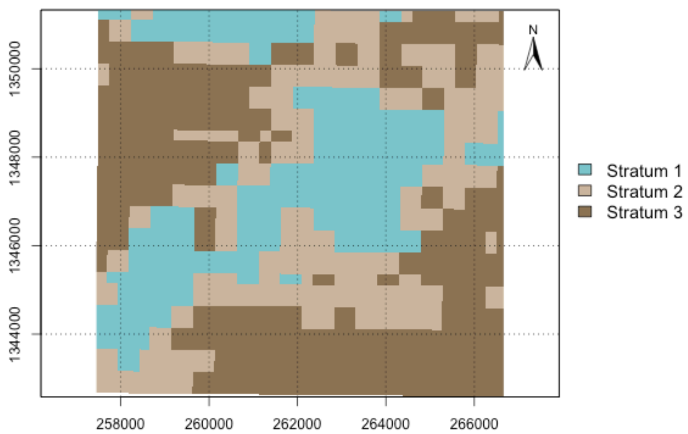
Amplitude strata EVI



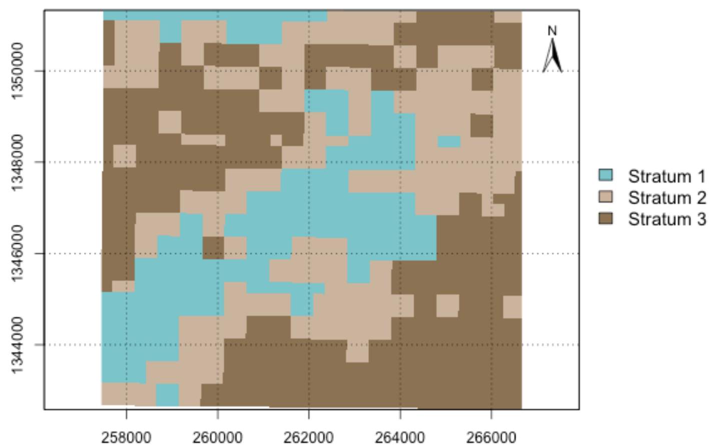
Amplitude strata NDVI



Seasonal strata EVI



Seasonal strata NDVI



Appendix E: Classification accuracies

Table E1. Classification results for the *amplitude* stratification method with the RF classifier on a single image (29th of July, 2014).

	OA	Kappa
Stratum 1 (valley)	0.6823899	0.5161932
Stratum 2 (breakaway)	0.6325301	0.5175389
Stratum 3 (plateau)	0.7075718	0.5951451

Table E2. Classification results for the *mean* stratification method with the RF classifier on a single image (29th of July, 2014).

	OA	Kappa
Stratum 1 (valley)	0.5217391	0.2379518
Stratum 2 (breakaway)	0.6358209	0.4987075
Stratum 3 (plateau)	0.6778656	0.5186071

Table E3. Classification results for the *seasonal* stratification method with the RF classifier on a single image (29th of July, 2014).

	OA	Kappa
Stratum 1 (valley)	0.6124498	0.4795196
Stratum 2 (breakaway)	0.6762402	0.5560251
Stratum 3 (plateau)	0.6603774	0.495468

Table E4. Classification results for the *amplitude* stratification method with the RF classifier on a single image (11th of October, 2015).

	OA	Kappa
Stratum 1 (valley)	0.7269939	0.6135022
Stratum 2 (breakaway)	0.769697	0.6961506
Stratum 3 (plateau)	0.8290155	0.7634057

Table E5. Classification results for the *mean* stratification method with the RF classifier on a single image (11th of October, 2015).

	OA	Kappa
Stratum 1 (valley)	0.3913043	0.05571848
Stratum 2 (breakaway)	0.7412481	0.6491685
Stratum 3 (plateau)	0.8314176	0.7609044

Table E6. Classification results for the *seasonal* stratification method with the RF classifier on a single image (11th of October, 2015).

	OA	Kappa
Stratum 1 (valley)	0.7757202	0.7060299
Stratum 2 (breakaway)	0.8481675	0.7942422
Stratum 3 (plateau)	0.748503	0.6345828

Table E7. Classification results for several classifiers (non-stratified) on a single image (29th of July, 2014).

	OA	Kappa
Random Forest	0.5621351	0.4224295
Multinomial logistic regression	0.5137615	0.3527069
Maximum Likelihood	0.5062552	0.3667048
k-Nearest Neighbour	0.5129470	-
Support Vector Machines	0.5696414	0.4292441

Table E8. Classification results for several classifiers (non-stratified) on a single image (11th of October, 2015).

	OA	Kappa
Random Forest	0.7049044	0.6058668
Multinomial logistic regression	0.4475874	0.2534148
Maximum Likelihood	0.5091514	0.3657287
k-Nearest Neighbour	0.6769611	0.5639671
Support Vector Machines	0.593178	0.454017

Table E9. Classification results for the RF classifier on a composite image (2014, combined by calculating the NDVI of each image) (stratified & non-stratified). The stratified images have been calculated using the *amplitude* method.

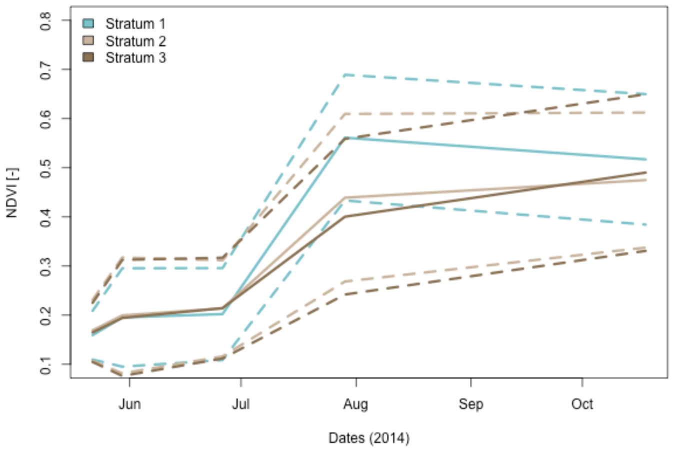
	OA	Kappa
Non-stratified	0.457429	0.2791257
Stratum 1 (valley)	0.7552632	0.5556338
Stratum 2 (breakaway)	0.6259542	0.5125551
Stratum 3 (plateau)	0.6431925	0.5030049

Table E10. Classification results for the RF classifier on a composite image (2015, combined by calculating the NDVI of each image) (stratified & non-stratified). The stratified images have been calculated using the *amplitude* method.

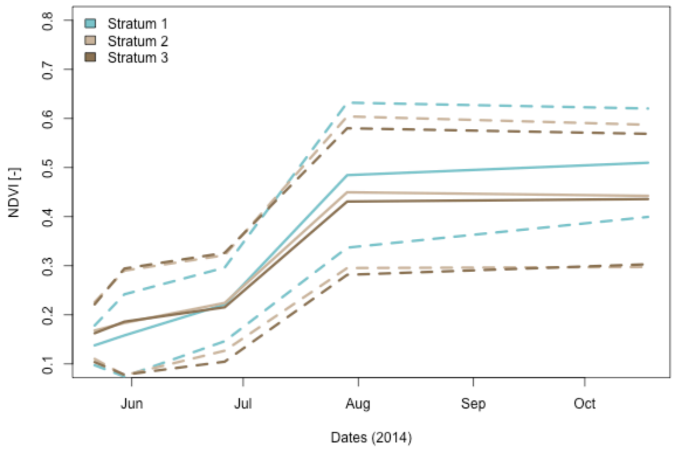
	OA	**Kappa**
Non-stratified	0.5784641	0.4378513
Stratum 1 (valley)	0.6552795	0.4741968
Stratum 2 (breakaway)	0.6506276	0.5359736
Stratum 3 (plateau)	0.718593	0.6085464

Appendix F. The seasonal profiles of different crops

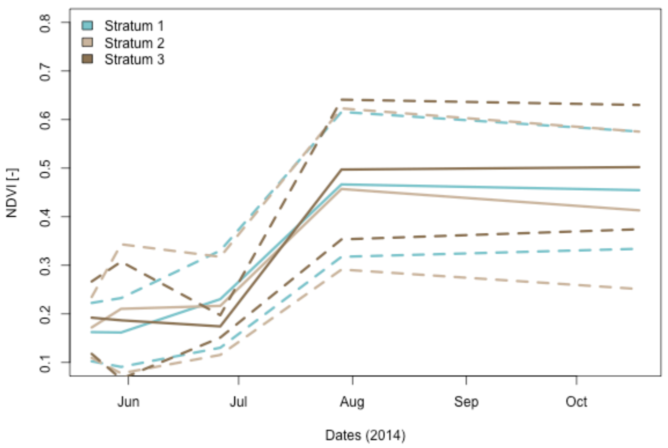
Maize seasonal profile
(stratification *Amplitude*, EVI)



Millet seasonal profile
(stratification *Amplitude*, EVI)



Peanut seasonal profile
(stratification *Amplitude*, EVI)



Sorghum seasonal profile
(stratification *Amplitude*, EVI)