CENTRE FOR GEO-INFORMATION

THESIS REPORT GIRS-2016-03

REMOTE SENSING AND CROP RECOGNITION

IMPROVING THE INFORMATION SYSTEM SUPPORTING SMALLHOLDERS IN MALI

W.L. van Ommeren

March 2016

WAGENINGEN UR
For quality of life

STARS

ICRISAT
Science with a human face

REMOTE SENSING AND CROP RECOGNITION

IMPROVING THE INFORMATION SYSTEM SUPPORTING SMALLHOLDERS IN MALI

Author
W.L. van Ommeren

Registration number 91 04 24 621 080

Supervisors
Dr.ir. J.G.P.W. Clevers
Dr.ir.ing. A.G.T. Schut

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

March 2016
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2016 -03
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

CONTENTS

List of Figures vii
List of Tables viii
Abstract ix
1. Introduction 1
1.1 Background 1
1.2 Problem definition 1
1.3 Research objective and questions 3
2. Methodology and materials 5
2.1 The study area 5
2.2 Data acquisition and pre-processing 6
2.2.1 UAV data collection 6
2.2.2 Satellite imagery 7
2.3 Distinguishing the crops in Samanko 8
2.4 Classification 9
2.4.1 Unsupervised method 9
2.4.2 Supervised method 9
2.4.3 Mono-temporal classification 11
2.4.4 Multi-temporal classification 11
2.4.5 Accuracy assessment 11
2.4.6 Per-field aggregation 12
2.5 Stratification of Sougoumba 12
2.6 Spatial and temporal resolutions 13
3. Results 14
3.1 Suitability of the UAV imagery 14
3.2 Classification of Samanko 14
3.2.1 Mono-temporal distinguishability and classification 14
3.2.2 Multi-temporal distinguishability and classification 15
3.2.3 Per-field aggregation 16
3.3 Classification of Sougoumba without stratification 16
3.3.1 Mono-temporal distinguishability and classification 16
3.3.2 Multi-temporal distinguishability and classification 17
3.4 Sougoumba with stratification 18
3.4.1 Mono-temporal distinguishability and classification 18
3.4.2 Multi-temporal distinguishability and classification 19
3.5 Per-field aggregation 20
3.6 Spatial and temporal resolutions 21
3.6.1 Optimal temporal resolution 21
3.6.2 Optimal spatial resolution 21
3.6.3 Optimal combination 22
4. Discussion 23
4.1 Suitability of the UAV imagery 23
4.2 Classification of Samanko 23
4.2.1 Mono-temporal distinguishability and classification 23
4.2.2 Multi-temporal distinguishability and classification 24
4.3 Classification of Sougoumba without stratification 25
4.3.1 Mono-temporal distinguishability and classification 25
4.3.2 Temporal distinguishability and classification 25
4.4 Sougoumba with stratification 26
4.4.1 Mono-temporal distinguishability and classification 26
4.4.2 Multi-temporal distinguishability and classification 27
4.5 Per-field aggregation 28
4.6 Spatial and temporal resolutions 29
4.6.1 Optimal temporal resolution 29
4.6.2 Optimal spatial resolution 29
4.6.3 Optimal combination 30
5. Conclusion 31
6. Recommendation 32
6.1 UAV data 32

6.2 Satellite data 32
6.3 Training data 32
6.4 Classification 33
6.5 Stratification 33
Acknowledgements 34
References 35
Appendix 1: Crop maps for Samanko 2015 and Sougoumba 2014 38
Appendix 2: Tetracam imagery specifications 39
Appendix 3: Tetracam calibration and alignment of the imagery 40
Appendix 4: Fieldwork protocol 41
Appendix 5: Satellite imagery specifications 43
Appendix 6: Elevation and buildup strata Sougoumba 44
Appendix 7: Mono- and multi-temporal classification results Samanko 45
Appendix 8: Mono and multi-temporal signatures and feature space plots for Sougoumba without stratification 46
Appendix 9: Mono- and multi-temporal classification results of Sougoumba without stratification 49
Appendix 10: Mono-temporal signatures and feature space plots for Sougoumba with stratification of soil 51
Appendix 11: Temporal signatures and feature space plots for Sougoumba with stratification of soil 54
Appendix 12: Mono- and multi-temporal classification results of Sougoumba with stratification 57
Appendix 13: Standard deviation of classes in Samanko 59
Appendix 14: Standard deviation of classes in Sougoumba without stratification 60
Appendix 15: Average standard deviation of classes in Sougoumba with a stratification of soil 62
Appendix 16: Per-field aggregation results 63
Appendix 17: Misclassifications due to within-field variability or misplaced boundaries 64
Appendix 18: Temporal resolution results 65
Appendix 19: Spatial resolution results 66
Appendix 20: Per-field results with optimal temporal resolution (5 dates) 67
Appendix 21: Per-field results with optimal spatial resolution (8 meter) 68
Appendix 22: Average stand. dev. of classes in Sougoumba with a soil stratification and an 8 meter resolution 69
Appendix 23: Removal of the temporal bands 70
Appendix 24: Confusion matrices of the per-field aggregations with the optimal temporal and spatial resolution. 71
Appendix 25: Sowing dates per crop in Sougoumba 72

LIST OF FIGURES

Figure 1. The concept of the improved information system around smallholders. From STARS (2015). 1
Figure 2. Within-field variability due to trees and a variable crop performance (dark red circles are trees, other red pixels represent crops or weed, green areas represent bare soil and the bright green spots are most likely caused by ants eating the crops). WorldView-2 image: 03-09-2015. 2
Figure 3. (A) Millet. (B) Sorghum. (C) Peanuts. (D) Maize. (E) Cotton. 2
Figure 4. Study sites of Samanko (the blue area is where the octocopter data collection took place) and Sougoumba. 5
Figure 5. Repeating field boundaries and misalignment of the fields on Tetracam images with a 2 cm spatial resolution. 6
Figure 6. The effect of Erdas AutoSync. An overlay of the first band of two Tetracam images with a 10 cm resolution (Blue layer: 21-10-2015, Red layer: 18-11-2015). The blue layer has a transparency of 50 percent. (A) Before Erdas AutoSync, the arrow indicates the required shift. (B) After Erdas AutoSync. 7
Figure 7. The effect of Erdas AutoSync. An overlay of the first band of two satellite images (Blue layer: 01-05-2014 (GeoEye-1), Red layer: 14-11-2014 (WorldView-2)). The blue layer has a transparency of 50 percent. (A) Before Erdas AutoSync. (B) After Erdas AutoSync. 8
Figure 8. (A) GeoEye-1 image of Sougoumba (01-05-2014). (B) Tree mask created with ISODATA. (C). A buffer of 5 m around the trees. 8
Figure 9. Crop map for Sougoumba 2015. WorldView-2 image: 19-10-2015. Adapted from de Schaetzen (2015). 10
Figure 10. (A) WorldView-2 image of Sougoumba. The major part of the blueish pixels represent buildup (09-05-2015). (B) Buildup mask created with ISODATA. 12
Figure 11. Soil types and strata in Sougoumba. Adapted from the soil dataset created by Projet Inventaire des Ressources Terrestres (1983). 13
Figure 12. Measured spectral reflectances and the reflectance of their reflectance standards in the Tetracam images with a 10 cm resolution. The Tetracam images were acquired at 15-10-2015, 13-11-2015 and 18-11-15. 14
Figure 13. NIR values of the WorldView-2 (832.5 nm) and aggregated Tetracam image (800 nm) of 13-11-2015 plotted against each other. The initial resolution of the Tetracam image was 10 cm. 14
Figure 14. (A) Feature space plot of the NIR (740nm) and RED (680nm) band with the 95 percent confidence ellipse. (B) Mean spectral profiles plotted with the lower and upper quartile. Based on 10 cm Tetracam image: 15-10-2015. 15
Figure 15. Mean NDVI profiles plotted with the lower and upper quartile. Based on Tetracam images of Samanko. 16
Figure 16. RGB image of temporal NDVI calculated on the 10 cm Tetracam dataset of Samanko. (A) Millet. (B) Sorghum. 16
Figure 17. Pixel-based temporal classification result with the K-NN method applied on the PVI dataset. RGB image: 12-11-2015. 16
Figure 18. (A) Feature space plot of the NIR (816.5nm) and RED (650nm) band with the 95 percent confidence ellipse. (B) Mean spectral profiles. Based on Quickbird image: 04-10-2014. 17
Figure 19. Overall accuracies per classifier for the single image classification of Sougoumba. Based on the satellite images of: 04-10-2014 (Quickbird) and 25-09-2015 (WorldView-2). 17
Figure 20. Temporal mean NDVI and PVI profiles. Based on satellite images of Sougoumba 2014. 17
Figure 21. Overall accuracies per classifier for the multi-temporal PVI classification of Sougoumba. 18
Figure 22. Feature space plot of the early-middle and middle-end season PVI classes of 2014 with the 95 percent confidence ellipse. 18
Figure 23. Feature space plots of the NIR (816.5nm) and RED (650nm) band with the 95 percent confidence ellipse. Based on Quickbird image: 04-10-2014. (A) Without stratification. (B) Plateau alluvium stratum. (C) Valley alluvium stratum. (D) Intermediate soils stratum. 18
Figure 24. Overall accuracies of the classification using a single date per soil stratum. The classifications are based on all training datasets in both years (using K-NN). 19
Figure 25. Feature space plots of the early-middle season PVI classes with the 95 percent confidence ellipse of 2015 (using fields of farmers in the STARS project). (A) Without stratification. (B) Plateau alluvium stratum. (C) Valley alluvium stratum. (D) Intermediate soils stratum. 19
Figure 26. Overall accuracies per soil strata for the classifications based on all training datasets in both years (using K-NN with the temporal PVI series). 20
Figure 27. Overall accuracy with different temporal resolutions for the pixel-based classifications based on all training datasets (using K-NN with the temporal PVI series). 21
Figure 28. Kappa's per stratum for the pixel-based temporal K-NN classifications based on all training datasets. The unsupervised Kappa's represent the classification results on the three temporal PVI images selected without knowing the within-variability or overlap of the classes. The supervised Kappa's represent the results when the three PVI images that showed the largest deviation between the crop classes were used. 21

Figure 29. Overall accuracies per strata for the classifications based on all training datasets in both years (with a factor 4 aggregation of the PVI using the K-NN classifier). 21
Figure 30. (A) Curved crop rows in the original image. (B) Curved crop rows in the optimized image. Tetracam image: 21-10-2015. 23
Figure 31. Temporal within-field (Peanut) variability in Sougoumba, visualised with the PVI. The temporal profile of the three markers is visualised in Figure 30. Based on the satellite images of 2015, from left to right; 03-06-2015, 09-07-2015, 03-09-2015, 25-09-2015, 19-10-2015 and 05-11-2015. 26
Figure 32. Temporal PVI profiles of the markers placed in Figure 29. 26
Figure 33 Classification improvements of Sougoumba, using the K-NN classifier (and PVI dataset for the temporal classification). 27
Figure 34. Within-field variability causing misclassification for Sougoumba 2014. The left images contain a RGB image of the PVI index, each colour represents a different date. The right images contain the resulting classified raster (using K-NN with the PVI). 29
Figure 35. Feature space plots of the early-middle season PVI classes with the 95 percent confidence ellipse of 2015 (using fields of de Schaetzen (2015)). (A) Plateau alluvium stratum, without aggregation. (B) Plateau alluvium stratum, aggregated with factor 4. (C) Intermediate soils stratum, without aggregation. (D) Intermediate soils stratum, aggregated with factor 4. 30

LIST OF TABLES

Table 1. Tested vegetation indices. 11
Table 2. Similarity of the Tetracam images with different resolutions (5 and 10 cm). Based on a linear model with intercept. * All results were significant at $p < 0.01$. 14
Table 3. Average confusion matrix of five spectral maximum likelihood classifications for Samanko 15-10-2015 15
Table 4. Confusion matrix of the per-field aggregation results. Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation). 20
Table 5. Average Kappa per spatial aggregation for the classifications based on all training datasets in both years (using K-NN with the temporal PVI series). 21
Table 6. Confusion matrix of the per-field aggregation results with five dates and a spatial resolution of 8 meters. Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation). 22

ABSTRACT

Creating a transparent information system is crucial to improve the position of smallholder farmers in Mali. Therefore, the project Spurring a Transformation for Agriculture through Remote Sensing was launched and one of the main goals of this project was crop type classification. Crop variability, crop similarity at certain growing stages, different land preparation practices and landscape characteristics were, however, posing challenges for a successful classification. Several classifiers, indices and strata were tested on their suitability for the classification of two study sites. Therefore, both UAV and satellite imagery was made available. Classifications were carried out on a single image and on multi-temporal datasets. In addition, the influence of the spatial and temporal resolution of the images was tested. It was discovered that when crops grew under optimal conditions, a classification with a single image, captured while the crops were still flowering, was sufficient (overall accuracy of more than 80 percent). For a more heterogeneous study site a stratified temporal classification was more suitable. Increasing the spatial resolution was only useful if the training dataset contained a relatively large number of extreme observations. However, increasing the temporal resolution was found to be useful in all classifications. A reduction from eight to five images did not have a large influence on the accuracy. The maximum overall accuracy for the heterogeneous study area was 68 percent when an extensive training dataset covering the entire study area was used. Crop classification on smallholder farms in Mali is thus possible. Moreover, the approach followed by this research is flexible and can be applied to other study areas.

Keywords: pixel-based classification, smallholder farms, Mali, multi-sensor, spatial resolution, temporal resolution, per-field aggregation, vegetation indices, multiple classifiers, mono-temporal classification, multi-temporal classification

1. INTRODUCTION

1.1 Background

About two thirds of the world's consumed food is produced on smallholder farms, which are mainly located in developing countries (Dioula et al., 2013). Therefore, securing the food production in these regions is essential. However, because smallholder farms are mostly located in resource-poor areas, this food security is affected by poverty, poor agricultural productivity and declining soil fertility (Giller et al., 2009; ICRISAT, 2015; Jayne et al., 2003; STARS, 2015).

Knowledge as a result of crop monitoring is one of the main drivers that could enhance the position of the smallholder farmers, as it can lead to sustainable intensification, a decreasing yield gap and higher quality products (STARS, 2015; van Ittersum et al., 2013). Moreover, crop monitoring can secure the investment climate and make markets more accessible. One reason for this is that unharvested crops can be used as collateral by farmers, but that requires that the expected crop productivity is known (Barbier, 1997; Deininger & Byerlee, 2011; Sanchez et al., 1997). At last, it can help in managing the national food supplies more effectively, as a nationwide accurate yield prediction can help to prevent food shortages or oversupply (STARS, 2015). Creating a transparent information system is thus crucial for smallholder farmers.

To improve the information system around smallholders the project Spurring a Transformation for Agriculture through Remote Sensing [STARS] was launched (supported by the Bill & Melinda Gates Foundation) (CGIAR, 2014; STARS, 2015). One of the main value propositions of STARS is crop type classification with remote sensing (Figure 1). The reason for this is that with remote sensing it is possible to monitor a large area in a relatively short period of time. In addition, remote sensing has already been used successfully for crop recognition at mostly larger farms (Bastiaanssen & Ali, 2003; Jakubauskas et al., 2002; Oleson et al., 1995). Conrad et al. (2010) have also classified crops at smallholder farms successfully. The focus in the study of Conrad et al. (2010) was on field boundary detection, per-field aggregation and the use of time series. Still, the optimal classification methods and the influence of spatial and temporal resolution for the classification accuracy are not known. One of the goals of this research is, therefore, to find and explain the optimal method and resolutions. In addition, different vegetation indices were compared with respect to their suitability for crop recognition. To do this, both unmanned aerial vehicle [UAV] and satellite imagery was used and tested on their suitability. Five major crop types were selected, including cotton, maize, sorghum, peanuts and millet. These crops were chosen because they are the most common crops in dryland systems (ICRISAT, 2015; STARS, 2015).

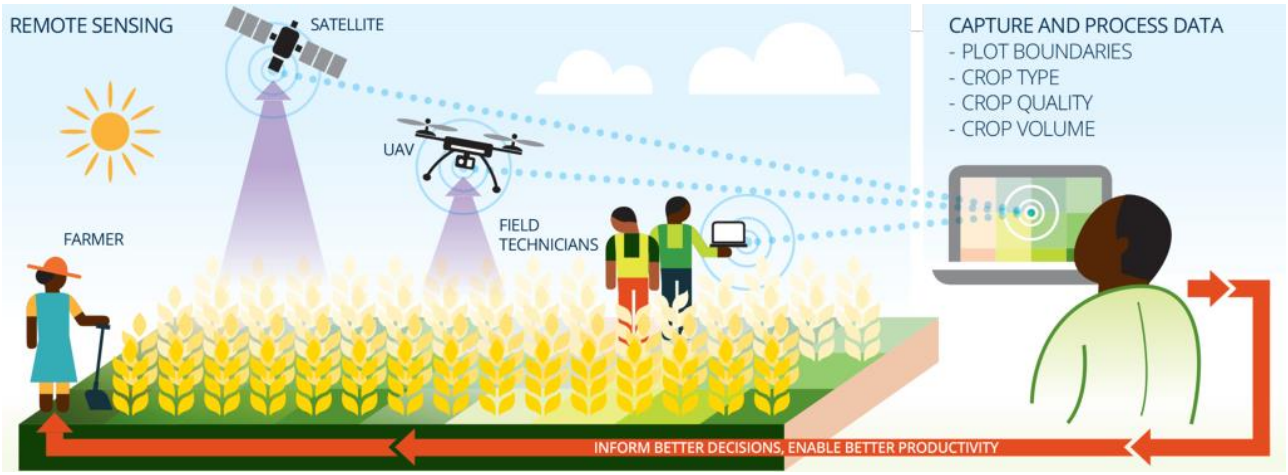


Figure 1. The concept of the improved information system around smallholders. From STARS (2015).

1.2 Problem definition

Recognizing and classifying crops accurately is essential for the improvement of the current information system around smallholder farms. However, when farmlands are heterogeneous, classification becomes challenging. Land preparation practices, farmer resource allocation decisions or a combination of these are causing between and within-field irregularities (e.g. farmers can grow several crops on a single field) (van Ittersum et al., 2013). In addition, landscape and crop characteristics affect the classification: different soil types influence the reflection, trees disturb the signal and the spectral reflection of different crop types can be similar at certain growing stages (CGIAR, 2014; Conrad et al., 2010; Schut et al., 2009; Smith & Fuller, 2001; STARS, 2015) (Figure 2 & Figure 3). In short, spectral variation of crops within and between small fields (average size of 1.5 hectares) is having a major impact on the distinguishability of different crop types.

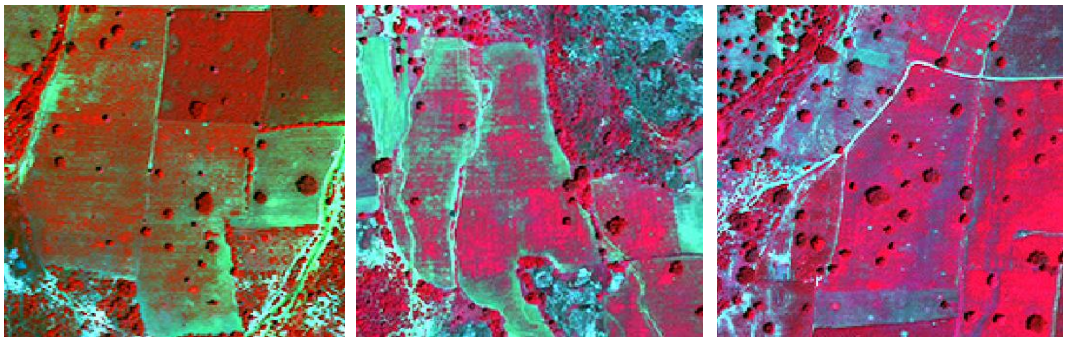


Figure 2. Within-field variability due to trees and a variable crop performance (dark red circles are trees, other red pixels represent crops or weed, green areas represent bare soil and the bright green spots are most likely caused by ants eating the crops). WorldView-2 image: 03-09-2015.

Figure 3. (A) Millet. (B) Sorghum. (C) Peanuts. (D) Maize. (E) Cotton.

Still, previous research has pointed out that an accurate classification is possible (e.g. Conrad et al., 2010). Crop type recognition has been successful with both object and pixel-based methods. Object-based classification proved to yield the most accurate crop type classifications. However, human input has a major role in this approach, as unique segmentation parameters have to be chosen manually (Conrad et al., 2010; Jensen et al., 2006; Lobo, 1997; Lobo et al., 1996). It is also possible to classify crops with a pixel-based approach. The advantage of a pixel-based classification is that it can be done systematically and without human input. However, it will most likely lead to misclassifications due to small scale variability within fields (Chuvieco & Huete, 2009). Both methods have their advantages and disadvantages, but it is also possible to combine and use the advantages of both methods. A pixel-based classification followed by a per-field aggregation is one way to classify crops systematically, while diminishing the misclassifications due to outliers in the fields. This aggregation is done by counting the class that occurs most frequent within a single field (Conrad et al., 2010).

When following the per-field aggregation, the first step in the crop recognition is thus a pixel-based classification. Several methods are designed to assign pixels to a category and each method can yield a different classification result (Chuvieco & Huete, 2009). Accuracy of these different methods varies, and there is no optimal approach, as varying crops and landscape characteristics are influencing the performance of the classification methods (Castillejo-González et al., 2009; South et al., 2004). Therefore, the performance of the classifiers is data driven.

The main challenge in accurately recognizing crops on smallholder farms is overcoming the large within-field variability. A small field size increases the classification complexity, therefore, spatial resolution of the imagery is of importance. With low resolution imagery a single tree on a field with bare soil can increase the measured reflectance in the near-infrared [NIR] and result in a misclassification. In other words, within-field variability cannot be detected with a too coarse resolution. With a higher spatial resolution it might be possible to remove the pixels from the field that are

causing most of the irregularities. Although improving the spatial resolution seems like an optimal solution for an accurate classification, there are two main drawbacks. Firstly, sensors with a high spatial resolution will, generally, have a lower temporal resolution (Al-Wassai & Kalyankar, 2013; Chuvieco & Huete, 2009). Secondly, the spatial resolution can simply be too high. On a very small scale leaves of a single crop type can even be heterogeneous, which makes it impossible to distinguish between different crops based on leaf reflectance.

In addition to spatial resolution, temporal resolution can also strongly affect the accuracy of a crop classification. Multi-temporal datasets have been proven to be superior to mono-temporal datasets when it comes to crop distinguishability (Conrad et al., 2010). During different growth stages plant structure, pigment and plant assemblages vary, which means that the spectral profile changes over time. Crops, which are spectrally similar at some moment in time, can thus differ at other moments, which can make the discrimination between these crops easier (Murthy et al., 2003; Simonneaux et al., 2008). However, there are disadvantages of a higher temporal resolution: the amount of data which needs to be processed increases with a higher temporal resolution, increasing the processing time and number of images needed (costs).

Besides multi-temporal information, other information sources can be used to improve the classification (Franklin, 1989). For example, digital elevation models [DEM] and soil maps have been used to enhance classifications (Gong & Howarth, 1990; Mas, 2004; Smith & Fuller, 2001). On distinct elevation levels or soil types different growing conditions might occur, which alter the spectral signature. In addition, different soil colors might influence the vegetation reflectance (Iverson & Cook, 2000). Another phenomenon, which has been found on smallholder farms, is that there are gradients of decreasing soil fertility with an increasing distance from homesteads (Samaké et al., 2005; Tittonell & Giller, 2013; Tittonell et al., 2005). Crops close to villages have better growing conditions, which might result in different mono or multi-temporal profiles.

Ancillary datasets can be added to the classification by stratifying the study area. In short, stratification can create a larger inter-variability between classes by dividing the study area in a meaningful way (Blaschke et al., 2000; Burnett & Blaschke, 2003). Stratification is achieved by splitting the study area into several parts based on specific rules. Afterwards each of these parts is classified separately (McRoberts et al., 2002). The accuracy will increase if the strata are explaining part of the variability in the area. If, for example, there are three different soil types in the area, three different strata can be defined. Subsequently, in each stratum different training areas can be selected for the same crop. This means that, in this case, a single crop has three training vectors with different spectral characteristics. Because one single crop type is represented in multiple classes the within class variability is most likely decreased, which leads to higher accuracies. The difficulty is to detect and apply relevant strata in the classification (Vintrou et al., 2012).

In addition to ancillary data, vegetation indices might be useful for crop type classification. Vegetation indices can, for example, be used to estimate the biomass in a certain pixel, but also to discriminate between different vegetation types. Several vegetation indices are developed for this purpose, but the suitability of these indices differs per study area (Carlson & Ripley, 1997; Elvidge & Chen, 1995; Huete, 1988). An advantage of vegetation indices is that they can reduce the amount of data in a temporal analysis, as these indices combine several bands in their calculation and their output is a single value per pixel. Thus, instead of using all bands for each time step a single band containing the index can be used. In the classification with temporal vegetation indices the same methods can be applied as with the single image classification. The only difference is that with the multi-temporal dataset each band represents a date instead of measured reflectances at a certain bandwidth.

1.3 Research objective and questions

There is a need to improve the current information system around smallholder farms. Crop type classification with remote sensing may help in this improvement. However, the heterogeneous landscape and fields poses challenges. A pixel-based method in combination with a per-field aggregation seems most suitable, but there are still uncertainties about the most optimal classification method, and the temporal and spatial requirements. Also ancillary data and vegetation indices might increase the accuracy, but their effect on the classification is still unknown. By identifying the requirements and developing a method to accurately distinguish between crop types, this knowledge gap in crop recognition on smallholder farms can be filled.

The principal aim of this study is to develop a method to accurately classify the five main crop types in Mali (cotton, maize, sorghum, peanuts and millet). The STARS project is promoting the use of remote sensing in Mali by testing crop monitoring on their research center (Samanko) and by involving local farmers in the field (Sougoumba). The classification of this research is based on UAV imagery, satellite imagery and ancillary data. The most suitable classification method, vegetation index and temporal and spatial requirements need to be determined. Moreover, the effect of stratification using different ancillary datasets has to be tested. These research objectives are induced in the following research question: How accurately can crop types (cotton, maize, sorghum, peanuts and millet) on smallholder farms in Mali be classified with satellite and UAV imagery? Based on the outcomes of the study of Conrad et al. (2010) it is expected that the classification will be relatively accurate when temporal information and stratification

are used. However, a slightly different method is used on a different study area; therefore, it is most likely that the results will not be completely similar.

To answer the main research question, the following four sub-questions were investigated. Firstly: how are the five crop types spectrally and temporally distinguishable in Samanko and Sougoumba? This question is necessary, as without distinguishability it is not possible to achieve an accurate classification. The assumption is that for Samanko both the single image and multi-temporal profiles will be separable, as crops were well managed and the landscape was homogeneous. For the more heterogeneous area of Sougoumba, however, the multi-temporal profiles are expected to show more deviation than the mono-temporal profiles, as growth patterns are containing information on the phenological development of the crops. It is expected that an increase in information increases the possibility to separate crops from each other. The second sub-question is: how do smallholder field classifications in Sougoumba compare to the classifications in Samanko? Because the distinguishability of different crop types is assumed to be lower in Sougoumba, the classification accuracy is also expected to be lower. The third question is: to what extent is stratification useful for the classification in Sougoumba? Since stratification is expected to decrease the within-class and between-class variability, it is assumed that a stratification will improve the classification accuracy of Sougoumba. The last sub-question is: how are temporal and spatial resolutions influencing the classification accuracy? For the temporal resolution it is expected that images with redundant information and images which were acquired on dates where crop types are highly similar can be removed from the classification without significantly affecting the accuracy. A higher spatial resolution is not per definition expected to always yield a higher accuracy. While a higher spatial resolution might be more suitable for removing trees and bare soil from the images, a high resolution is also expected to increase the variability within the crop classes. By assessing these questions, this study details challenges and accuracies for crop classification in smallholder landscapes.

2. METHODOLOGY AND MATERIALS

2.1 The study area

The data collection of this research took place in Mali at two sites (Figure 4). Both sites were part of the STARS project and millet, sorghum, peanuts, cotton and maize were the five dominant crop types.

The first study area was ICRISAT's research station in Samanko, located 15 kilometers south of Bamako. ICRISAT's laboratory and fields covered a total area of 124 hectares. This area was controlled by researchers and crops grew under controlled conditions. One part of the study area was used to determine optimal crop management. A single crop in this part was expected to show a relatively homogeneous spectral and temporal signature, as there were hardly any within-field irregularities (ICRISAT, 2015). The crops were located in 14 fields, with an average field size of 152 square meters (Appendix 1A). All crops were sowed at closely related dates and each crop was represented in at least two fields were different fertilization doses were applied. In this study area high resolution UAV images were used for the crop classification. These images were acquired by flying an octocopter (the UAV used for this research) over the fields (Paragraph 2.2.1). Unfortunately, only three crops were present in these fields when the octocopter data was collected, namely: sorghum, millet and cotton. Peanuts and maize were already harvested before the experiment.

The second study site covered an area of about 30 square kilometers around Sougoumba, a small village near the border of Burkina Faso. The landscape of this area can be described as "highly and intermediately cultivated tree-shrub savannah area" (Woittiez et al., 2015, p. 53). The main occupation in this area is farming and about 50 farm fields were included in the STARS project. Each of the five crops was represented in about ten fields. The heterogeneous landscape made this area suitable for this study; different soil types, field sizes and planting densities were all occurring. In addition, there were differences in elevation, as part of the area was covered by a flood plain and part by an old plateau (Blaes et al., 2014; STARS, 2015).

There were three reasons for the two different study sites in this research. Firstly, this choice was made to be able to study the crop types in detail. In the area of Samanko crops grew under optimal conditions, small differences between crop types might be discovered and found to be sufficient for an accurate classification. Secondly, a comparison between the two sites might confirm the importance of stratification. In Samanko spectral and temporal signatures will most probably show smooth stable curves and distinguishing between different crop types might be relatively easy. However, for the Sougoumba site this might not be the case, as different growing conditions will most likely cause variability and overlap in the spectral and temporal signatures of the crops. Thirdly, the UAV was stationed in Samanko to minimize the consequences of a failure in the flying system. Samanko is close to the international airport of Mali and new parts could be flown in relatively quickly if needed. This made Samanko the only study site at which it was possible to create an extensive temporal and spatial dataset with the UAV. Therefore, to measure the requirements and possibilities of an accurate classification this study area was necessary.

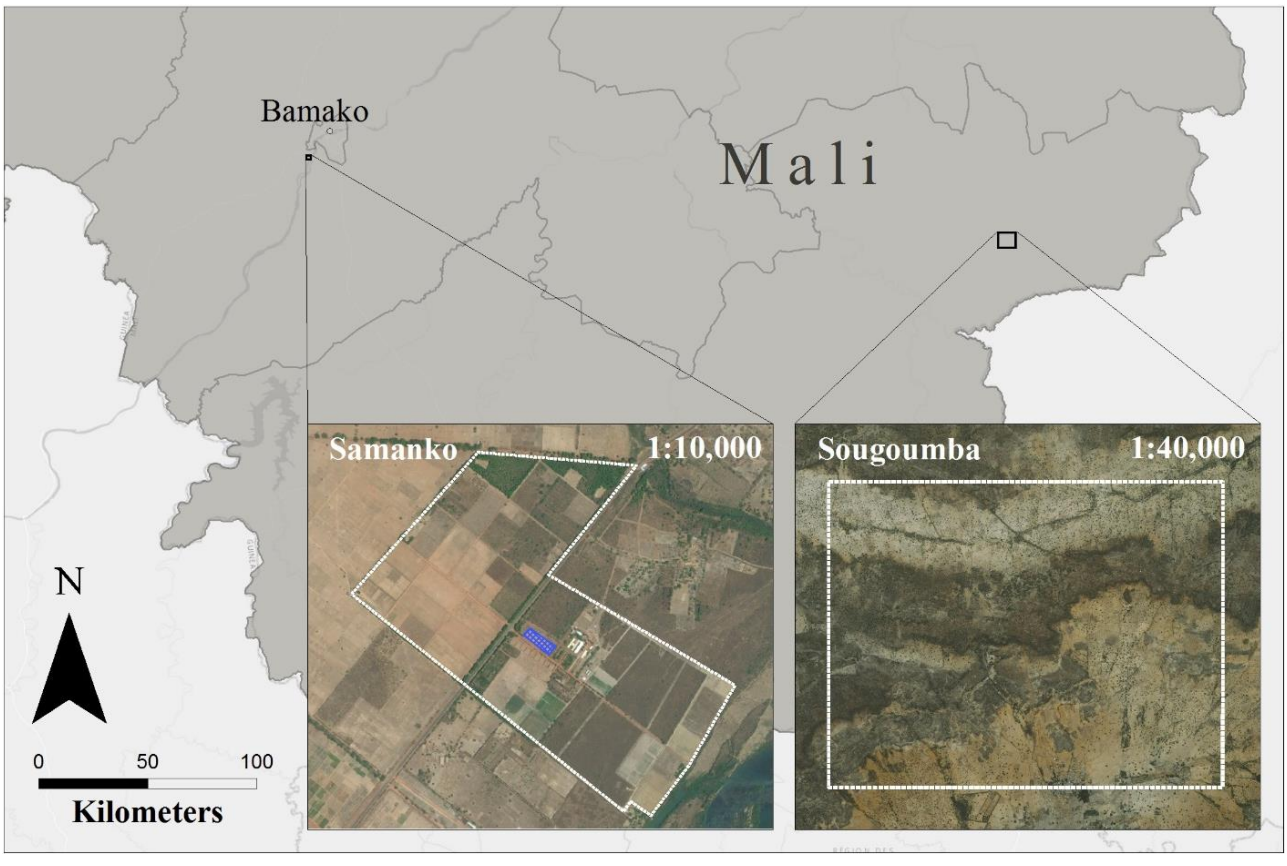


Figure 4. Study sites of Samanko (the blue area is where the octocopter data collection took place) and Sougoumba.

2.2 Data acquisition and pre-processing

2.2.1 UAV data collection

From the beginning of October 2015 until the end of November 2015 an extensive UAV dataset of Samanko was created. Therefore, an octocopter Geo-x8000 mounted with a Tetracam was flown over the area four times a week, at heights of 35, 90 and 185 meter (spatial resolutions of respectively 2, 5 and 10 cm on the ground). For each height a flight plan was created with an overlap of 80 percent. To minimize the influence of shadow, all flights were performed around the same time (1:30 pm – 2:30 pm). The flights at 90 and 185 meter were carried out within one day, because they were short enough to be conducted on a single set of batteries. For this research, the Tetracam with a cmos sensor was measuring in five narrow spectral bands (10 nm bandwidth), which were located within the visible and NIR regions (Appendix 2). In addition, one sensor was pointed upward to measure the incoming radiation.

The Tetracam images were validated in three steps. Firstly, the incidence light sensor [ILS] was calibrated three times during the data acquisition (Appendix 3A). With measurements of the ILS sensor it was possible to directly convert the raw images to images with reflectance values. In addition to the calibration of the ILS, the images were validated by putting reflectance panels alongside the fields and comparing the measured reflectances with the known reflectance standards. These reflectance panels were professionally created and tested by the University of Twente. The check with the panels was carried out on the images at 5 centimeter resolution, as spectral mixing at the borders of the panels will be limited (because the image has a higher resolution). Because the calibration panels were dirty at the start of the data acquisition, they were replaced after the first calibration. Unfortunately, the white panel (80 percent reflectance standard) in the replacement set had two shades. In the report delivered with the panels it was reported that this panel was painted into two colors, sanded and repainted (ITC Faculty, 2014). Nonetheless, the shades were still visible and, therefore, this panel was not used for the calibration. This will not influence the validation of the images as there were three other suitable panels. The second validation step was the comparison of the Tetracam images of different resolutions. And the final step was the comparison of UAV images with semi-synchronously acquired satellite imagery (Appendix 5A). For the final two steps the approach of Matese et al. (2015) was followed: the degree of similarity between the images was calculated by masking the images to the study area. Thereafter, they were resampled to a similar resolution, using bilinear interpolation. Finally, the Pearson's r statistic between the images was calculated.

During the fieldwork a strict protocol was followed which is shown in Appendix 4. Before each flight the ground control points [GCP], Tetracam and octocopter were checked. After the flight the raw data were transferred to a computer in the research centre. PixelWrench 2 was used to convert the images to tiffs, which contained reflectance values scaled to the image bit space (8-bit, scale from 0 – 255). For each flight an orthophoto was created using the software Pix4d. To create reliable orthophotos the settings for the highest accuracy were chosen (Appendix 3B). Afterwards, the image was georeferenced using four GCPs, which were measured with a real-time kinematic device. The datum was set to WGS 84 / UTM zone 29N. In addition to the GCPs, markers were placed in the image on static objects to optimize the alignment (10/20 markers for respectively 10/5 centimeter spatial resolution). Unfortunately, Pix4d was not able to align the images with a 2 centimeter spatial resolution (35 meter flight height). Slight movements of vegetation and a major change in relative position of static objects to the camera caused blurry and noisy orthophotos (Figure 5).

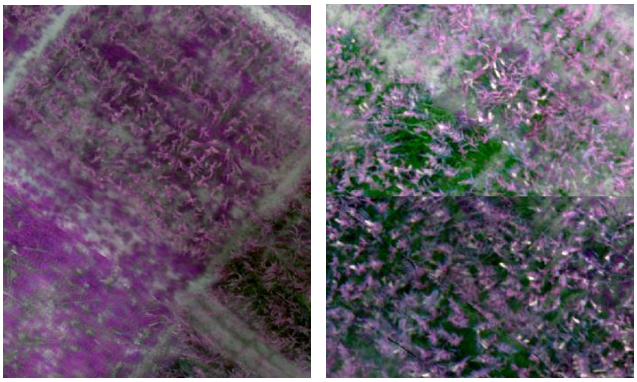


Figure 5. Repeating field boundaries and misalignment of the fields on Tetracam images with a 2 cm spatial resolution.

Due to the manual placement of the GCPs on the orthophotos a placement error was created; as a result, the GCPs and fields in all orthophotos were not located on the exact same coordinates. To decrease this georeferencing error the automatic sync station in Erdas IMAGINE 2015 was used (Figure 6). With this tool the orthophotos were optimized by aligning them automatically to one reference orthophoto. This alignment was based on tie points created with an automatic point matching algorithm (Watson, 2006). The photo with a 5 centimeter resolution and the lowest root mean square error between the modelled and measured GCPs was chosen as reference image. All other orthophotos of 5 and 10 centimeter resolution were re-aligned to this image (keeping the original resolution). Again the settings

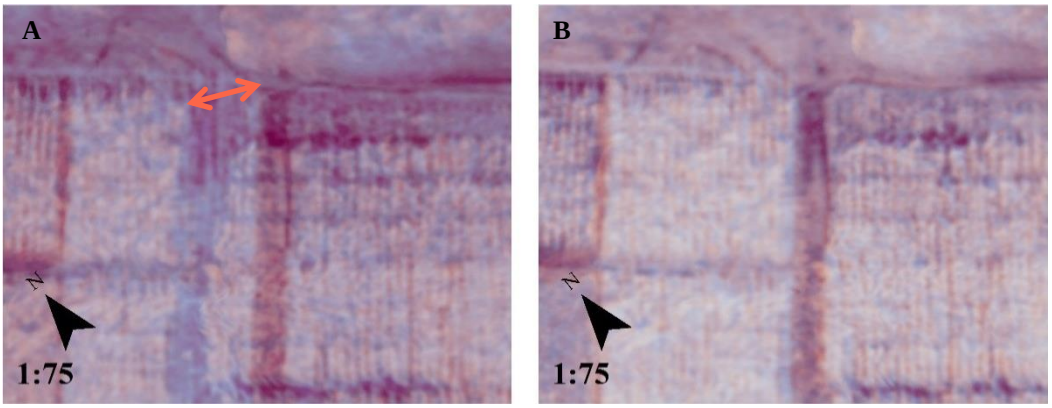


Figure 6. The effect of Erdas AutoSync. An overlay of the first band of two Tetracam images with a 10 cm resolution (Blue layer: 21-10-2015, Red layer: 18-11-2015). The blue layer has a transparency of 50 percent. (A) Before Erdas AutoSync, the arrow indicates the required shift. (B) After Erdas AutoSync.

were set to obtain the maximum accuracy (Appendix 3C). A polynomial order of 3 was chosen, because, on top of the shift between the images, there were some distortions visible within the area of the four GCPs.

After the optimization in Erdas, the images were prepared for the analysis in four steps. Firstly, all images were masked to the study area, removing all unnecessary edges and reducing the amount of data. Secondly, a standard grid was created and all orthophotos were resampled to this grid (using nearest neighbor to avoid spectral mixing). The grid cell size was in accordance with the obtained orthophoto cell size (5 and 10 cm). Without this step it would not be possible to stack vegetation indices derived from the imagery of several dates. Thirdly, the reflectance values were rescaled and trimmed to values between 0 and 1 to be able to compare the UAV images with satellite images. The final step was the removal of bare soil from the images. This bare soil would have affected the spectral signatures of the crops, as a random point selection algorithm was used for the creation of the training data (Paragraph 2.4.2). To remove the bare soil the normalized difference vegetation index [NDVI] had to be calculated first (Paragraph 2.4.3). All pixels with a NDVI lower than 0.2, were removed from the images.

2.2.2 Satellite imagery

In addition to the octocopter imagery, satellite data were used in this research (Appendix 5). The satellite images were delivered after orthorectification and atmospheric correction by the University of Twente. A dataset of WorldView-2 images, acquired semi-synchronously with the octocopter data, was made available for the site of Samanko. As mentioned, the main goal of these images was to validate the UAV imagery. In addition to the Samanko satellite imagery, very high resolution images of the growing seasons in 2014 and 2015 were acquired for Sougoumba. This imagery was used for the classification of this study site. All images had to be pre-processed before they were suitable for this research.

The first pre-processing step was removing the images with clouds covering the training areas. However, images with cloud shadows which were covering the training areas were kept, as ratio-based vegetation indices will partly level out the effect of this shadow (Chuvieco & Huete, 2009). Moreover, most clouds occurred in the growing season, which is the period at which crops can distinguish themselves in their green up speed; therefore, the removal of all images containing cloud shadows would most likely have resulted in a loss of relevant information and have affected the classification negatively.

Secondly, not required bands were removed from the imagery. All satellite images contained four sensor bands similar to the Tetracam bands. Especially, the WorldView imagery of type MULTI required band removal, since these images contained eight bands. To reduce the amount of data and ease up the calculations the non-overlapping bands were removed from the satellite and Tetracam images. This resulted in similar sized datasets for all satellite and Tetracam images.

The third step was re-alignment of the images, as after overlaying the images it became clear that the orthorectification was not without errors (Figure 7). Therefore, IMAGINE AutoSync was used to optimize the alignment, with the same settings as before except the polynomial order was set to 2. The reason for this was that the images only needed to be corrected for a shift. All images were re-aligned to an image without cloud cover acquired just before the start of the growing season. This increased the chance that the automated tie-point mechanism chose objects which were static during the entire growing season to re-align the image, as hardly any vegetation was present in the fields.

After the re-alignment of the imagery the same pre-processing chain as the UAV chain was applied to all images. The grid cell size was set to 2 meter, as this was the most common cell size of the initial imagery. In this way the resampling error stayed limited, because in theory cells only had to be shifted (with the nearest neighbor assignment).

Bare soil pixels were also removed from the images. However, removing the bare soil from all imagery would be incorrect, as the phenological development of a crop is also visible in the speed at which it is increasing ground cover. Therefore, only the bare soil pixels present in the peak of the season were removed.

The pre-processing chain of the satellite images of Sougoumba required an extra step, because trees were causing variation in the field (Figure 8). Especially when determining training data with a random point mechanism, the removal of pixels that are not representing the specific crop type is important to create valid signatures. In the early months of the growing season trees were clearly visible in the landscape, as they were the only green vegetation visible in the study area. Therefore, trees were detectable with the Iterative Self-Organizing Data Analysis Technique [ISODATA] algorithm in ERDAS (Paragraph 2.4.1). From the total of five classes, the classes that did not represent trees were set to 0 and the tree classes to 1. Afterwards the raster was converted to polygons which only represented trees. To ensure that the complete trees would be filtered out of all images, a buffer of 5 meters was created around each tree polygon. This was done because motion parallax was present in the images. This means that the apparent position of stationary objects (trees) is different when the viewing angle of the satellite image is different (Lillesand et al., 2014).

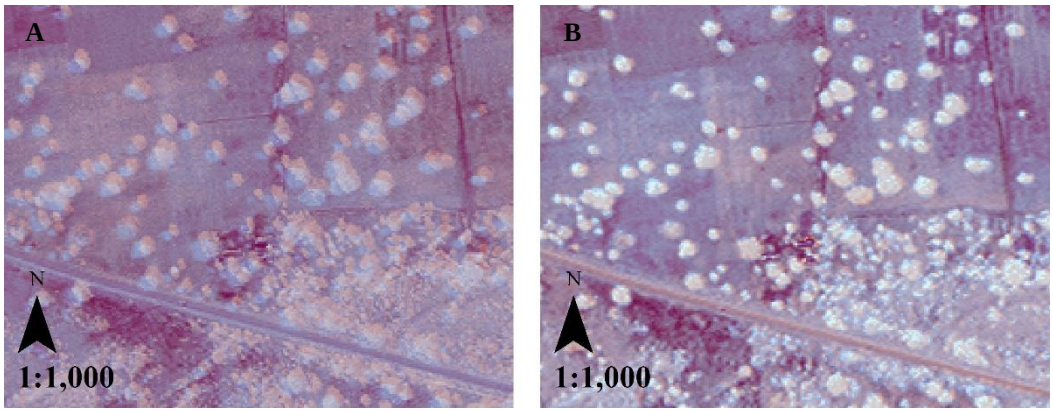


Figure 7. The effect of Erdas AutoSync. An overlay of the first band of two satellite images (Blue layer: 01-05-2014 (GeoEye-1), Red layer: 14-11-2014 (WorldView-2)). The blue layer has a transparency of 50 percent. (A) Before Erdas AutoSync. (B) After Erdas AutoSync.

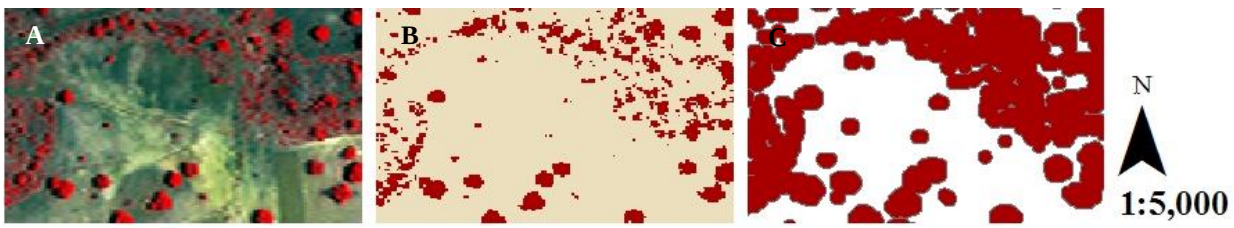


Figure 8. (A) GeoEye-1 image of Sougoumba (01-05-2014). (B) Tree mask created with ISODATA. (C). A buffer of 5 m around the trees.

2.3 Distinguishing the crops in Samanko

Before the classification it was important to test if the crops in Samanko were actually separable. To check the separability, feature space plots and spectral signatures were created for each crop. The plots and profiles were based on the images with a 10 centimeter resolution, because this image should contain the same (averaged) values as the image with a 5 centimeter resolution (validation of the imagery will underpin this). Because just three of the five crops were present at the time of data collection, only the distinguishability of these crops could be tested. In this research crop class separability was tested with a single spectral image and with a multi-temporal dataset.

For the creation of spectral profiles of a single image the UAV image of 15 October 2015 was used, as the crops were still flourishing at that time. Therefore, they were expected to be separable. When spectral signatures were extracted on pixel level it was also possible that green pixels, which represented weed, would have been selected. Therefore, for each of these crops the mean spectral profiles were measured at field level (after the removal of bare soil). These weed pixels were now for sure included in the single image profiles, but were averaged out by reflectance values of the crop. In addition to the mean profile, the upper and lower quantiles were measured to check the overlap between the classes. Finally, the standard deviation within the classes was calculated to measure the within-class variability.

In addition to the crop class separability at a single date, the multi-temporal distinguishability was checked. To create temporal profiles and a feature space plot, the data were transformed to a single band per time step. For this the NDVI was calculated for each of the images. For Samanko it was expected that the soil does not have a major influence

on crop appearance, as the soil is homogeneous and fertilizer was applied when needed. In addition, the NDVI is a ratio-based index, which means that shadow effect will be partly leveled out and, as this research was dealing with very high resolution imagery, leaf shadows were influencing the measured reflectances (Chuvieco & Huete, 2009). The NDVI profiles were also aggregated to field level and, similar to the mono-temporal measurements of separability, the quantiles and standard deviations were calculated.

2.4 Classification

In this research several classification methods were used. An unsupervised method was used to define strata and to detect trees and buildup in the satellite images. Supervised methods were used for the actual classification. Because there is no optimal classifier, five classifiers were tested on their suitability for the classifications.

2.4.1 Unsupervised method

Several times in this study an unsupervised pixel-based classification was performed. With this approach no prior knowledge is needed, as ISODATA (the method which was used) defines its own classes based only on the input image. ISODATA is a common unsupervised classification technique where initially arbitrary cluster centers are spread randomly in feature space. Each pixel can only fall into a single cluster. In each iteration, pixels are classified based on minimum distance to cluster centers in feature space, and new means for clusters are determined. This complete process is repeated until the percentage of pixels changed in the process reaches a certain threshold (convergence threshold), which is specified by the user. After the classification the clusters still need to be named to the class they represent (Chuvieco & Huete, 2009; Sohn & Rebello, 2002; Zhong et al., 2006). The ISODATA classification was carried out in Erdas IMAGINE 2015. Clusters were formed with the standard settings. Only the maximum number of iterations was changed to 25 (to ensure that the convergence threshold of 0.95 was reached).

2.4.2 Supervised method

In contrast to the unsupervised method, supervised classifications depends on prior knowledge of the study area. Two steps are needed for a supervised classification. The first and most important step is defining the classes through training. Ground truth maps were used, which were representing field boundaries of the crops to create the training data. It is not possible to select an average number or trend that represents a specific class, as a single class always occupies a range of values. For example, similar crop types can have slightly different spectral signatures because they are in a different phenological state. Therefore, it is important to catch this possible variability within the training data (Chuvieco & Huete, 2009).

For the imagery of Samanko a field boundary dataset was created and validated based on field work (Appendix 1A). For Sougoumba, the second study site, polygons had already been created. In 2014 and 2015 a field campaign was covered by the ICRISAT research program and they collected data during the growing season (Blaes et al., 2014) (Appendix 1B). All this data were available in polygon format. The polygons were created with a handheld GPS and manually adjusted with UAV imagery. Nonetheless, for 2015 the most optimal dataset was created by de Schaetzen (2015) (Figure 9). This third dataset also included farms that were not involved in the STARS project and contained a total of 504 fields relevant for the classification. However, to be able to compare the classification results of both years also the STARS training dataset was used in the classification of 2015. The reason for this is that both STARS datasets had a similar number of fields and 90 percent of farmers in 2015 were also participating in 2014. The latter is an advantage, as most likely the farming practices of these farmers will be similar in both years.

The training phase requires human input and thus needed to be systematic. To diminish the effect of human error, overlapping borders and GPS inaccuracies the size of all polygons was decreased. In Samanko the diameter of all polygons was reduced with 1 meter and in Sougoumba with 10 meters. This was done with an inside buffer which was clipped from the polygons. The choice for the different distances was based on the fact that there were only a few small training areas available in Samanko and a reduction of 10 meters would remove a major part of these areas. For Sougoumba, enough larger training areas were available. In addition, the borders of the training datasets of Sougoumba were not checked individually, which made the decrease of 10 meters a safer option. The next step in systematically defining the training data was the creation of random points per crop in the fields. From these points the training data could be extracted. Regarding the sample size, a number of 100**no_of_classes* pixels would be sufficient according to several scholars (Chuvieco & Huete, 2009; Mather & Tso, 2009; Shahshahani & Landgrebe, 1994). Therefore, 100 pixels per crop type were randomly selected in all three field boundary datasets. This data were randomly split, without replacement, into a training and test dataset. The training dataset contained 80 percent of the selected points per crop type.

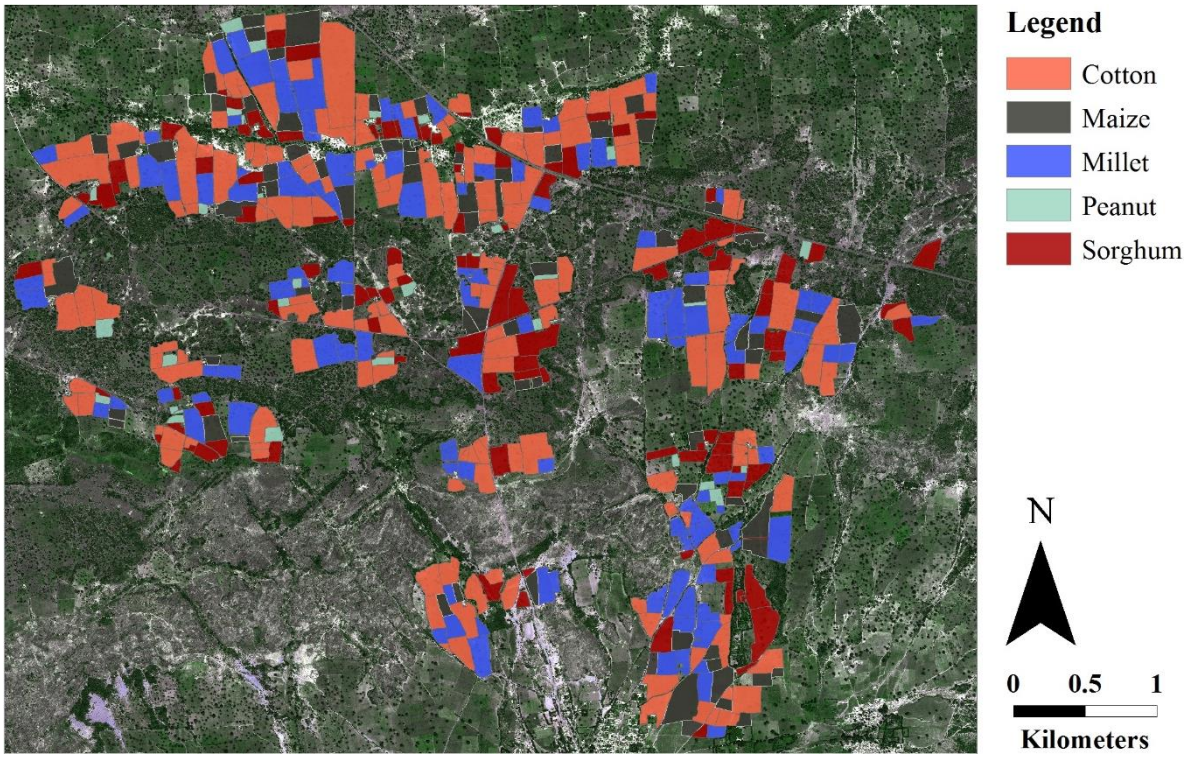


Figure 9. Crop map for Sougoumba 2015. WorldView-2 image: 19-10-2015. Adapted from de Schaetzen (2015).

The final step of a supervised classification is the assignment phase (Chuvieco & Huete, 2009; Sohn & Rebello, 2002). The purpose of this phase is to actually transform the digital numbers to categories based on the training data. In this study five classification methods were tested on their performance, which were:

- The parametric minimum distance, which assigns a pixel to the class from which the center in feature space is nearest. To find the nearest class center the Euclidean distance is used. Therefore, this classifier is relatively sensitive to outliers and overlap between classes (Chuvieco & Huete, 2009);
- The parametric maximum likelihood, which assigns pixels to classes based on their mean vector and variance-covariance matrix. The probability that these pixel characteristics belong to a certain class is the determining factor in the classification (Chuvieco & Huete, 2009);
- The non-parametric spectral angle classifier, which is based on the angle between the spectra of the classes. This makes this method sensitive to outliers and overlapping classes. In addition, spectral mixing of pixels will negatively influence the result of this classifier, as it assumes that a single pixel represents a single ground cover type (Rashmi et al., 2014);
- The non-parametric k-nearest neighbor classifier, which uses the class which is occurring the most in the *k* closest samples to classify a pixel. The assignment of a class is only based on these closest samples and excludes the other samples. This means the algorithm is less sensitive to outliers and overlapping classes (Song et al., 2007). In this research, the *k*-value was set to the square root of the training samples, rounded to the nearest odd integer (Jönsson & Wohlin, 2004). In this research 100 sample points were created per crop class (total of 500). From these samples 80 percent was used for training (total of 400). The *k*-value was thus the square root of 400 rounded to the nearest odd integer, which is 19;
- The non-parametric regression tree classifier, which predicts the dependent variable (crop) with independent variables using decision rules. The classifier is using all relevant training vectors to define a single class. However, for each class only the most suitable bands are used in the classification. Bands in which classes are overlapping can therefore be excluded (Chuvieco & Huete, 2009).

The first three methods were carried out in Erdas IMAGINE and the latter two in R. All methods were tested on their accuracy for the Samanko datasets and the most accurate method(s) were used for the classification of Sougoumba. As a major part of the classification was dependent on the random training data, all classifications were carried out five times. From the five classified rasters the most common pixel value (mode) was selected and used as value in de final classified raster. In this way, the risk of a significantly lower or higher accuracy due to a misplacement of the random training and testing points was reduced.

2.4.3 Mono-temporal classification

The suitability of the classifiers was tested on mono and multi-temporal datasets. A classification solely based on the spectral information of a single image was carried out first. For the mono-temporal classification of Samanko, all classification methods were carried out on the UAV dataset of 15 October 2015, the date closest to the peak of the growing season. Only the dataset with a 10 centimeter resolution was classified in this step. The reason for this was that this image should contain the same information as the higher resolution datasets, but requires roughly half the computing power. The most accurate classifiers which were applied on a single image classification of Samanko, were also used in the mono-temporal classification of Sougoumba. For the mono-temporal classification of Sougoumba, the images closest to 15 October were used to be able to compare the results to the Samanko classification.

2.4.4 Multi-temporal classification

In addition, a classification based on temporal information was carried out. Differences in phenological development might be enough to be able to separate spectrally similar crops. To create the temporal profiles vegetation indices were used. The main goal of vegetation indices is to enhance the vegetation signal (Chuvieco & Huete, 2009). An advantage of these indices is that they combine multiple bands to calculate a single index image which contains the most relevant information.

Probably the most well-known vegetation index is the NDVI, which uses the NIR and red bands to characterize vegetation. Other vegetation indices are also using these spectral bands, as they are indicators for absorption by chlorophyll (red) and light scattering by air/water particles determined by leaf structure (NIR). In addition to the NDVI, there are indices which add extra information by including soil parameters in the calculation. Tested soil-adjusted indices and their formulas are listed in Table 1. Both orthogonal and ratio-based soil-adjusting vegetation indices were compared. Orthogonal indices measure the perpendicular distance of a pixel to the soil line, while ratio-based indices are measuring the slope of the line between the soil-adjusted origin of the NIR-red feature space plot and the NIR-red pixel values (Chuvieco & Huete, 2009).

To be able to apply the soil-adjusted indices the characteristics of the soil line need to be known (Baret et al., 1993). These characteristics were extracted from the images. The most suitable method determined by Xu and Guo (2013) was used, which is known as the (R, NIR_{min}) method. This method uses the NIR and red band to extract the soil line from a two-dimensional scatter plot. This is done by finding the minimal NIR value for each interval of 0.005 red reflectance. Thereafter, the soil line is derived from the relationship between the red and NIR reflectances for the identified minimum values per interval by fitting a linear model. With this calculation the necessary soil characteristics were known and the vegetation indices were calculated.

Time series of vegetation indices were used for classification. To find the most suitable classifier for the multi-temporal classification, all classification methods were applied using the NDVI. Thereafter, each vegetation index was tested independently on the Samanko dataset with the most accurate classification method(s). The classification method(s) and vegetation index(/indices) that yielded the highest accuracy were used in the rest of the research.

Table 1. Tested vegetation indices.

Abbreviation	Formula
NDVI	$\frac{NIR - RED}{NIR + RED}$
SAVI	$\frac{NIR - RED}{NIR + RED + L} \cdot (1 + L)$
TSAVI	$\frac{a \cdot (NIR - a \cdot RED - b)}{a \cdot NIR + RED - a \cdot b + X \cdot (1 + b^2)}$
PVI	$\frac{1}{\sqrt{1 + a^2}} \cdot (NIR - a \cdot RED - b)$
WDVI	$\mathrm{NIR} - a \cdot \mathrm{RED}$

Note. Adapted from Huete (1988); Tucker (1979); Baret et al. (1989); Richardson and Weigand (1977); Clevers (1988)**.**

2.4.5 Accuracy assessment

The accuracy of the classifications is not only dependent on the training phase but also on the classification method. To test the accuracy of each method it is required to have an independent test dataset (Mather & Tso, 2009). In this research the remaining 20 percent of the randomly selected points in the training fields, was used as validation data (See Paragraph 2.4.2). The accuracy was calculated by comparing the classified image with the classes in the ground truth

datasets at the test pixel locations. The overall, producer and user accuracy were calculated per classified raster using a confusion matrix. In addition, the Kappa index was calculated, which also takes the random chance of a correct classification into account (Chuvieco & Huete, 2009; Congalton, 1991; Mather & Tso, 2009).

2.4.6 Per-field aggregation

As mentioned, a pixel-based classification can be improved with a per-field aggregation. For example, it was highly likely that in some of the heterogeneous fields noise was caused by an aberrant crop in a field. Thus the pixel of this aberrant crop might be classified correctly; however, it still is decreasing the accuracy of the classification, as the training and test data were selected based on the assumptions that the pixels in the fields were all representing a single crop. To diminish the influence of outliers in the field the most frequent class occurring in a field was assigned to that field. This per-field majority filter was, thus, firstly applied to remove noise from the classified image. Secondly, the per-field aggregation was applied to locate the fields which were misclassified (at least the majority of the pixels in the field was misclassified). In this way, the reason for this misclassification can be determined by looking at the field characteristics. To measure the accuracy of the per-field aggregation again the confusion matrix and Kappa were calculated. This aggregation was only applied on the results of the most accurate classification methods for both study areas.

2.5 Stratification of Sougoumba

Overlapping classes in feature space are increasing the difficulty of separating these classes from each other. Stratification has been used in previous research to improve the classification with overlapping classes (Hansen & Wendt, 2000; Hoppus & Lister, 2001; McRoberts et al., 2005; McRoberts et al., 2002). The main purpose of stratification is increasing the variability between classes and decreasing the variability within a single class (Blaschke et al., 2000; Burnett & Blaschke, 2003; Ricchetti, 2000). Stratification was not necessary for Samanko, as crops grew under similar conditions in this area; however, for Sougoumba the effect of stratification was assessed. Different landscape characteristics in the study area were known to have an effect on the crop performance. Strata of these characteristics have been tested on their influence on the classification accuracy.

Elevation was the first of three characteristics used for stratification. As mentioned, the study area of Sougoumba consisted of a higher plateau and a lower valley. The higher plateau was less suitable for farming as a hard crust made the soil almost impenetrable. Down in the valley the soft sandy alluvium was much easier to cultivate. To create the elevation strata the DEM from the Shuttle Radar Topography Mission [SRTM] was used (NASA, 2015). The elevation strata were categorized with ISODATA (Appendix 6A). A number of two classes was chosen, one for the plateau and one for the valley.

Soil was the second characteristic used to stratify the study area. This characteristic was also important for crop performance, as soil fertility, soil depth and soil structure are influencing crop growth. In addition, soil color can influence the measured reflectance. To create soil strata, the dataset created by Projet Inventaire des Ressources Terrestres (1983) was used. Although this dataset was created more than 30 years ago, it was still the most detailed and accurate soil dataset available for the study site of Sougoumba. For the soil strata the division was based on the polygons in the dataset; however, to be able to capture training data of all crop types per soil stratum, some soil types were merged based on their elevation and influence on crop performance. Relatively fertile alluvium soils were separated from the other, less fertile, soils. Furthermore, the alluvial soils were split in a lower and higher stratum (Figure 11).

The third landscape characteristic used for stratification was distance to buildup. As specified before, there is a decreasing soil fertility with an increasing distance from homesteads. Therefore, the Euclidian distance to buildup was calculated. The buildup itself was detected with the ISODATA technique applied on images early in the season (Figure 10). Classes which were mainly representing buildup were selected manually, pixels that clearly did not represent buildup in these classes were removed. After the buildup detection, the Euclidean distance to all buildings was calculated. Not all buildup was captured with this approach; however, in most of the larger homesteads some buildings were detected. As a result, the Euclidian distance to buildup was expected to be reasonably accurate. The distance raster was also stratified with ISODATA, this time a number of three classes was chosen (Appendix 6B).

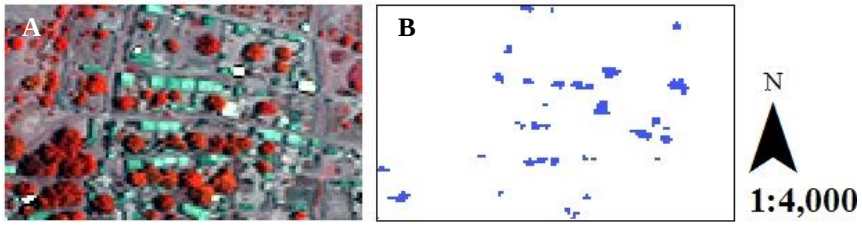


Figure 10. (A) WorldView-2 image of Sougoumba. The major part of the blueish pixels represent buildup (09-05-2015). (B) Buildup mask created with ISODATA.

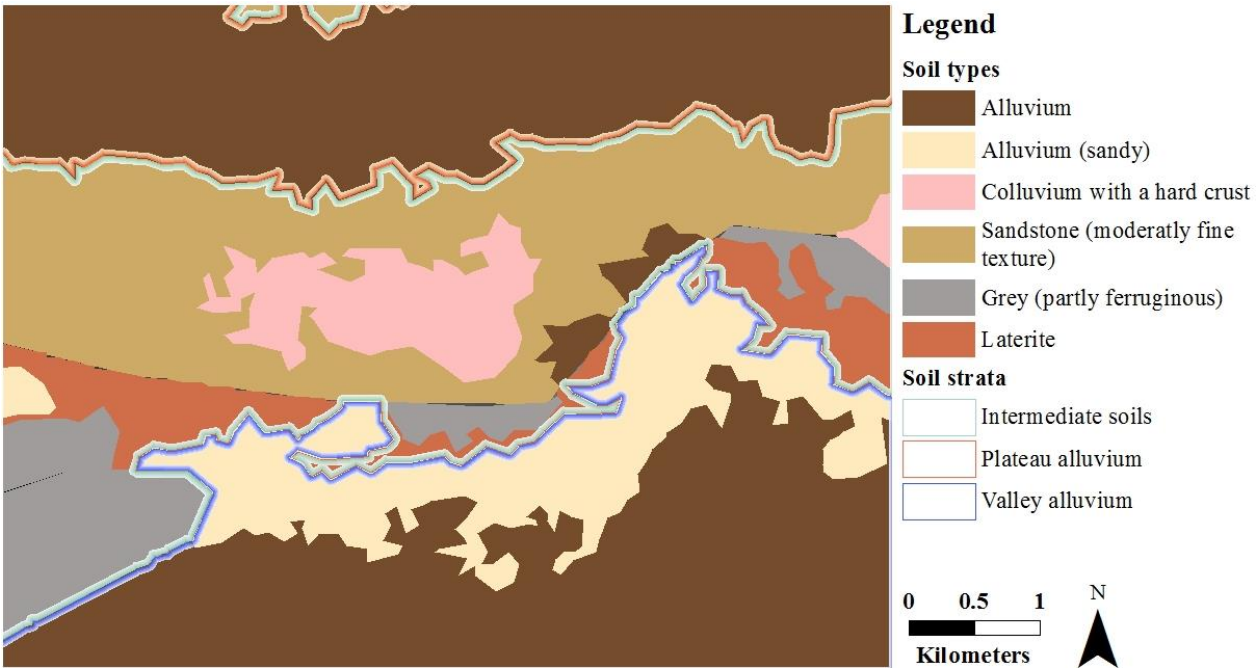


Figure 11. Soil types and strata in Sougoumba. Adapted from the soil dataset created by Projet Inventaire des Ressources Terrestres (1983).

2.6 Spatial and temporal resolutions

Different sensors in this research had variable spatial and temporal resolutions. To find the optimal resolutions of the different sensors in this research, three steps were taken. Classifier performance with different temporal resolutions was tested by removing bands, which represented dates, from the multi-temporal datasets systematically. Firstly, the temporal bands of the images of 2014 were removed until this dataset had an equal number of bands compared to the 2015 dataset. Bands representing imagery at dates closest to the dates in the 2015 dataset were kept (images of 29-07, 4-10 and 14-11 were removed). Secondly, the bands that were expected to be least informative were removed (Appendix 23A). For this the standard deviations between the mean training sample values of the classes were calculated for each date. The calculation was done with the training samples without stratification, because otherwise each stratum could have had different lowest standard deviations per class. The dates with the lowest standard deviation were removed first, as these represented dates at which crop classes were most similar. However, for this method the initial number of images has to be relatively large, while a lower number of images will save costs. Therefore, thirdly, this study tested if it was possible to only use images at the green up, peak of the season and senescence phase, without knowing the standard deviation between the classes (Appendix 23B).

In addition to the decrease in temporal resolution, the spatial resolution of the pre-processed images was aggregated stepwise (with a factor of: 2, 3 and 4). The effect of these aggregations and the temporal band removal on the accuracy of the most optimal classification was measured by applying the same (most optimal) methods that were used on the original data. Due to time constraints the tree- and bare soil mask of the un-aggregated images were used in this classification. The last step in investigating the spatial and temporal resolutions, was the combination of both aggregations. The accuracy of the two aggregations and the combination was also tested in a per-field aggregation. With this step the search for the optimal classifiers, vegetation indices and resolutions was completed and the most suitable classification method for crops on smallholder farms in Mali was determined.

3. RESULTS

3.1 Suitability of the UAV imagery

The suitability of the Tetracam images was checked by comparing the spectral signatures of reflectance panels in the imagery to the expected reflectance values for these panels. In Figure 12 it is visible that the lower measured reflectance values were in accordance with the reflectance standards (5 and 20 percent). In the blue and green parts of the spectrum the higher reflectances (45 percent) showed some deviation; however, this deviation is not having a large impact on crop recognition, as crops typically have a low reflectance in these wavelengths and the vegetation indices used are not based on wavelengths in the blue and green part of the spectrum.

In addition, the UAV images of different resolutions were compared to each other and to WorldView-2 imagery (Table 2 & Figure 13). Images of different resolutions had a strong correlation. The images of October 21st were showing the lowest, but still acceptable, correlation. Similarity with the satellite image was also reasonable, as the Pearson's correlations between the NIR and red bands of the satellite and UAV image used were respectively 0.62 and 0.74. This means that there was a moderate to strong positive relationship between the images of both platforms. These correlations were significant at a p-value smaller than 0.01. For the comparison to the satellite imagery, the UAV image with a 10 centimeter resolution, created at 13 November was used for three different reasons. Firstly, because this image needed to be aggregated less than the image with a 5 centimeter resolution. Secondly, because this image was captured on the exact same date as one of the satellite images. Finally, because both Tetracam images at this date showed the largest similarity. The correlation between the satellite and UAV imagery was thus strong. Moreover, the UAV images were correctly calibrated in the relevant spectral regions and showed a strong similarity to each other. As a result the images were found suitable for this study.

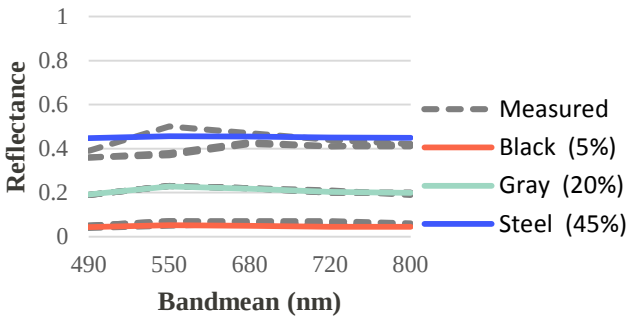


Figure 12. Measured spectral reflectances and the reflectance of their reflectance standards in the Tetracam images with a 10 cm resolution. The Tetracam images were acquired at 15-10-2015, 13-11-2015 and 18-11-15.

Table 2. Similarity of the Tetracam images with different resolutions (5 and 10 cm). Based on a linear model with intercept. * All results were significant at $p < 0.01$.

Date	Pearson's r	R^2
15-10-15	0.82*	0.68
21-10-15	0.75*	0.56
23-10-15	0.80*	0.64
26-10-15	0.89*	0.79
29-10-15	0.87*	0.76
02-11-15	0.93*	0.86
05-11-15	0.92*	0.85
09-11-15	0.89*	0.78
13-11-15	0.95*	0.90
18-11-15	0.91*	0.82

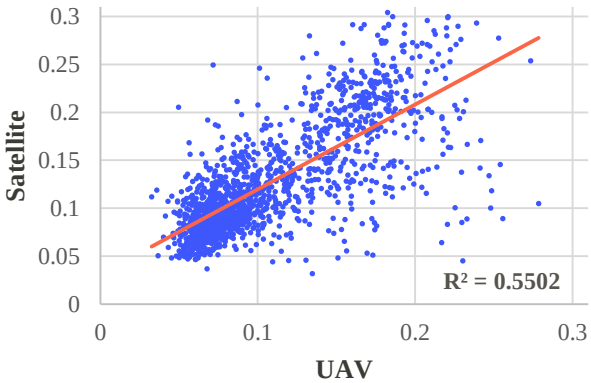


Figure 13. NIR values of the WorldView-2 (832.5 nm) and aggregated Tetracam image (800 nm) of 13-11-2015 plotted against each other. The initial resolution of the Tetracam image was 10 cm.

3.2 Classification of Samanko

3.2.1 Mono-temporal distinguishability and classification

In the area of Samanko the within-field variation was low; therefore, the spectral signatures of the crops were expected to be separable. After the calibration of the UAV images, it was possible to validate this assumption and derive the

spectral crop characteristics. Visible in Figure 14B is that the assumption was correct, as the mean spectral signatures of the crops were distinguishable. When looking at the feature space plot, it becomes clear that distinguishing sorghum from the other crops was most challenging due to the spectral overlap with millet (Figure 14). Cotton however, was clearly brighter in the near-infrared region compared to the other crops. Another interesting finding, which is shown in the feature space plot, was that sorghum had a relatively large number of outliers outside the 95 percent confidence ellipse.

Nonetheless, based on the spectral profiles it was expected that a single image classification would be relatively accurate, as the classes were clearly separable in the NIR region. Indeed, all classifications based on spectral information had an overall accuracy of above 72 percent (Appendix 7A). Of these three crops the most difficult crop to classify was sorghum (Table 3). From the five tested classifiers the minimum distance and spectral angle were least accurate. K-NN, regression tree and maximum likelihood were more promising classifiers. These methods classified more than 80 percent of the validation data correct.

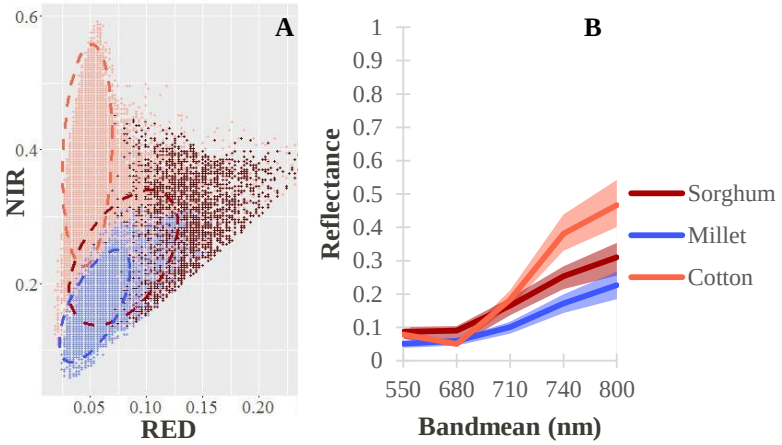


Figure 14. (A) Feature space plot of the NIR (740nm) and RED (680nm) band with the 95 percent confidence ellipse. (B) Mean spectral profiles plotted with the lower and upper quartile. Based on 10 cm Tetracam image: 15-10-2015.

Table 3. Average confusion matrix of five spectral maximum likelihood classifications for Samanko 15-10-2015

	Sorghum	Millet	Cotton	Total	User accuracy
Sorghum	**15.80**	3.20	1.40	20.40	0.77
Millet	3.80	**16.80**	0.00	20.60	0.82
Cotton	0.40	0.00	**18.60**	19.00	0.98
Total	20.00	20.00	20.00	60.00	
Producer accuracy	0.79	0.84	0.93		
Overall accuracy	0.85				
Kappa	0.78				

3.2.2 Multi-temporal distinguishability and classification

Adding temporal information was expected to increase the variability between and decrease the variability within the classes. For Samanko it was only possible to monitor the senescence of the crops; still, differences between crops were already visible in this phenological phase (Figure 15). Cotton was again the crop that was clearly separable. In addition, the mean and interquartile range of sorghum and millet were showing no overlap in the later part of the senescence phase, as the NDVI of millet was decreasing faster. The different senescence rates of sorghum and millet is also clearly visible in a Red-Green-Blue [RGB] image of the temporal NDVI (Figure 16). As with the spectral signatures of a single image, the mean temporal sorghum profile was located between the profiles of the other crops. As a result, sorghum pixels were sometimes misclassified as cotton or millet in the classified raster (Figure 17). For Samanko these misclassifications were acceptable, as still the majority of the pixels in a field were assigned to the right class (Paragraph 3.2.3).

Like the spectral profile, the temporal profile of the crops showed some overlap. However, the within-class variability was decreased, the standard deviation of each temporal class was relatively low compared to the standard deviation of the classes in a single image (Appendix 13). Therefore, the classification accuracy increased with an average of 1.5 percent for the K-NN and regression tree method (Appendix 7B). The maximum likelihood classifier

was, however, clearly underperforming in the multi-temporal classification compared to the K-NN and regression tree classifiers. As a result, only the K-NN and regression tree classifiers were used in the classifications of Sougoumba.

In this research five vegetation indices were used to create the multi-temporal datasets (Appendix 7C). From all these datasets the NDVI dataset was performing the worst. The accuracy of all the soil-adjusted vegetation indices was almost similar. To be able to analyze the performance of both types of vegetation indices, the most accurate ratio and orthogonal-based vegetation indices were used on all following classifications in this study (PVI and TSAVI).

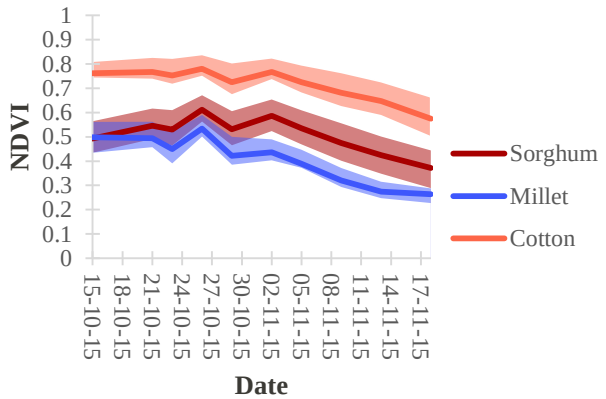


Figure 15. Mean NDVI profiles plotted with the lower and upper quartile. Based on Tetracam images of Samanko.

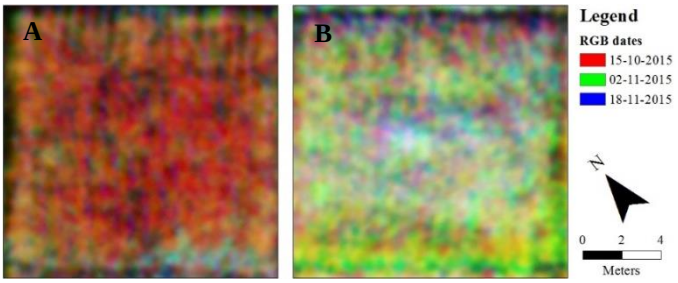


Figure 16. RGB image of temporal NDVI calculated on the 10 cm Tetracam dataset of Samanko. (A) Millet. (B) Sorghum.

3.2.3 Per-field aggregation

The last step was the per-field majority filter of the study site of Samanko. For this the classified rasters of the most accurate methods and vegetation indices were used. For all per-field aggregations the result was 100 percent accurate. In Figure 17 it is clearly visible that most frequent class in each field was indeed the correct class.

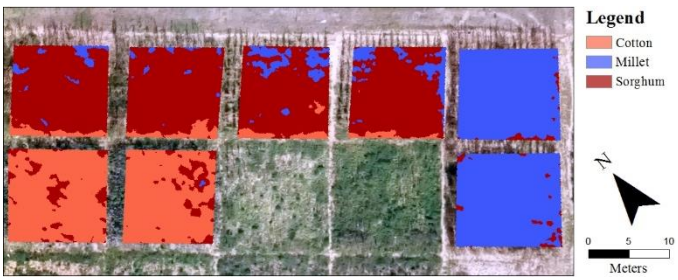


Figure 17. Pixel-based temporal classification result with the K-NN method applied on the PVI dataset. RGB image: 12-11-2015.

3.3 Classification of Sougoumba without stratification

3.3.1 Mono-temporal distinguishability and classification

Without stratification crops were showing a lot of spectral overlap in Sougoumba (Figure 18 & Appendix 8A-B). Classes were containing a large range of reflectance values and the mean spectral signatures were almost similar. Because of this inseparability, the classification was expected to be less accurate compared to the classification of Samanko.

Both 2014 and 2015 were classified using the two classifiers that were best for the classification of the area of Samanko. The accuracies of the classifications of Sougoumba were relatively low compared to the accuracies of Samanko. Secondly, the classification of 2015, based on the training dataset of de Schaetzen (2015), was clearly underperforming. With the initial classification of 2015 the fields of de Schaetzen (2015) were used. When, however, the STARS training dataset was used for 2015, the overall accuracy increased with about 7 percent (when using the K-NN method), but was still a lower accuracy than the overall accuracy of 2014 (Figure 19, Appendix 9A-C).

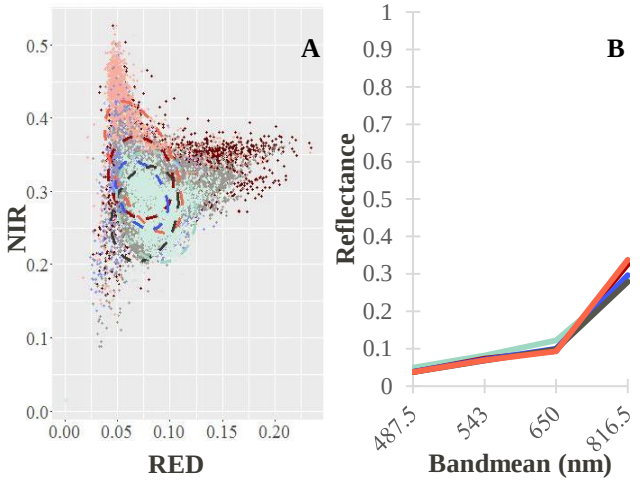


Figure 18. (A) Feature space plot of the NIR (816.5nm) and RED (650nm) band with the 95 percent confidence ellipse. (B) Mean spectral profiles. Based on Quickbird image: 04-10-2014.

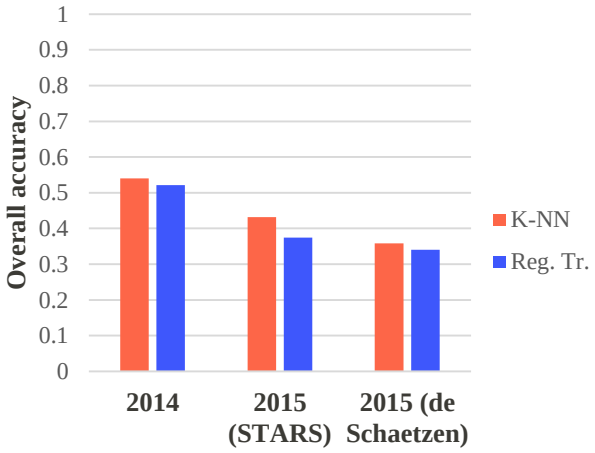


Figure 19. Overall accuracies per classifier for the single image classification of Sougoumba. Based on the satellite images of: 04-10-2014 (Quickbird) and 25-09-2015 (WorldView-2).

3.3.2 Multi-temporal distinguishability and classification

The temporal profiles of the crop classes in Sougoumba showed an increase in distinguishability compared to the spectral profiles based on a single image (Figure 20, Appendix 8C & Appendix 8E). At the start of the season the vegetation index values were, however, almost similar; in the green up, flowering and senescence phase the temporal signature started to deviate slightly. In Sougoumba the same trend was visible as in Samanko. In the senescence phase cotton, sorghum and millet followed different average trends. Again the NDVI of millet was decreasing faster than the NDVI of sorghum. This confirms the validity of both measurements.

However, the multi-temporal classification of Sougoumba was based on the two indices which were most optimal in the multi-temporal classification of Samanko (Figure 21, Appendix 9D-F). When the PVI was used, the overall classification accuracy improved with more than 10 percent for all classifiers compared to the single image classification. The classification with the TSAVI resulted in slightly lower overall accuracies compared to the PVI classification (difference of 0-2 percent). In addition, the classification based solely on the farmers participating in the STARS project yielded higher accuracies. Nevertheless, differences in overall accuracy decreased between both years even when the fields of de Schaetzen (2015) were used. Moreover, the classification based on this dataset yielded the largest improvement compared to the single image classifications. Although the accuracies became more similar, the differences between the two optimal methods increased: the regression tree classifier was clearly performing worse in the multi-temporal classification.

To be able to explain the improvement in accuracy, feature space plots were created based on the classification with the PVI (Figure 22 , Appendix 8D & Appendix 8F). The classes showed less overlap in the 95 percent confidence ellipse compared to the mono-temporal classes. What was striking is that the multi-temporal dataset of 2015 was less extensive, but yielded similar accuracies as the classification of 2014 when only the STARS fields were used. Apparently, a higher temporal resolution does not have to yield a higher accuracy. This finding was tested in depth in Paragraph 3.6.1.

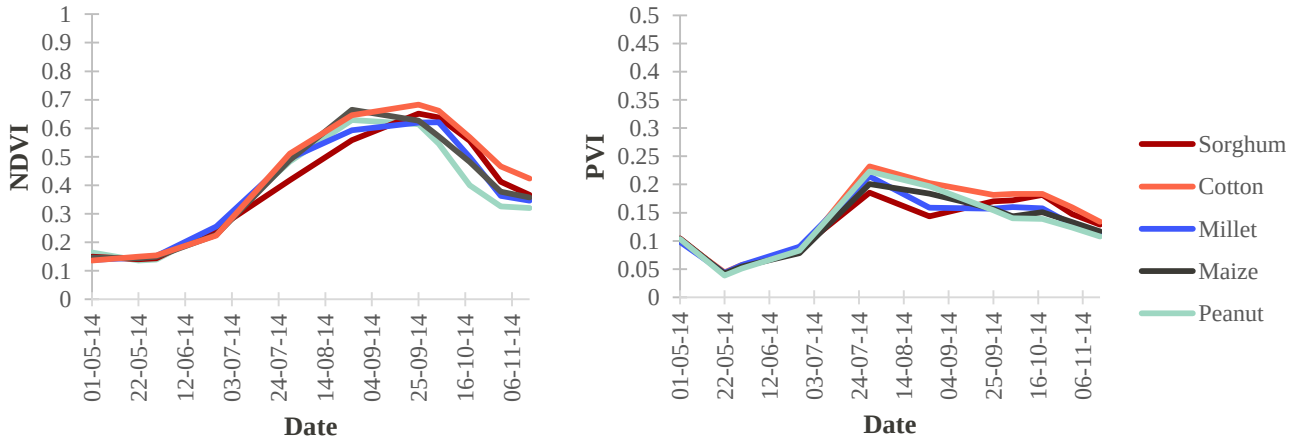


Figure 20. Temporal mean NDVI and PVI profiles. Based on satellite images of Sougoumba 2014.

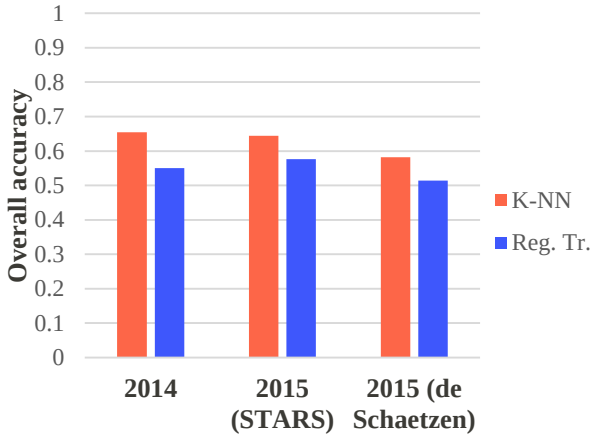


Figure 21. Overall accuracies per classifier for the multi-temporal PVI classification of Sougoumba.

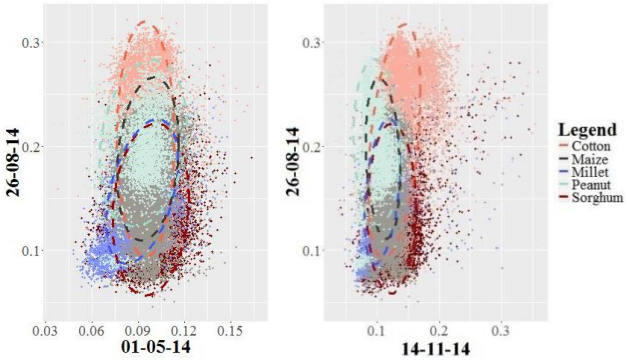


Figure 22. Feature space plot of the early-middle and middle-end season PVI classes of 2014 with the 95 percent confidence ellipse.

3.4 Sougoumba with stratification

3.4.1 Mono-temporal distinguishability and classification

Stratification was expected to enhance the class separability and classification accuracy even more. Therefore, the study area of Sougoumba was stratified using soil types, distance to buildup and elevation. Stratifying the area improved the single image classification with an average of 8 percent, when the most optimal stratification of soil was applied. The overall accuracy with this strata was 2-3 percent higher compared to the other stratifications (Appendix 12A-C). The disadvantage of the soil stratification was that in 2014 there were no peanut fields present in the second soil stratum. However, this was not visible in the accuracies, as the classification of 2014 yielded a slightly lower accuracy (compared to the classification of 2015 with the STARS fields). In all cases the K-NN method was more optimal than the regression tree method. Furthermore, the classifications based on the smaller field boundary datasets of the STARS farmers yielded better results. Finally, the difference in accuracy between the classifications based on the two training datasets of 2015 was decreased.

Feature space plots of the soil strata show that the increase in accuracy was caused by an improvement in class separability and compactness (Figure 23 & Appendix 10). The stratum with the valley soils was most easy to classify, followed by the stratum with the alluvium soils located on the plateau. The last stratum, located between the other two strata, yielded the worst accuracies (Figure 24). Despite the fact that stratification increased the accuracy of the mono-temporal classification, the multi-temporal classification without stratification was still more accurate (average difference of 10.7 percent in overall accuracy).

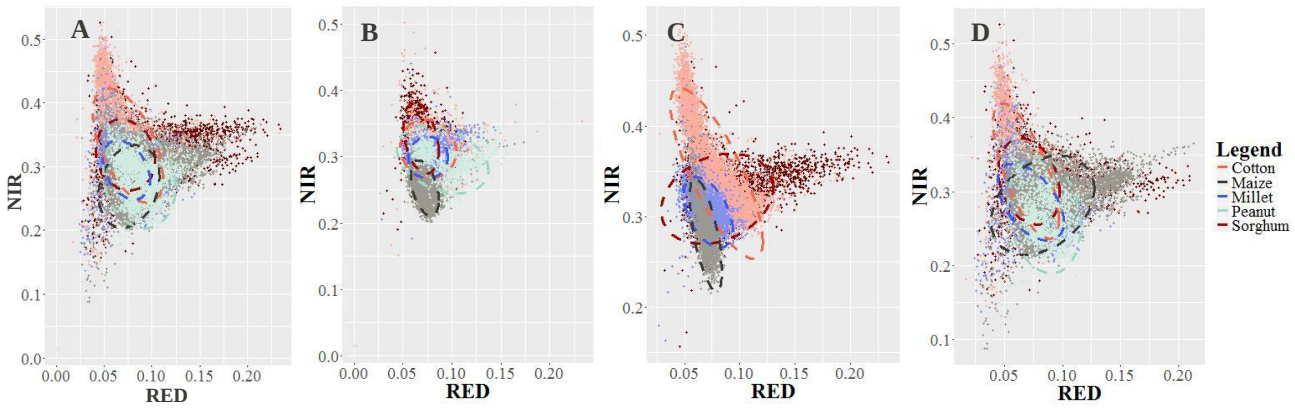


Figure 23. Feature space plots of the NIR (816.5nm) and RED (650nm) band with the 95 percent confidence ellipse. Based on Quickbird image: 04-10-2014. (A) Without stratification. (B) Plateau alluvium stratum. (C) Valley alluvium stratum. (D) Intermediate soils stratum.

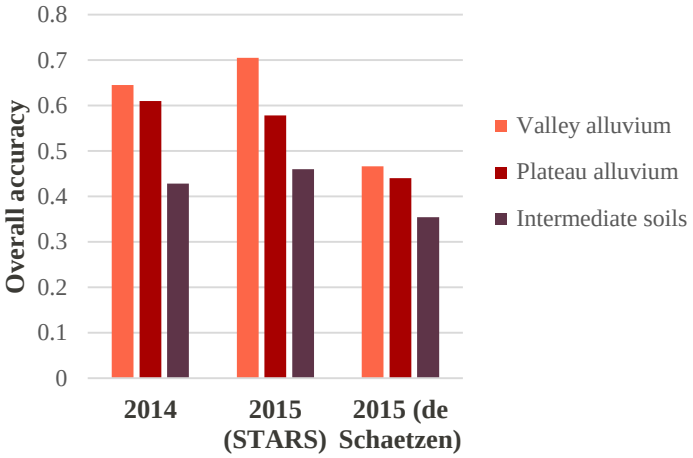


Figure 24. Overall accuracies of the classification using a single date per soil stratum. The classifications are based on all training datasets in both years (using K-NN).

3.4.2 Multi-temporal distinguishability and classification

Compared to the unstratified multi-temporal classification, all stratified classifications increased the overall accuracy with an average of 10 percent when the most optimal approaches were used (Appendix 12D-F). Again, the K-NN method proved to be superior to the regression tree method and the differences between classifications, based on temporal TSAVI and PVI datasets, were minimal. The stratification of soil was, however, not the most optimal stratification for all classifications, as for the dataset of de Schaetzen (2015) stratification of elevation yielded a higher accuracy. In addition, differences between the results based on the three different training datasets were larger compared to the unstratified multi-temporal classification. Another result was that the classification of the year 2014 again was more accurate than the classification of 2015 when the STARS dataset was used.

Feature space plots of the crop-classes are shown in Appendix 11 and Figure 25**.** The classes were more compact and showed less overlap compared to the unstratified classes (Figure 22). Especially for the dataset of de Schaetzen (2015), the differences between the strata were clearly visible. Same as with the single image classification the classification accuracy was highest in the stratum with the valley alluvium soils, followed by the other strata (Figure 26). Although the differences in classification accuracy increased between the classifications based on the three training datasets, the stratification and classification of multi-temporal data yielded higher accuracies than the mono-temporal stratified classification.

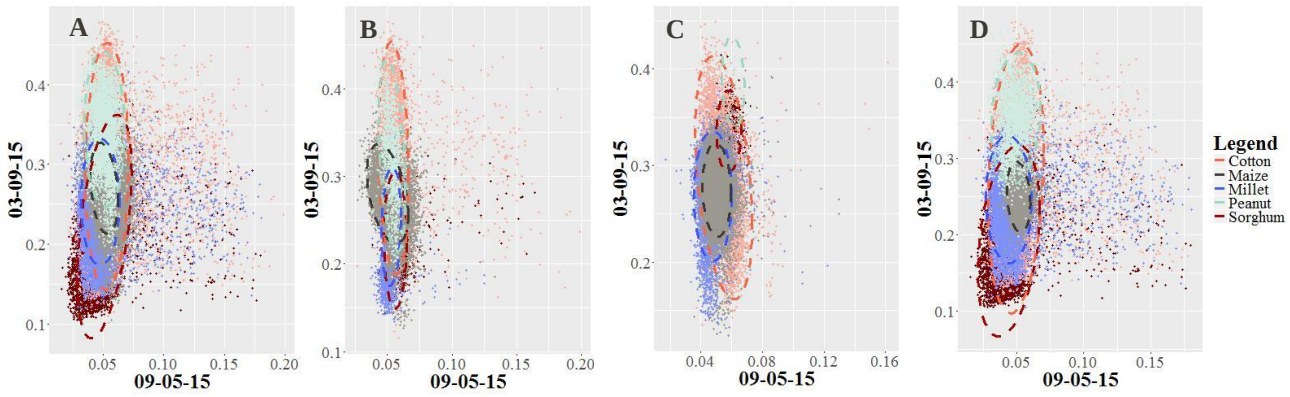


Figure 25. Feature space plots of the early-middle season PVI classes with the 95 percent confidence ellipse of 2015 (using fields of farmers in the STARS project). (A) Without stratification. (B) Plateau alluvium stratum. (C) Valley alluvium stratum. (D) Intermediate soils stratum.

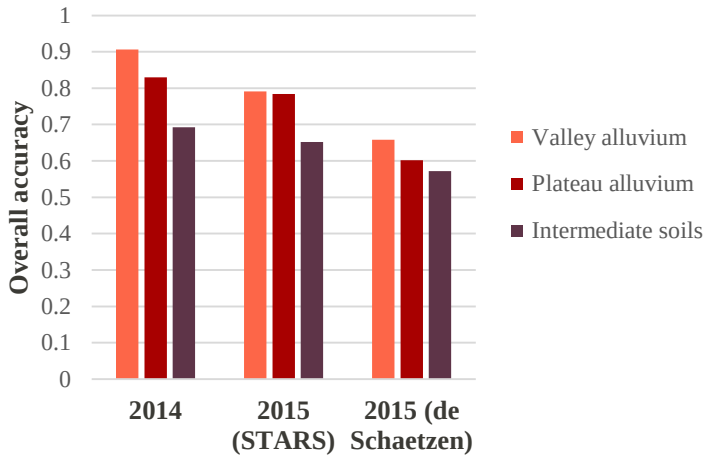


Figure 26. Overall accuracies per soil strata for the classifications based on all training datasets in both years (using K-NN with the temporal PVI series).

3.5 Per-field aggregation

From the previous paragraphs it is clear that in all classifications the multi-temporal classification with the K-NN method was superior. Stratification increased the accuracies even more and the PVI and the soil strata were in most cases optimal for the classification of Sougoumba. Therefore, the multi-temporal classification result per soil stratum with this method and index was used for the per-field aggregation.

With the per-field majority filters of 2014 and 2015, based on the STARS dataset, the overall accuracy was increased with 4 and 8 percent, respectively (to 85 and 82 percent) (Table 4 & Appendix 16A-B). When the dataset of de Schaetzen (2015) was used, the accuracy was increased with 5 percent. For 2014 peanuts had the lowest producer accuracy, while millet was the least reliable class. In 2015, for the classification with training data derived from the fields of de Schaetzen (2015), the cotton fields were most difficult to classify; however, the reliability of the classified cotton fields was relatively high. Peanuts and sorghum were the two crops that were both showing a relatively low user and producer accuracy in this case. When the STARS fields were used to create training data for 2015, the cotton fields were misclassified the most. All per-field aggregations thus had deviating results.

Because two training datasets were available for 2015, the classified raster of the dataset that yielded the highest accuracy in the pixel-based classification (STARS dataset) was also used in the per-field aggregation of the other dataset (de Schaetzen (2015) dataset). However, this aggregation was less successful and resulted in an overall accuracy of 51 percent (Appendix 16C).

Table 4. Confusion matrix of the per-field aggregation results. Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation).

	Sorghum	**Cotton**	**Millet**	**Maize**	**Peanuts**	**Total**	**User accuracy**
Sorghum	**68**	26	20	0	3	117	0.58
Cotton	11	**98**	2	7	5	123	0.80
Millet	23	8	**88**	8	2	129	0.68
Maize	5	27	8	**60**	1	101	0.59
Peanuts	3	5	1	6	**19**	34	0.56
Total	110	164	119	81	30	504	
Producer accuracy	0.62	0.60	0.74	0.74	0.63		
Overall accuracy	0.66						
Kappa	0.56						

3.6 Spatial and temporal resolutions

3.6.1 Optimal temporal resolution

To measure the influence of different temporal resolutions datasets of 2014 and 2015 were matched to each other. The number of bands was reduced to eight for the 2014 dataset and the optimal classification approach was applied to this new dataset. As a result, the overall accuracy of the classification of 2014 decreased from 81 to 76 percent. Next the temporal PVI bands were removed in the order presented in Appendix 23A. For each removed band the classification was carried out again. In Figure 27 it is visible that a lower temporal resolution was not decreasing the accuracy in all cases. Even when only five images were used the overall accuracies of the pixel-based classification and per-field aggregation stayed almost the same (Appendix 18 & Appendix 20).

As mentioned it was also tested if it was possible to only use images acquired at the green up phase, peak of the season and senescence phase (Appendix 23B). These images were chosen without knowing the within-variability or overlap of the classes in feature space. Classification based solely on these dates was relatively inaccurate and resulted in a reduction in overall accuracy of 10-16 percent. Almost all Kappa's per stratum show a similar pattern when both classifications with three images are compared (Figure 28).

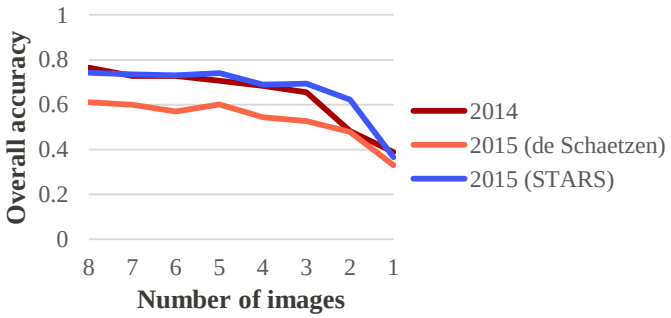


Figure 27. Overall accuracy with different temporal resolutions for the pixel-based classifications based on all training datasets (using K-NN with the temporal PVI series).

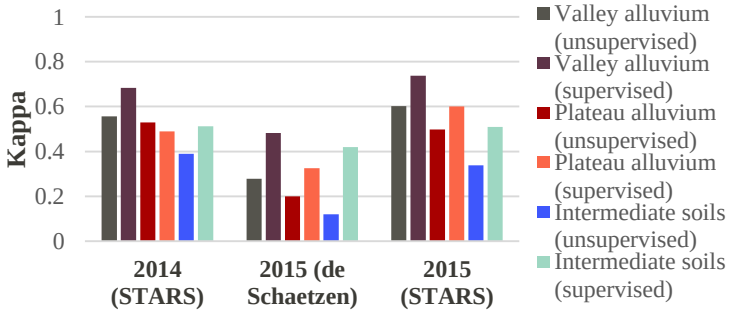


Figure 28. Kappa's per stratum for the pixel-based temporal K-NN classifications based on all training datasets. The unsupervised Kappa's represent the classification results on the three temporal PVI images selected without knowing the within-variability or overlap of the classes. The supervised Kappa's represent the results when the three PVI images that showed the largest deviation between the crop classes were used.

3.6.2 Optimal spatial resolution

In addition to the temporal resolution, different spatial resolutions were expected to result in different accuracies. To be able to validate this, the PVI rasters were aggregated to resolutions of 4, 6 and 8 meters (a factor of 2, 3 and 4). On those aggregated datasets the best classifier was applied. With each classification the Kappa increased (Table 5). Aggregation with a factor 4 was most promising, as for both classifications of 2015 the overall accuracies increased with an average of 6 percent (Appendix 19). For 2014 there was, however, no significant improvement. Distribution of the accuracies in each stratum and average Kappa was in accordance with the un-aggregated stratified classification (Figure 29). An interesting result is that if the spatially aggregated rasters were used in the per-field aggregation, only the accuracy of the classification based on the de Schaetzen (2015) training data was improved (to 68 percent overall accuracy). Moreover, the per-field aggregation for 2014 yielded lower accuracies after the aggregation (Appendix 21).

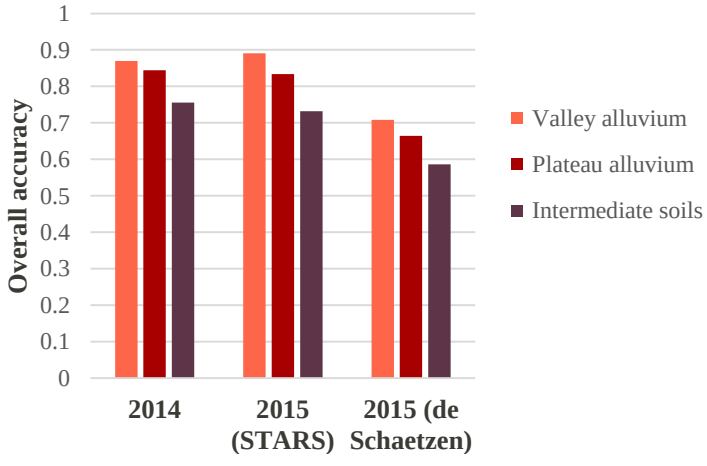


Figure 29. Overall accuracies per strata for the classifications based on all training datasets in both years (with a factor 4 aggregation of the PVI using the K-NN classifier).

Table 5. Average Kappa per spatial aggregation for the classifications based on all training datasets in both years (using K-NN with the temporal PVI series).

Training dataset	2 m	4 m	6 m	8 m
2014 (STARS)	0.74	0.75	0.75	0.77
2015 (de Schaetzen)	0.51	0.53	0.55	0.57
2015 (STARS)	0.68	0.75	0.79	0.78

3.6.3 Optimal combination

To check if the spatial aggregation also worked on the datasets with a lower temporal resolution, a pixel-based classification with the combined optimal resolutions was carried out. This resulted in overall accuracies of 79 percent for both pixel-based classifications based on the STARS training datasets. The accuracy of the pixel-based classification, based on the training data of de Schaetzen (2015), increased from 60 to 63 percent. A per-field aggregation improved the overall accuracies with the optimal resolution to 79, 67 and 86 for respectively the classifications based on training datasets of STARS (2014), de Schaetzen (2015) and STARS (2015) (Table 6 & Appendix 24).

Table 6. Confusion matrix of the per-field aggregation results with five dates and a spatial resolution of 8 meters. Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation).

	Sorghum	**Cotton**	**Millet**	**Maize**	**Peanut**	**Field Count**	**User accuracy**
Sorghum	**67**	24	22	0	3	116	0.58
Cotton	17	**104**	2	8	5	136	0.76
Millet	19	6	**88**	6	1	120	0.73
Maize	6	23	5	**59**	1	94	0.63
Peanut	1	7	2	8	**20**	38	0.53
Field count	110	164	119	81	30	504	
Producer accuracy	0.61	0.63	0.74	0.73	0.67		
Overall accuracy	0.67						
Kappa	0.57						

4. DISCUSSION

4.1 Suitability of the UAV imagery

Based on different tests the UAV imagery was found to be suitable for this research. The Tetracam images were in most cases strongly correlated to each other and to WorldView-2 imagery. Nevertheless, three deviations were found. Firstly, the Tetracam images showed some small deviations with the reflectance standard of 45 percent in the visible part of the spectrum (especially the blue and green region). This was possibly caused by errors made in the sensor calibration, image transformation or panel creation. A second finding was the relatively low correlation between the two UAV images captured at the 21st of October 2015 (of 5 and 10 centimeter spatial resolution). This was most likely caused by orthophoto alignment issues. In Figure 6 it is visible that the image of the 21st of October was originally misaligned to the reference image. Figure 30 shows that the fields indeed had curving crop lines while in reality they should have been straight. The re-alignment with the polynomial order of 3 has shifted the image, but some erroneous curves were not fully corrected. Therefore, the pixels of both images were not overlapping completely, which most likely has caused the lower correlation. Thirdly, the comparison with the satellite image was successful, but showed a lower correlation than the comparison between UAV images.

Five factors can explain the difference between the images of both platforms. Firstly, the sensor band means and widths were dissimilar (Appendix 2 & Appendix 5). Because sensors were acquiring images at different wavelengths, the measured reflectances slightly differed. Secondly, the spatial resolution of the sensors were dissimilar. Therefore, the UAV imagery with a 10 centimeter resolution had to be resampled to match the satellite resolution of 2 meter. As values were resampled using bilinear interpolation, the resampled Tetracam image was smoother than the WorldView-2 image. With interpolated resampling pixels that are initially surrounding the resampled pixels are influencing the value assigned to that pixel. Eventually this method should be more accurate than a nearest neighbor assignment (which loses information by only assigning the closest cell value of the original image to the resampled image), but also causes boundaries in the image to fade (which makes the images less similar). Thirdly, the images were not aligned to each other, which influenced the similarity negatively. Fourthly, satellite images are influenced by atmospheric conditions. Although the images were corrected the atmosphere might still have negatively influenced the accuracy of the satellite imagery. At last, the overpass time of the satellite is most likely not completely similar to the time at which the UAV captured the images. This means that the shadow sizes differed in both images, making them less similar.

Matese et al. (2015) found comparable correlations between UAV and satellite imagery and stated that the choice for a specific platform is, therefore, dependent on the goal of the research. With satellite imagery it is harder to measure the within-field heterogeneity, but easier to create a temporal dataset over multiple seasons and multiple years. Still, the precision and resolution of UAV imagery is, in general, more optimal, which makes the UAV imagery highly suitable when the possibilities of a crop classification are researched. However, when a larger area like Sougoumba is classified, satellite imagery is still more useful. The range, reliability and price of satellite images are still outweighing the advantages of the UAV imagery.

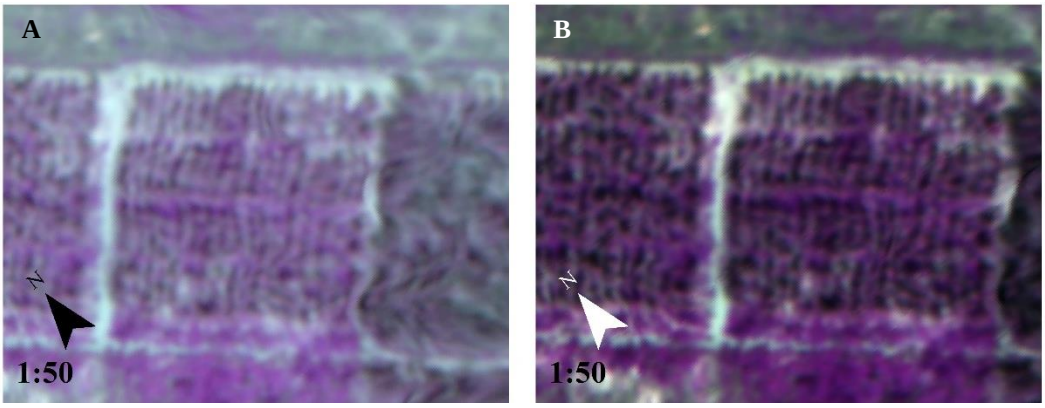


Figure 30. (A) Curved crop rows in the original image. (B) Curved crop rows in the optimized image. Tetracam image: 21-10-2015.

4.2 Classification of Samanko

4.2.1 Mono-temporal distinguishability and classification

In Samanko crops were clearly separable based on spectral information of a single Tetracam image. In the visible region the crops were showing similar reflectances. However, in the NIR region there was hardly any overlap. Cotton was brightest in the NIR, which indicated that this crop had a different structure or was healthier at the time of data

acquisition. Most likely the crop structure was having the largest impact on this difference in reflection. The reason for this is that sorghum and millet are grain crops with a tall stem, while cotton is a bush with narrow leaves (Figure 3).

Sorghum was most difficult to classify with a single image. In the feature space plot it is visible that this crop-class had extreme values in the NIR bands. The most likely cause of these outliers was the presence of weed in the fields. When looking at an RGB image created later in the season, indeed weed was present in the sorghum fields (Appendix 1A). Because only bare soil was removed from the image, these greener pixels were still influencing the classification of this crop. In addition, the spectral signature of this class was located between the signature of cotton and millet. As a result, sorghum cells with a higher or lower spectral profile than the class range had a higher chance to be misclassified. While, for example, a spectral profile which was significantly higher than the average cotton profile still was classified as cotton.

Several classification methods have been tested on their accuracy. The minimum distance and spectral angle mapper methods were performing worst. Most likely because these classifiers are more sensitive to overlapping classes and outliers than the other three classification methods which were more accurate. The K-NN method assigns a pixel to a certain class by looking at the *k* nearest neighbors. The sensitivity is lower because multiple neighbors are taken into account. The regression tree method is less sensitive because only the most informative bands are used per class, reducing the influence of class overlap. From all classifiers the maximum likelihood classification was most accurate. Because this method uses probabilities to assign pixels to a specific class the problem of overlapping classes was mostly overcome.

Hubert-Moy et al. (2001) also found the minimum distance classifier to be the poorest classifier when classes had overlapping boundaries, while the maximum likelihood was more optimal. However, in the research of Castillejo-González et al. (2009) the maximum likelihood classifier yielded the lowest accuracy. This confirms the statement done by South et al. (2004, p. 91): "There is no ideal classification routing to suit all needs and requirements". Per study the suitability of multiple methods should thus be investigated to obtain the most optimal results**.** Also in further research, using the same study area, it is recommended test multiple classifiers, as different training classes will influence the performance of each method.

4.2.2 Multi-temporal distinguishability and classification

For Samanko, the temporal distinguishability was slightly higher than the spectral distinguishability in a single image. This was mainly caused by a lower within-class variability. The crops had a comparable phenological development and therefore the classes were still showing overlap. However, even with multi-temporal classifications on larger fields than Samanko, this remains an issue which is caused by crop characteristics and farmer management practices (Wardlow et al., 2007). The reason for this is that crops sowed on closely related dates will most likely follow a roughly similar phenological trend; in addition, irrigation and fertilization practices are influencing the variability within a single crop class. Although temporal profiles followed similar trends in Samanko, they were separable. Cotton kept the highest average vegetation index values over time. This implies that cotton has a slower senescence phase compared to the other crops, as all crops were sowed at the same time and managed under similar conditions. In addition, millet had relatively high index values at the start of the measurements, but the values decreased faster compared to sorghum. Millet was thus decaying faster, which is most likely caused by crop characteristics. Most likely the decreased within-class variability (compared to the mono-temporal classes), caused the multi-temporal classification results to be slightly more accurate than the mono-temporal results.

Again the K-NN and regression tree methods were most accurate in recognizing crop types for all indices. The reasons for their performance are similar to the reasons given in the previous paragraph. The maximum likelihood classifier was, however, clearly performing worse in the multi-temporal classification. Because the maximum likelihood method is assuming normally distributed classes this classifier was yielding lower accuracies. When looking at Figure 15 it is visible that the mean of both cotton and sorghum was smaller than the median. This means that the classes have a negatively skewed distribution. Non-normally distributed classes are common in the classification of vegetation (Foody, 1996; Varela et al., 2008). One way to cope with this issue is to normalize the classes (Holmgren & Persson, 2004). Another method is to use non-parametric classifiers, which are not assuming normal distributions (Aci et al., 2010; Foody, 1996). The current research has not normalized the distribution of the classes and, therefore, it might be possible that with a transformation the parametric classifiers are more suitable for a multi-temporal classification. Future research is needed to investigate how parametric classifiers applied on normalized training data are comparing to the non-parametric classifiers. Nevertheless, when the approach of this research is followed, non-parametric classifiers are more optimal.

From the five tested vegetation indices the NDVI was least suitable, which most likely was caused by the fact that this index is not correcting for underlying soil. Although all bare soil pixels were removed before the classification, pixels which partly consisted out of bare soil might not have been removed. Soil-adjusted vegetation indices are reducing the possible negative impact of these pixels and are, therefore, more suitable for classification of the crops in Samanko. In addition, the NDVI tends to saturate at higher vegetation densities, as a result crops growing under optimal conditions with a high leaf area index could be harder to separate (Chuvieco & Huete, 2009). A single optimal vegetation index can, however, not be defined. Indices perform different under varying conditions (different study

areas, imagery etc.). For this reason, researchers, that investigated the suitability of vegetation indices more in depth, did find deviating results. That is why scholars as Broge & Leblanc (2001) and Rondeaux et al. (1996) state that the choice for the most optimal vegetation index should always be based on the characteristics of the study area, purpose of the study and the amount of a priori information available. If a similar research would be conducted it is thus of importance that the most suitable vegetation index is determined in a similar way, instead of using the same indices found in this study.

It was expected that classifications based on a single image are sub optimal for crop classification compared to multi-temporal classifications (Conrad et al., 2010; Van Niel & McVicar, 2004). However, for Samanko this was not the case, as the difference in overall classification accuracy was only 3 percent. In the current research the temporal classes were more compact than the single date classes, but the difference was minimal. This research, therefore, partly undermines the findings of previous researchers. It must be stated that these scholars were classifying areas where the within and inter-variability of the classes was relatively high. Nevertheless, in areas where crops are growing under optimal conditions a multi-temporal classification is not necessarily more optimal. Of course the spectral signature of the crops has to be distinguishable at certain moments in time. For Samanko the spectral profiles were distinguishable while the crops were still flowering. In other study areas, crops might be separable in other phenological phases (Wardlow et al., 2007). Therefore, it is important to notice that this research is not claiming that flowering crops growing under optimal conditions are always classifiable with a single image. When, however, the date at which crops are separable is known, a mono-temporal classification is sufficient for well-managed crops. Still, when smallholder fields with a high variability are classified a multi-temporal classification is highly likely to outperform a classification based on a single image.

4.3 Classification of Sougoumba without stratification

4.3.1 Mono-temporal distinguishability and classification

Compared to Samanko, the overlap was relatively large for the mono-temporal classes in Sougoumba. Variation in the fields caused most classes to cover a wide range of red and NIR reflectances. Because bare soil and trees were removed, the overlap and within-variability were most likely caused by weed or mixed crops in the fields. In addition, spectral mixing of vegetation with bare soil most probably has disturbed the measured spectral reflectances. For Sougoumba this mixing played a bigger role than for Samanko, as the resolution of the imagery was higher. In addition, more classes were used in Sougoumba compared to Samanko, which decreased the separability. Also in the study of Wardlow et al. (2007) it was found that a smaller number of classes increases the distinguishability. Van Niel & McVicar (2004) handled this issue by classifying and, afterwards, manually removing the pixels of highly accurately classified crops from further classification. A possible next question is if this method will also be effective for Sougoumba.

The difference in accuracy between the single image classifications of 2014 and 2015 (using de Schaetzen (2015)) was remarkable. Most likely, the reason for this difference was that the training samples for each crop were containing a larger range of measured spectral reflectances in 2015 when the training dataset of de Schaetzen (2015) was used (Appendix 8A-B). This was due to the fact that there were more training fields present in this training dataset of de Schaetzen (2015). More fields means a more complete dataset, but with the disadvantage of a higher variability. In Figure 18, Appendix 8A-B and Appendix 14A-C it is therefore visible that the classes were less variable when only the fields of the STARS dataset were used for the classification. This could be caused by a variety of reasons; crop management is one of them. However, there was no ancillary data attached to the dataset of de Schaetzen (2015). For future research it is, therefore, recommended to collect ancillary data about the field management for all training datasets, as this will help in explaining the differences between the accuracies of the classifications.

When comparing Appendix 14A-C with Appendix 13A it is visible that the class standard deviations in Sougoumba were not lower than the standard deviations in Samanko. However, the feature space plots of Sougoumba show a larger overlap in the 95 percent confidence ellipse (compare Appendix 8A-C with Figure 14). Understanding of factors that cause this inter-variability and using them for a stratified classification was thus highly relevant.

4.3.2 Temporal distinguishability and classification

The class distinguishability was increased by adding temporal information, as the temporal classes showed less overlap than the classes in a single image. However, at the start of the growing season the vegetation index values were almost similar. The reason for this is that the fields were still bare and, therefore, the measured index only represented the soil. Later in the season the temporal signatures started to deviate in the green up and senescence phase. This increase in separability in these phenological phases was also found by others (Doraiswamy et al., 2003; Wardlow et al., 2007). Still, crops were hard to separate as temporal signatures were very similar. Firstly, because more crops were classified for the area of Sougoumba, smaller distances and more overlap between the classes was created in feature space. Secondly, the sowing dates of the crops were variable (Appendix 25). Thirdly, due to within-field irregularities the class range (standard deviations) was relatively large compared to the single image class range (Appendix 14). Figure 31 and Figure 32 show that indeed within one field, a crop can have a totally different temporal profile**.** Nonetheless, the multi-temporal classification was more accurate than the single image classification.

Moreover, the differences between the classifications based on the STARS training datasets and the de Schaetzen (2015) training dataset decreased. Appendix 8C-F shows that the mean temporal profiles of both datasets in 2015 were almost similar. In addition, temporal information was decreasing the influence of within-field variability (Appendix 14D-F). Thus, the spectral variability within a specific class can be partly overcome by adding temporal information. When crops grow under less optimal conditions the statement of Conrad et al. (2010), Van Niel & McVicar (2004) and Wardlow et al. (2007) is confirmed in this work; multi-temporal crop classification is more suitable than a classification using a single image.

From the two tested methods, applied at the Sougoumba imagery, the K-NN clearly outperformed the regression tree method. Most likely this was caused by deviations in the fields. Because the K-NN method does the assignment based on the *k* nearest neighbors in feature space, a few aberrant cells in a training dataset will have a relatively low impact on the classification. However, with the regression tree method used in this research, cells are classified solely on their spectral variables. Each splitting node in the tree only uses a single variable per splitting rule. The splitting rule is derived from all cells in the training dataset and applied on all individual cells that need to be classified (Therneau & Atkinson, 1997). Therefore, aberrant cells are more likely to be misclassified. Using multiple neighbors in feature space to assign a class to a cell was most suitable for this research.

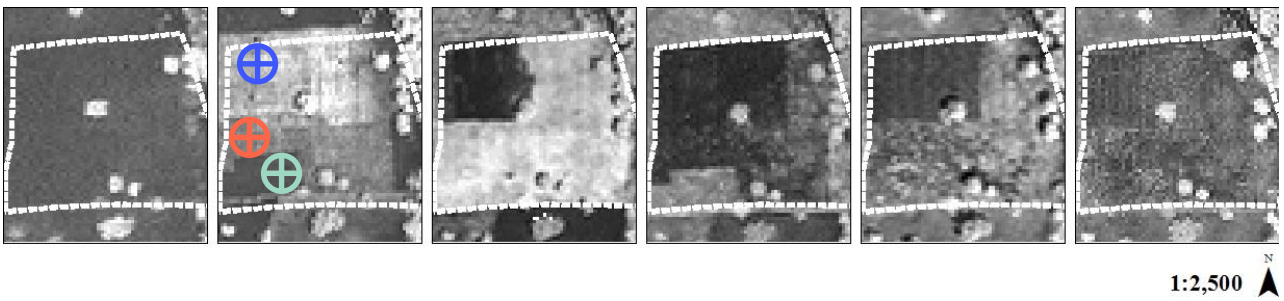


Figure 31. Temporal within-field (Peanut) variability in Sougoumba, visualised with the PVI. The temporal profile of the three markers is visualised in Figure 32. Based on the satellite images of 2015, from left to right; 03-06-2015, 09-07-2015, 03-09-2015, 25-09-2015, 19-10-2015 and 05-11-2015.

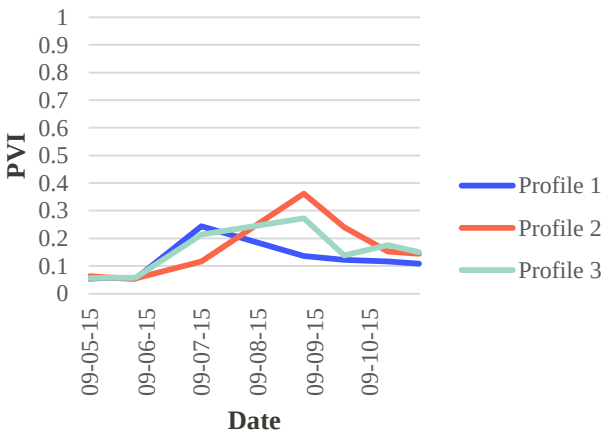


Figure 32. Temporal PVI profiles of the markers placed in Figure 31.

4.4 Sougoumba with stratification

4.4.1 Mono-temporal distinguishability and classification

Stratification increased the accuracies of the single image classification. When the study area of Sougoumba was split up based on the soil strata the highest classification accuracy was achieved. Firstly, this was caused by the fact that these strata were also based on the elevation data. Therefore the soil strata were actually combining two landscape characteristics, which were causing variability, into a single set of strata. Secondly, the village of Sougoumba was located in the lower valley. Distance to this village was thus also slightly represented in the soil strata, as with these strata also the most fertile soil was located close to the village. Although it might be more cumbersome for other study areas, combining multiple landscape characteristics into one strata dataset is a plausible way to increase the classification accuracy.

That the stratification of soil was successful is underpinned by the fact that each stratum had different accuracies. These differences in accuracies were caused by different inter-variability and within-variability of the classes in each stratum (Figure 23 & Appendix 15A-C). The variability was lowest in the most fertile stratum with the

valley soils. The stratum with the plateau alluvium was harder to classify; but, for the STARS training datasets, the crop classes were still relatively compact compared to the third stratum. The third stratum contained all fields located between the other two strata. In the feature space plots it is visible that for this stratum the crop classes contained a major part of the NIR and red wavelength reflectances and were thus highly variable. In addition, the classes showed major overlap, which is also visible in the mean spectral profiles that followed a similar profile for each crop in this stratum. The overall accuracies per stratum confirm these findings (Figure 24).

Differences in accuracy between classification results of the three training datasets were decreased, compared to the unstratified mono-temporal classification of Sougoumba. In some cases the class compactness in the training dataset of de Schaetzen (2015) was even better than the class compactness in the STARS dataset. This implies that, to a certain extent, stratification of soil was able to diminish the effect of landscape characteristics causing crop variability. Nonetheless, although the accuracy was increased, spectral similarity and within-field variation was still a problem with stratification. Where spectral information of a single image was sufficient for Samanko, for Sougoumba more information was needed. The reason for this is that even the unstratified classification based on temporal information yielded higher accuracies, which is in accordance with the statement that a classification based on a single image is less suitable for a classification of heterogeneous areas (Conrad et al., 2010; Van Niel & McVicar, 2004). Thus, to achieve a larger classification precision a stratified multi-temporal classification is more promising.

4.4.2 Multi-temporal distinguishability and classification

The multi-temporal classification with a stratification of soil yielded the highest accuracies (Figure 33). One difference with the stratified mono-temporal classification was that, for the dataset of de Schaetzen (2015), a stratification of elevation was more optimal. This also underpins the importance of soil for crop performance. In this study site, soil composition changed with elevation. It can be assumed that crop variability is caused by this soil composition and not by the difference in elevation. The suitability of the elevation was most likely caused due to misclassifications in the soil stratum with an intermediate elevation. This stratum was divided over both elevation strata and, therefore, it is possible that the effect of the low classification accuracy in this soil stratum was partly leveled out.

In addition, the differences in accuracy between the three different datasets were higher compared to the stratified classification based on a single image. This was most probably caused by the fact that unstratified mono-temporal classes were created in the peak of the growing season. All crops were thus flowering, decreasing the within-variability of the crop-classes. Variability in the multi-temporal classes might be caused by the fact that, during a longer time interval, there are more processes which are influencing the measured crop reflectances. According to Guerschman et al. (2003), other processes causing variability might be: cultural practices (e.g. sowing, fertilization, replanting, weeding or harvesting) or pests. Appendix 25 shows that indeed crops were sown on different dates in both years. Tittonell et al. (2008) investigated the influence of soil fertility and crop management on yields in dryland areas. They found that soil fertility variables were subservient to crop management variables. Conrad et al. (2010) state that within field heterogeneity is decreased in rain fed agricultural areas, which implies that irrigation practices can influence the classification accuracy. Moreover, different farming practices are also affecting the accuracy of classifications of larger fields where crops grow under more optimal environmental conditions (Wardlow et al., 2007). Capturing these crop management variables in different strata might, however, not be possible, as they are dependent on the farmers

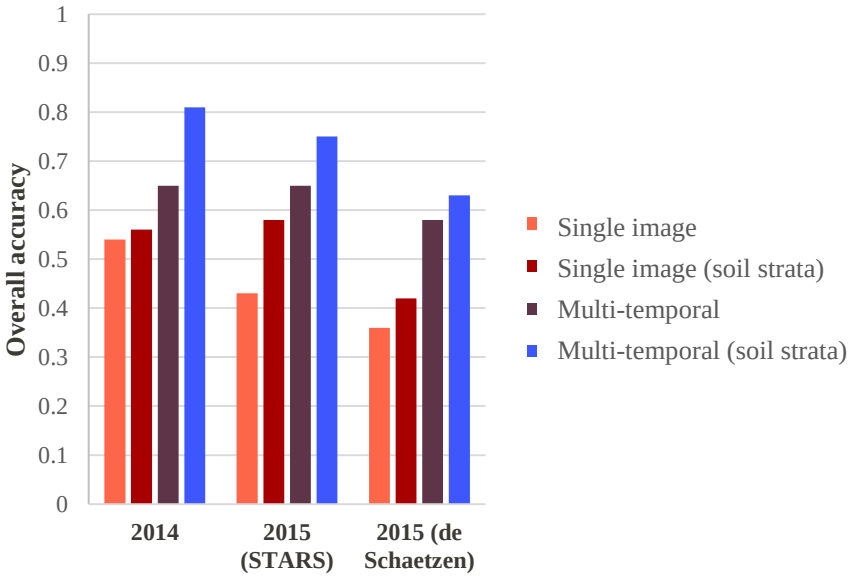


Figure 33 Classification improvements of Sougoumba, using the K-NN classifier (and PVI dataset for the temporal classification).

themselves and not necessarily spatially correlated. However, if change detection techniques are applied, they might be able to filter out some of the noise caused by crop management. If in a field the vegetation index value abruptly changes this probably means that a crop is removed, weeded or replanted (Figure 31). This development will most likely result in a misclassification and therefore pixels that deviate from the overall temporal crop profile should be removed from the analysis. Future research should be conducted to research the effect of the removal of these aberrant pixels.

The accuracy of the stratified multi-temporal classification was significantly higher than the stratified mono-temporal classification. However, for multi-temporal classifications stratification of soil or elevation was less successful in decreasing the differences between both classifications of 2015. When a larger training dataset is used it is more likely that fields subject to the previously mentioned noise are included, increasing the temporal variability of the classes. What was also striking was that the differences between the classifications based on both STARS training datasets increased**.** The reason for this is the absence of peanuts in the soil stratum with a low elevation for 2014. With the K-NN method this stratum had an overall accuracy of 91 percent in 2014. However, when the STARS fields were used in the classification of 2015 the overall accuracy was just 79 percent for this stratum (Figure 26). Indeed the class overlap seemed to be lower in the feature space plots of 2014. Having sufficient training data in all strata is thus important. Therefore, in the case of a continuation of this study more fields should be included in the STARS datasets. In addition, for the classification of 2014 more dates were used. Without stratification this, however, did not seem to influence the accuracy. This phenomenon will be further explained in paragraph 4.6.

4.5 Per-field aggregation

The multi-temporal K-NN classification in combination with the PVI and a stratification of soil was, in almost all cases, the most suitable method for a classification in Sougoumba. When the classified rasters of this method were used in the per-field majority filter, the accuracies increased even more. In 2014 the improvement of this aggregation was the lowest, the reason for this was, firstly, the relatively low producer's accuracy for peanuts. Two of the eight fields were misclassified, which was caused by misplaced field borders in the training dataset and within-field variability (Figure 34). Borders most likely contained some errors, as at the start of the season differences within fields were already clearly visible. Secondly, within-field variability was causing the other misclassifications. This within-field variability and border misplacement is most probably caused by errors in the creation of the field maps or crop management. It is possible that another crop was present in the field or that part of the field was sowed, harvested or replanted at a different time. In addition, because the weeding in the fields happens manually it is possible that fields were not completely weeded within a single day. A partly weeded field increases the variability, which will affect the classification accuracy. When looking at differences between temporal profiles of three pixels in a single field it becomes clear that it was almost impossible to classify some fields correctly (Figure 32). Misclassified fields in the classifications of 2015 showed similar characteristics (Appendix 17).

As mentioned the classification based on the STARS dataset was superior compared to the classification based on the de Schaetzen dataset for 2015. Therefore, the classified raster created with training data of this dataset was also tested in the per-field aggregation with the dataset of de Schaetzen (2015). However, this resulted in a relatively low accuracy, which confirms the statement that it is important to capture the variability of each crop in the training data. However, it must be said that the per-field aggregations of the STARS datasets with the STARS training data were biased. Firstly, because the fields in the STARS dataset did not accurately represent the crops in the region of Sougoumba in each stratum. Not all crops were present in the second soil stratum and the number of fields per stratum was relatively low. Secondly, farmers involved in the STARS project might have been chosen based on their farming practices or they might have adapted their crop management. Thirdly, the training data of the pixel-based classification were also located in all fields used in the per-field aggregation. In future research this bias can be decreased by choosing sufficient randomly distributed training samples in each stratum without knowing the field boundaries. The distribution of the training data is essential, as a training dataset which does not capture all the variability will result in a low classification accuracy.

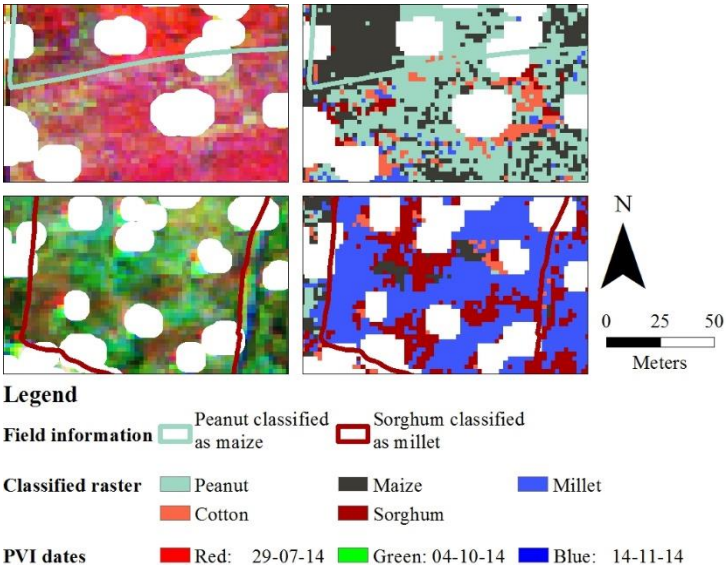


Figure 34. Within-field variability causing misclassification for Sougoumba 2014. The left images contain a RGB image of the PVI index, each colour represents a different date. The right images contain the resulting classified raster (using K-NN with the PVI).

4.6 Spatial and temporal resolutions

4.6.1 Optimal temporal resolution

Diminishing the variability was expected to be easier with an extensive temporal dataset. However, temporal datasets can also contain redundant information. Removing vegetation index images early in the season, where all fields only represented bare soil was not decreasing the accuracy significantly. The remaining images explained a major part of the differences between the crop types. All vegetation index images located in the phenological phases where crops started to deviate could not be removed without significantly affecting the classification results. This means that for crop classifications in Sougoumba it is necessary to have images from the green up phase until the end of the growing season.

It is, however, not possible to have an accurate classification with three images without knowing if crops are separable on these images. Images selected were assumed to show a relative large distinguishability between the classes. Nonetheless, the classification results were showing relatively low overall accuracies. Therefore, it is important to first check the class overlap and variability before proceeding with the classification. Van Niel & McVicar (2004) successfully classified their study area with only three images, but they only used the images which were relatively accurate in a single image classification. Therefore, they still needed an extensive temporal dataset. Five images in the growing, peak of the season and senescence period of the crops should, however, be sufficient for classification. If, in future research, saving costs is of importance, it is still possible to first classify the area with only three images, which are expected to contain separable classes. When the accuracy is not sufficient, the temporal resolution can be increased with images acquired in the previously mentioned crop phases.

4.6.2 Optimal spatial resolution

Increasing the spatial resolution improved the classification in 2015. Aggregation decreased the within-class variability when the training data of de Schaetzen (2015) was used (Appendix 15D-F & Appendix 22). For the other classification of 2015 the standard deviations were only slightly decreased by the aggregation. On the contrary, the within-class variability increased for 2014. This difference in compactness was also visible in the accuracies; the aggregation was especially useful for the classification based on the de Schaetzen (2015) training data, because there were more outliers (Appendix 11). In the classification without aggregation the soil strata with a high and intermediate elevation contained the most variable classes. Therefore, the result of the aggregation is clearly visible when feature space plots of these strata are compared (Figure 35). The feature space plots indeed show that the variation in the classes decreased (keep in mind that the axis are scaled on the minimal and maximum PVI values). The reason for this is that in case of a factor 4 aggregation, four cells are combined to a single cell and the new cell gets the average value. As a result, the influence of outliers decreases. Classifications with compacter classes containing less extreme observations are, therefore, benefitting less from an aggregation. This was also visible in the per-field majority filter, where only the classification based on the dataset of de Schaetzen (2015) had a higher accuracy after the spatial aggregation.

It needs to be noted that trees and bare soil pixels were removed before aggregation. Therefore, these findings are not transferable to coarser resolution imagery. In this study, an aggregation was increasing the accuracy, but high resolution imagery was needed for the classification to be successful. An extension of this study might point out if this 'need' is indeed valid, or if spatially aggregated images are also suitable for bare soil and tree removal.

4.6.3 Optimal combination

Combining the five best images with the spatial aggregation with a factor 4, resulted in slightly lower accuracies than the classification with only a spatial aggregation. Thus, the reduction in accuracy is mainly caused by the loss of temporal information. Still, the classification result was almost similar for 2014 compared to the classification without aggregation and with all temporal information. Moreover, for both classifications in 2015 the overall accuracy did improve slightly. This finding underpins the statement that a higher temporal resolution is not necessarily more suitable for classification. Only the informative bands, in which crops are distinguishing themselves, are useful. Removing redundant bands speeds up the calculation and has a relative small impact on the classification accuracy.

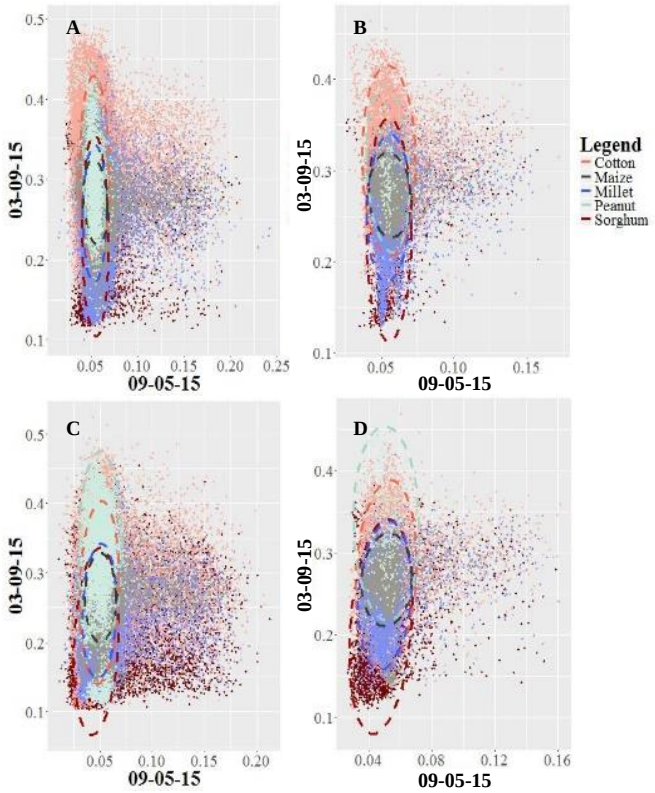


Figure 35. Feature space plots of the early-middle season PVI classes with the 95 percent confidence ellipse of 2015 (using fields of de Schaetzen (2015)). (A) Plateau alluvium stratum, without aggregation. (B) Plateau alluvium stratum, aggregated with factor 4. (C) Intermediate soils stratum, without aggregation. (D) Intermediate soils stratum, aggregated with factor 4.

5. CONCLUSION

The main goal of this study was to create an accurate method for the classification of crops on smallholder farms in Mali (cotton, maize, sorghum, peanuts and millet). For the classification, three training datasets were used: two datasets were derived from the field boundaries of farmers participating in the STARS project and a more extensive one was created by de Schaetzen (2015). In all classifications the more extensive dataset of de Schaetzen (2015) was less accurate. This was because this dataset contained a larger number of fields and, therefore, the class variability was increased. However, it must be noticed that the classifications with this dataset were closer to reality as fields were distributed over the complete study area.

In the process of coming to an accurate classification, first the mono and multi-temporal distinguishability of crops was measured. This measurement was done at two different study sites. In the study site of Samanko, crops were distinguishable in the NIR wavelengths and the phenological senescence phase. The per-field aggregation of this area proved that for well-maintained crops a single spectral image is sufficient to classify all crops correctly. However, in the more heterogeneous area of Sougoumba the crop classes were harder to separate. This was caused by class overlap and within-class variability due to variable crop management, errors in field boundaries, landscape characteristics and within-field variability. For Sougoumba, mean spectral profiles derived from a single image were completely overlapping each other. However, crop classes showed some differences in the multi-temporal profiles, which was in agreement with the expectations.

Accuracies of the classifications were displaying the importance of class separability. For Samanko the optimal classifications resulted in overall accuracies above 80 percent. Evenmore, with a per-field aggregation all fields in this study site were classified correctly. The classifications of Sougoumba yielded lower overall accuracies. In this study site the multi-temporal classification was clearly superior. A single image pixel-based K-NN classification resulted in a maximum overall accuracy of 54 and 43 percent with the smaller training datasets of the STARS farmers of respectively 2014 and 2015. When the dataset of de Schaetzen (2015) was used for training, the overall accuracy was only 36 percent. The multi-temporal classification with the most suitable vegetation index (PVI) improved the accuracies with on average 10 percent. The reason for this is that crops are differentiating themselves from each other in different phenological phases, while in a mono-temporal classification crop classes can have the same spectral ranges. Still, classes were overlapping and contained outliers. Therefore, the K-NN classification, which is less sensitive to these class characteristics, was the most optimal classifier.

Although the accuracies improved with a multi-temporal classification, within-field variabilities were still causing misclassification. Part of this variability was caused by landscape characteristics and to diminish this influence stratification was needed. Stratification of soil was most optimal, as these strata were combining multiple landscape characteristics. Especially, the class overlap decreased as a result of the splitting up of the study area. Nonetheless, a stratified classification with a single image was still less accurate than an unstratified multi-temporal classification. A single image classification is thus unsuitable for a classification of heterogeneous areas. Stratification of multi-temporal datasets yielded the best results as the overall accuracy improved with an average of 10 percent compared to the unstratified classification; therefore, stratification is highly useful in the classification of smallholder crops.

An aggregation of the classification results to field level was successful in decreasing the effect of outliers in the classifications. In addition, aggregation of the spatial resolution diminished this influence especially for the multi-temporal classification based on the training dataset of de Schaetzen (2015). On the contrary, for the classifications with the STARS fields the accuracy decreased when the classifications with the spatially aggregated rasters were used in the per-field aggregation. Spatial and per-field aggregation is only suitable if there are many extreme observations within the classes caused by a high within-field variability. Still, the original high resolution imagery was needed for tree- and bare soil removal. Furthermore, decreasing the temporal resolution to five images from different dates was not affecting the classification accuracies. For classification, it is only needed to have imagery at the dates at which crops start to deviate from each other. When both optimal resolutions were combined the accuracy slightly decreased, which was caused by the loss of temporal information. If the highest possible accuracy is required it is, therefore, advisable to keep a high temporal resolution. Nonetheless, the classification result of the combined aggregation was higher than the classification result without any aggregations. To come to an accurate classification, with a smaller (cheaper) dataset, this approach is highly suitable.

With a maximum overall accuracy of above 80 percent for the classifications with the STARS fields and an accuracy of 68 percent with the de Schaetzen (2015) dataset, the classifications in this research were successful and reasonably accurate. Some small improvements in the training datasets would increase the accuracy even more. Within-field variation, however, remains a challenge. As the classification accuracy is decreasing with an increasing class variability. Still, crop monitoring is possible in the region of Sougoumba. Moreover, the approach followed by this research is flexible and can be applied to other study areas. For STARS this means that the creation of a transparent information system supporting smallholder farmers in Mali is one step closer.

6. RECOMMENDATION

Several recommendations for possible future research have already been done in the discussion. In this chapter these and some more recommendations will be summed up and explained more in depth.

6.1 UAV data

- There were quite some issues with the Tetracam. Therefore, a manual was written and presented to ICRISAT. When errors occur this manual should be consulted.
- The calibration of the UAV imagery could be improved by using a newly created 80 percent reflectance panel. In addition, the reflectance values were only checked for 3 dates, because with the ILS it was possible to transform the DN-values to reflectances. However, to be safe, it is recommended to put the reflectance panels alongside the fields with each flight.
- Alignment of the imagery with a 2 centimeter resolution was not possible. As classifications with the imagery with a 5 and 10 centimeter resolution were accurate, it is not necessary to acquire imagery with a 2 centimeter resolution in further research.
- The alignment of the images was done with PixelWrench 2. However, the standard program used at the ICRISAT research center was Agisoft Photoscan. It is possible to do the orthorectification with Photoscan, but the results were less accurate (software failed to align some images and results were different when images were aligned multiple times using the same method) and the process is more time consuming. When Tetracam imagery has to be aligned PixelWrench 2 is recommended.
- The temporal dataset of Samanko was incomplete. It is recommended to acquire imagery of the whole crop calendar in future research. Firstly, because than all crops can be included in the research. Secondly, because new differences between crops might be discovered in different phenological phases.
- It is, however, not necessary to acquire UAV images four times a week. With the classification of Sougoumba only 5 images acquired during the green up, peak of the season and senescence phase were sufficient. Therefore, a lower temporal resolution for Samanko would also be suitable.
- Although satellite imagery is still more suitable to map the area of Sougoumba, it would be useful to acquire UAV images of some fields in Sougoumba. When this is done, these images can be classified and it can be tested if higher spatial resolutions (than the satellite imagery resolutions) are more suitable for classification.

6.2 Satellite data

- Satellite data of Samanko was only used to validate the UAV imagery. In an extension of this research it is not necessary to include this step, as the images can also be validated with reflectance panels. Moreover, a classification of Samanko will not be possible with satellite images, as fields are too small to have sufficient training data.
- The satellite images of Samanko were not all aligned to each other. This was not a major problem, as Erdas AutoSync was quite easy to understand. However, it should be noted that a check of the imagery should always be performed, even if the imagery is delivered after atmospheric correction and orthorectification.
- This research has pointed out that it is possible to come to an accurate classification with 5 images spread over the growing season. It is, however, necessary that in these images crops are differentiating themselves from each other. Therefore, images at the start of the season are not suitable. It is recommended to acquire these images from the start of July until the end of October. The number of images can be increased when the accuracy is too low.
- In this study it was found that an aggregation of factor 4 (to a resolution of 8 meter) increased the classification accuracy. Therefore, images with this resolution are also suitable in a smallholder crop classification. It is recommended to use this imagery with a lower resolution in future research, as it will most likely be cheaper. However, two images with a high resolution are still necessary to filter out the bare soil and trees. These images should be acquired in the early season for tree detection (May) and peak of the season for bare soil removal (August). It might also be possible that trees and bare soil can be detected successfully in images with a lower resolution. However, this should be tested in future research.

6.3 Training data

- The training data created with the STARS datasets was incomplete and the training data created with the de Schaetzen (2015) dataset was not completely correct. Training data could be improved in three ways. Firstly, it is important to define the strata beforehand, so that it is possible to ensure that training data of each crop is

present in each stratum. Secondly, the fields used to create training data should be spread randomly over the area. Not only fields of farmers involved in the STARS project should be used, to avoid a biased classification. Thirdly, it should be checked at the end of the season if farmers re-planted part of the field with another crop.

- In addition, it is useful to create a more extensive dataset about the farmer management practices. For the STARS fields this dataset is already present, but this should also be done for the extended training dataset. Information on sowing, harvesting and weeding could, namely, explain misclassifications.
- Training data were created with a random point algorithm. Bare soil and trees were removed, but it was still possible that pixels representing weed or aberrant crops were selected. Creating training data by doing field-work would resolve this issue, although this will be time consuming.
- To avoid training pixels being located on trees, trees were removed from the images with a buffer of 5 meter. However, the influence of this buffer size was not tested. It might be possible that a different buffer size might be more optimal. Therefore, it is recommended that in further studies different buffer sizes are tested.

6.4 Classification

- When crop classes are not following a normal distribution non-parametric classifiers are more suitable. However, this research has not tested parametric classifiers on normalized classes. An interesting question following on this research would be if parametric classifiers will also be suitable after this normalization. The easiest way to come to an accurate classification is, however, to use a non-parametric classifier like the K-NN.
- Although, the K-NN classifier was most suitable for the region of Sougoumba, it must me noted that for other study areas or with different imagery other classifiers might be more optimal. Therefore, it is recommended to always test multiple classifiers on their performance.
- In addition, multiple vegetation indices should be tested. In this study the PVI was most suitable, however the differences with other indices was relatively small.
- Van Niel & McVicar (2004) first classified crops that were relatively easy to classify and manually removed the pixels of these crops from further classification. In their research this method was increasing the accuracy. Therefore, it is highly interesting if this method is also suitable for a classification in Sougoumba.
- A per-field aggregation is recommended if there are a large amount of outliers in the classes. In addition, the influence of misplaced crops in the fields is also diminished, if the major part of the field is classified as the correct crop.
- Removal of pixels with a sudden change in the temporal signature might possibly increase the classification accuracy. Therefore, it is interesting if removal with change detection techniques is increasing the accuracy in further classifications.

6.5 Stratification

- Stratification is a useful tool to increase the classification accuracy in heterogeneous landscapes. In this study, a stratification of soil was most optimal, partly because soil strata were overlapping with other landscape characteristics. However, since the soil dataset was 33 years old, an update or validation of this dataset is recommended if it is going to be used in future studies.
- Moreover, management practices are also an important cause for variability. Capturing these practices in a new strata might also be useful. Furthermore, other unknown factors might be important. Therefore, in future research it is recommended to measure the suitability of other stratifications on the classification accuracy.
- In this research, the soil strata partly included other landscape characteristics (distance to buildup and elevation). Because the stratification of soil was most successful, it is recommended that multiple landscape characteristics that cause variation are included in the stratification.

ACKNOWLEDGEMENTS

This study was supported by ICRISAT (Mali) and Wageningen University. It was carried out within the framework of the STARS project, funded by the Bill & Melinda Gates Foundation. I would like to thank these instances for the opportunity given to me, as I learned a lot on top of the MGI master program followed at the Wageningen University.

I would like to thank my supervisors Dr.ir.ing. A.G.T. Schut and Dr.ir. J.G.P.W. Clevers for their advice during my thesis work. Their insight and expertise greatly improved this research. Without them I would most likely still be focused on improving instead of explaining the classification accuracy. For my stay in Mali, I would like to thank Dr.ir P.C.S. Traoré and my other colleagues at ICRISAT. I gained a lot of knowledge on the octocopter flight system and big data analysis. In addition, my stay in Mali was an unforgettable experience.

REFERENCES

Aci, M., İnan, C., & Avci, M. (2010). A hybrid classification method of k nearest neighbor, Bayesian methods and genetic algorithm. *Expert Systems with Applications, 37*(7), 5061-5067.

Barbier, E. B. (1997). The economic determinants of land degradation in developing countries. *Philosophical Transactions of the Royal Society B: Biological Sciences, 352*(1356), 891-899.

Baret, F., Guyot, G., & Major, D. (1989). *TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation.* Paper presented at the Geoscience and Remote Sensing Symposium, 1989. IGARSS'89. 12th Canadian Symposium on Remote Sensing., 1989 International.

Baret, F., Jacquemoud, S., & Hanocq, J. (1993). The soil line concept in remote sensing. *Remote Sensing Reviews, 7*(1), 65-82.

Bastiaanssen, W. G., & Ali, S. (2003). A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture, ecosystems & environment, 94*(3), 321-340.

Blaes, X., Traore , P. C. S., Schut, T., Ajeigbe, H. A., Chome, G., Boekelo, B., Diancoumba, M., & Goita, K. (2014). STARS-ISABELA 2014-2015: Field data collection protocol (v. 7). *Unpublished Protocol*.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., & Zeil, P. (2000). Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. *Environmental Information for Planning, Politics and the Public, 2*, 555-570.

Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment, 76*(2), 156-172.

Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological modelling, 168*(3), 233-249.

Castillejo-González, I. L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J. M., Jurado-Expósito, M., de la Orden, M. S., & González-Audicana, M. (2009). Object-and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture, 68*(2), 207-215.

CGIAR. (2014). Research initiative improves smallholder agriculture with remote sensing. Retrieved from http://drylandsystems.cgiar.org/content/research-initiative-improves-smallholder-agriculture-remote-sensing

Chuvieco, E., & Huete, A. (2009). *Fundamentals of satellite remote sensing*. Boca Raton: CRC Press Inc.

Clevers, J. (1988). The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sensing of Environment, 25*(1), 53-69.

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

Conrad, C., Fritsch, S., Zeidler, J., Rücker, G., & Dech, S. (2010). Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sensing, 2*(4), 1035-1056.

de Schaetzen, F. (2015). Field limits and classification Sougoumba 2015 [shapefile]. Scale Not Given. "Land cover classification of fields within clusters". November 2015. Using: ArcGIS [GIS software]. Version 10.3. Mali, ICRISAT, 2015. Internship report, Wageningen University & Research Centre.

Deininger, K. W., & Byerlee, D. (2011). *Rising global interest in farmland: can it yield sustainable and equitable benefits?* Washington: World Bank Publications.

Dioula, B. M., Deret, H., & Morel, J. (2013). Enhancing the role of smallholder farmers in achieving sustainable food and nutrition security. *Food and Agriculture Organization*.

Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment from remote sensing. *Photogrammetric Engineering & Remote Sensing, 69*(6), 665-674.

Foody, G. (1996). Fuzzy modelling of vegetation from remotely sensed imagery. *Ecological modelling, 85*(1), 3-12.

Franklin, S. E. (1989). Ancillary data input to satellite remote sensing of complex terrain phenomena. *Computers & Geosciences, 15*(5), 799-808.

Giller, K. E., Witter, E., Corbeels, M., & Tittonell, P. (2009). Conservation agriculture and smallholder farming in Africa: the heretics' view. *Field Crops Research, 114*(1), 23-34.

Gong, P., & Howarth, P. (1990). The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing, 56*(1), 67-73.

Guerschman, J., Paruelo, J., Bella, C. D., Giallorenzi, M., & Pacin, F. (2003). Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. *International Journal of Remote Sensing, 24*(17), 3381-3402.

Hansen, M. H., & Wendt, D. G. (2000). Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. *North Central Research Station. U.S. Department of Agriculture - Forest Service*.

Holmgren, J., & Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment, 90*(4), 415-423.

Hoppus, M. L., & Lister, A. J. (2001). *A statistically valid method for using FIA plots to guide spectral class rejection in producing stratification maps.* Paper presented at the Proceedings of the 3d annual forest inventory and analysis symposium.

Hubert-Moy, L., Cotonnec, A., Le Du, L., Chardin, A., & Perez, P. (2001). A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment, 75*(2), 174-187.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*(3), 295-309.

ICRISAT. (2015). International Crops Research Institute for the Semi-Arid Tropics. Retrieved from http://www.icrisat.org/

ITC Faculty. (2014). CALIBRATION REFERENCE PANELS: STARS. *University of Twente, 1.0*.

Iverson, L. R., & Cook, E. A. (2000). Urban forest cover of the Chicago region and its relation to household density and income. *Urban Ecosystems, 4*(2), 105-124.

Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture, 37*(1), 127-139.

Jayne, T. S., Yamano, T., Weber, M. T., Tschirley, D., Benfica, R., Chapoto, A., & Zulu, B. (2003). Smallholder income and land distribution in Africa: implications for poverty reduction strategies. *Food Policy, 28*(3), 253-275.

Jensen, J. R., Garcia-Quijano, M., Hadley, B., Im, J., Wang, Z., Nel, A. L., Teixeira, E., & Davis, B. A. (2006). Remote sensing agricultural crop type for sustainable development in South Africa. *Geocarto International, 21*(2), 5-18.

Jönsson, P., & Wohlin, C. (2004). *An evaluation of k-nearest neighbour imputation using likert data.* Paper presented at the Software Metrics, 2004. Proceedings. 10th International Symposium on pn Software Metrics, Chiaco, USA.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). *Remote sensing and image interpretation*. Seventh edition. Hoboken: John Wiley & Sons.

Lobo, A. (1997). Image segmentation and discriminant analysis for the identification of land cover units in ecology. *IEEE Transactions on Geoscience and Remote Sensing, 35*(5), 1136-1145.

Lobo, A., Chic, O., & Casterad, A. (1996). Classification of Mediterranean crops with multisensor data: per-pixel versus per-object statistics and image segmentation. *International Journal of Remote Sensing, 17*(12), 2385-2400.

Mas, J. (2004). Mapping land use/cover in a tropical coastal area using satellite sensor data, GIS and artificial neural networks. *Estuarine, Coastal and Shelf Science, 59*(2), 219-230.

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. *Remote Sensing, 7*(3), 2971-2990.

Mather, P., & Tso, B. (2009). *Classification methods for remotely sensed data*. Boca Raton: CRC press.

McRoberts, R. E., Holden, G. R., Nelson, M. D., Liknes, G. C., & Gormanson, D. D. (2005). Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest Service. *Canadian Journal of Forest Research, 35*(12), 2968-2980.

McRoberts, R. E., Wendt, D. G., Nelson, M. D., & Hansen, M. H. (2002). Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment, 81*(1), 36-44.

NASA. (2015). U.S. Releases Enhanced Shuttle Land Elevation Data. Retrieved from http://www2.jpl.nasa.gov/srtm/

Oleson, K., Sarlin, S., Garrison, J., Smith, S., Privette, J., & Emery, W. (1995). Unmixing multiple land-cover type reflectances from coarse spatial resolution satellite data. *Remote Sensing of Environment, 54*(2), 98-112.

Projet Inventaire des Ressources Terrestres (1983). Les ressources terrestres au Mali. *New York, Government of Mali/USAID/TAMS.*

Rashmi, S., Addamani, S., & Ravikiran, S. (2014). Spectral Angle Mapper Algorithm for Remote Sensing Image Classification. *International Journal of Innovative Science, Engineering & Technology, 50*(4), 201-205.

Ricchetti, E. (2000). Multispectral satellite image and ancillary data integration for geological classification. *Photogrammetric Engineering and Remote Sensing, 66*(4), 429-435.

Richardson, A. J., & Weigand, C. (1977). Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing, 43*(12).

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment, 55*(2), 95-107.

Samaké, O., Smaling, E., Kropff, M., Stomph, T., & Kodio, A. (2005). Effects of cultivation practices on spatial variation of soil fertility and millet yields in the Sahel of Mali. *Agriculture, Ecosystems & Environment, 109*(3), 335-345.

Sanchez, P. A., Shepherd, K. D., Soule, M. J., Place, F. M., Buresh, R. J., Izac, A.-M. N., Mokwunye, A. U., Kwesiga, F. R., Ndiritu, C. G., & Woomer, P. L. (1997). Soil fertility replenishment in Africa: an investment in natural resource capital. *Replenishing Soil Fertility in Africa*, 1-46.

Shahshahani, B. M., & Landgrebe, D. (1994). The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Transactions on Geoscience and Remote Sensing, 32*(5), 1087-1095.

Smith, G., & Fuller, R. (2001). An integrated approach to land cover classification: an example in the Island of Jersey. *International Journal of Remote Sensing, 22*(16), 3123-3142.

Sohn, Y., & Rebello, N. S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing, 68*(12), 1271-1282.

Song, Y., Huang, J., Zhou, D., Zha, H., & Giles, C. L. (2007). Iknn: Informative k-nearest neighbor pattern classification . *Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, ser. PKDD 2007. Berlin, Heidelberg: Springer , 248–264.

South, S., Qi, J., & Lusch, D. P. (2004). Optimal classification methods for mapping agricultural tillage practices. *Remote Sensing of Environment, 91*(1), 90-97.

STARS. (2015). Spurring a Transformation for Agriculture through Remote Sensing. Retrieved from http://www.stars-project.org/en/

Therneau, T. M., & Atkinson, E. J. (1997). An introduction to recursive partitioning using the RPART routines: Technical Report 61. Retrieved from http://www. mayo. edu/hsr/techrpt/61. pdf.

Tittonell, P., & Giller, K. E. (2013). When yield gaps are poverty traps: the paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research, 143*, 76-90.

Tittonell, P., Shepherd, K. D., Vanlauwe, B., & Giller, K. E. (2008). Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—An application of classification and regression tree analysis. *Agriculture, Ecosystems & Environment, 123*(1), 137-150.

Tittonell, P., Vanlauwe, B., Leffelaar, P., Shepherd, K. D., & Giller, K. E. (2005). Exploring diversity in soil fertility management of smallholder farms in western Kenya: II. Within-farm variability in resource allocation, nutrient flows and soil fertility status. *Agriculture, Ecosystems & Environment, 110*(3), 166-184.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127-150.

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance—a review. *Field Crops Research, 143*, 4-17.

Van Niel, T. G., & McVicar, T. R. (2004). Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. *Computers and Electronics in Agriculture, 45*(1), 91-108.

Varela, R. D., Rego, P. R., Iglesias, S. C., & Sobrino, C. M. (2008). Automatic habitat classification methods based on satellite images: A practical assessment in the NW Iberia coastal mountains. *Environmental Monitoring and Assessment, 144*(1-3), 229-250.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment, 108*(3), 290-310.

Watson, W. (2006). Automated Georeferencing Automated Georeferencing for Rapid Data Production. *Photogrammetric Engineering & Remote Sensing*, 337.

Woittiez, L., Katrien, D., & Ken, G. E. (2015). Adoptability of sustainable intensification technologies in dryland smallholder farming systems of West Africa; Research Report No. 64.

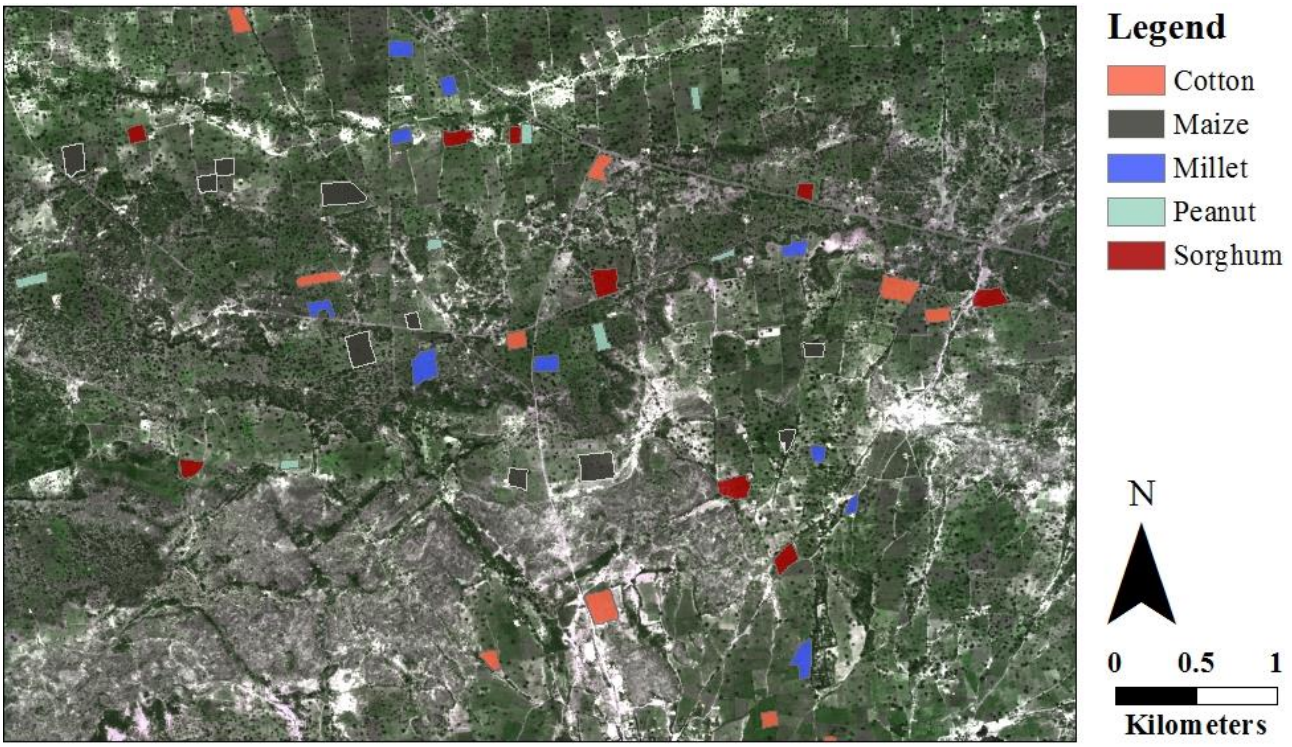
Xu, D., & Guo, X. (2013). A study of soil line simulation from landsat images in mixed grassland. *Remote Sensing, 5*(9), 4533-4550.

Zhong, Y., Zhang, L., Huang, B., & Li, P. (2006). An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing, 44*(2), 420-431.

APPENDIX 1: CROP MAPS FOR SAMANKO 2015 AND SOUGOUMBA 2014

(A) Crop map for Samanko 2015. RGB image: 12-11-2015.

(B) Crop map for Sougoumba 2014. Quickbird image: 04-10-2014.



APPENDIX 2: TETRACAM IMAGERY SPECIFICATIONS

Name	TETRACAM MINI-MCA6		
Sensor bands (nm)	545-555		
	675-685		
	705-715		
	735-745		
	795-805		
Resolution (cm)	2	5	10
Acquisition dates	14-10-15	15-10-15	15-10-15
	15-10-15	21-10-15	21-10-15
	17-10-15	23-10-15	23-10-15
	21-10-15	26-10-15	26-10-15
	24-10-15	29-10-15	29-10-15
	27-10-15	02-11-15	02-11-15
	30-10-15	05-11-15	05-11-15
	03-11-15	09-11-15	09-11-15
	10-11-15	13-11-15	13-11-15
		18-11-15	18-11-15

APPENDIX 3: TETRACAM CALIBRATION AND ALIGNMENT OF THE IMAGERY

(A) Tetracam calibration

Camera alignment
At the start of the data collection the Tetracam was properly calibrated. First the cameras were aligned using PixelWrench 2, a software package developed for the Tetracam. The images of all cameras are lined up to the master image. The master image is the image captured by the master channel, which is the channel used for the calculation of exposure requirements and the synchronizing of the other cameras (slaves). With the alignment differences in rotation, scaling, translation and vignetting were minimized for the different flying heights.

ILS sensor calibration
In addition to the camera alignment the ILS sensor was calibrated three times during the data collection. Normally this calibration should only be done once; however, as the original calibration was not satisfying the sensor was calibrated multiple times to ensure the data quality. PixelWrench 2 was used again for this purpose. First a reflectance panel with a reflectance of 45 percent was photographed with the Tetracam. This was done in clear sunlight and no shadows were cast on the panel or the camera. In PixelWrench 2 an area of the panel on the photos was selected and the conversion parameters were calculated by the program (down-welling sensor value and pixel scaling factor).

(B) Pix4d settings

Pix4d settings
- Processing Options Template: Agriculture
- Keypoints Image Scale: Full
- Image Scale: 1
- Multiscale: Checked
- Point Density: High
- Minimum Number of Matches:6

(C) Erdas AutoSync settings

IMAGINE AutoSync settings
- APM Strategy:
  Find points with: Defined pattern
  Intended Number of Points/Pattern: 8
  Maximum Blunder Removal Iterations: 20

- Geometric Model:
  Output Geometric Model Type: Polynomial
  Maximum Polynomial Order: 2 or 3
  RMS Threshold: 0.5

- Projection:
- Same as Input Image

- Output:
  Geocorrection: Resample
  Resample Settings: Nearest Neighbor; Cell Size same as Input Image

- Remove points with maximal error contribution until the Error Std. Dev is below 0.5 pixels

APPENDIX 4: FIELDWORK PROTOCOL

Charging batteries

The best place to charge the batteries is a cool and dry place. Temperatures of 5-10 degree Celsius are optimal.

1. First set up the charging station before plugging the batteries in. Attach the charger to the power supply and attach the battery charger plate to the charger.
2. Power up the station.
3. Check if the voltage of the power supply is set to 24 Volt (When the charger is attached the voltage will drop to 23 volt, which isn't a problem because the charger uses 1 volt).
4. Take the batteries with the same number written on the battery label.
5. Connect the batteries to the gold plugs and attach the balancer cable (order is not important).
6. Set the charger by pushing the big round button and selecting the Lido option.
7. Select charge in the following menu and click yes.
8. After about 25-30 minutes the batteries will be charged. You will hear an alarm and a message DONE will be shown on the screen of the charger. The charged batteries should have a voltage of about 4.2.
9. Remove the batteries directly after the alarm (do NOT forget them to avoid overcharging).
10. AFTER the removal of the batteries you can dissemble the station.

Preparing flight plan

Requirements

- Laptop with a windows installation
- USB Antenna, delivered with the octocopter
- Senselink USB dongle, delivered with the octocopter (small purple USB-stick)

1. Start GS-QGS
2. Draw a Polygon around the flight area
3. Set camera to Tetracam
4. Set desired flight altitude
5. Set overlap to 80 percent
6. Keep in mind the maximum flight length and flight distance

Prepare Tetracam

1. Format the FC cards before each flight
2. Calibrate the ILS if necessary
3. Align the camera if necessary

Prepare field

1. Bring the octocopter, batteries and laptop with the previously mentioned requirements
2. Check if the GCPs are clear
3. Mount fully charged batteries
    a. Take 4 batteries that have the same numbers written on the battery label.
    b. Place all batteries under the elastic clip on the octocopter, do NOT set them on the gold plugs yet.
    c. Place both batteries that are NOT having the balancer cable in the battery compartment on the connector plugs (So the battery circuit isn't closed and the octocopter isn't started yet, thus the propellers shouldn't be moving after this step).
    d. THIS STEP SHOULD BE COMPLETED WITHIN 5 SECONDS! Attach the balancer cable to one of the batteries which is not places on the battery, AFTERWARDS plug the battery on the connector plugs (now the propellers should be moving, because one circuit is closed).
    e. Check if all propellers are working.
    f. Attach the remaining battery to the balancer cable and place it on the gold plugs.
4. Check if the Tetracam is capturing images (flashing green/red led)

Flight protocol

1. Load the flight plan to the octocopter with GS-QGS
2. Check the octocopter and battery status in GS-QGS

3. Check if no one is near the octocopter
4. Check if there are no obstacles in the flight route
5. Power up the octocopter
6. Ascend to the desired flight altitude
7. Start the auto pilot
8. Start the Tetracam before the first waypoint
9. Stop the Tetracam after the last waypoint
10. Check if no one is in the landing zone
11. Land the octocopter
12. Remove both batteries with the balancer cables by first removing the cable and then unplugging the batteries. You can remove them by lifting the battery at the front end. After the removal of both batteries the octocopter is turned off.
13. Store the batteries behind the gold plugs.
14. Unplug the remaining two batteries and store them with the other two in the appropriate place in the octocopter box.
15. Turn of the remote control
16. Bring the gear back to the lab

Data transfer

1. Extract al CF cards from the Tetracam
2. Copy all files in the TTC folder to a new folder with an appropriate name (the date that the imagery is acquired D:\GEO-X8000\yyyymmdd\RAW)

Prepare Tetracam for next flight

1. Format all drives by connecting them individually to the pc.
2. Choose the file system FAT32
3. Choose an allocation unit size of 32 Kb (if that is not possible 16 Kb)
4. Do NOT change the name of the disk
5. Insert all disks back in the appropriate slot

APPENDIX 5: SATELLITE IMAGERY SPECIFICATIONS

(A) Samanko 2015

Name	WORLDVIEW-2 MULTI
Sensor bands (nm)	400 - 450 450 - 510 510 - 580 585 - 625 630 -690 705 - 745 770 - 895 860 - 1040
Resolution (m)	2
Acquisition dates	13-11-15 04-11-15

(B) Sougoumba 2014

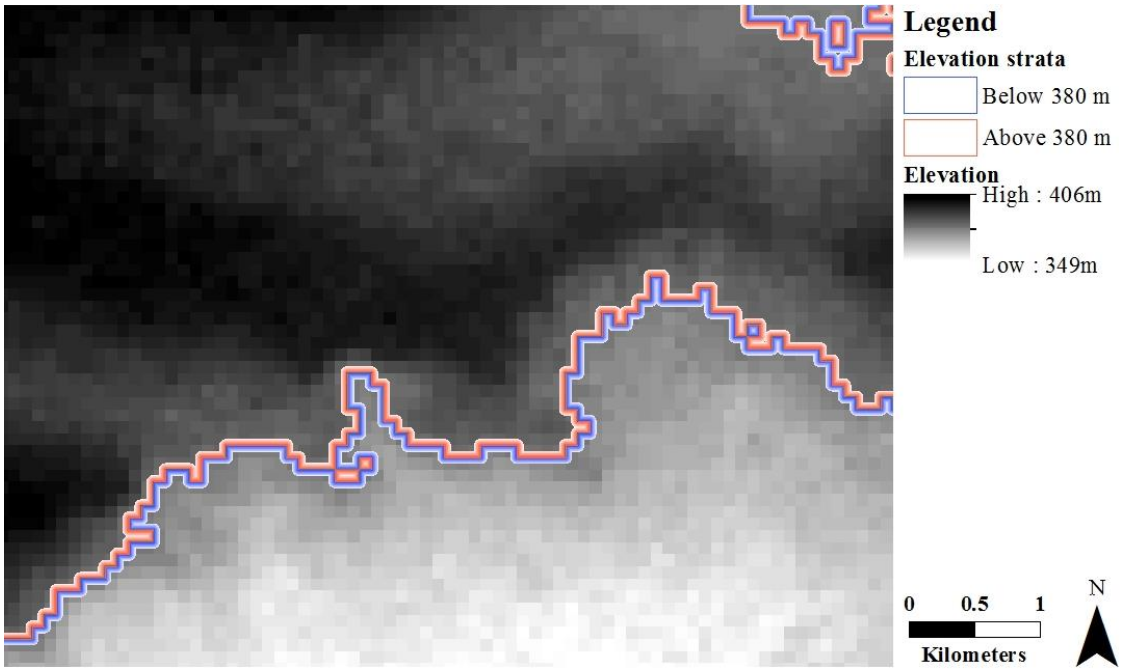
Name	WORLDVIEW-2 MULTI	WORLDVIEW-2 MS1	GEOEYE-1	QUIKBIRD
Sensor bands (nm)	400 - 450 450 - 510 510 - 580 585 - 625 630 -690 705 - 745 770 - 895 860 - 1040	450 - 510 510 - 580 630 -690 770 - 895	450 - 510 510 - 580 630 -690 770 - 895	430 - 545 466 - 620 590 - 710 715 - 918
Resolution (m)	2	2	2	2.4
Acquisition dates	22-05-14 30-05-14	26-06-14 29-07-14 18-10-14 01-11-14 14-11-14	01-05-14 08-07-14 25-09-14	26-08-14 04-10-14

(C) Sougoumba 2015

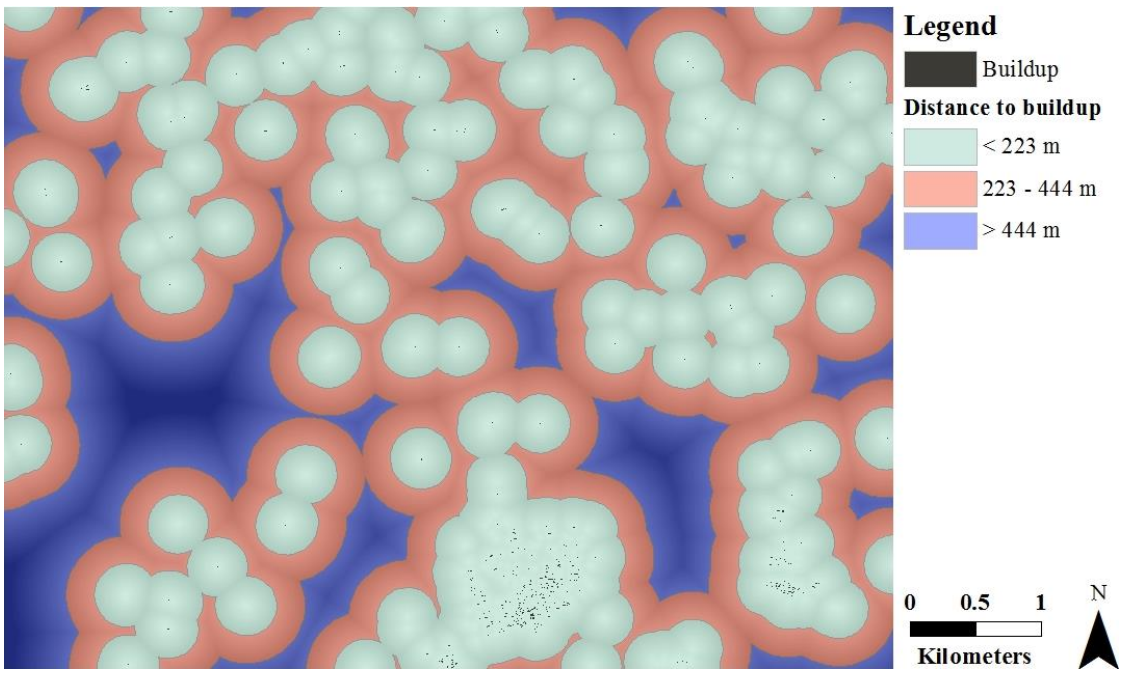
Name	WORLDVIEW-2 MULTI	WORLDVIEW-2 MS1	WORLDVIEW-3 MULTI	WORLDVIEW-3 MS1
Sensor bands (nm)	400 - 450 450 - 510 510 - 580 585 - 625 630 - 690 705 - 745 770 - 895 860 - 1040	450 - 510 510 - 580 630 - 690 770 - 895	400 - 450 450 - 510 510 - 580 585 - 625 630 - 690 705 - 745 770 - 895 860 - 1040	450 - 510 510 - 580 630 -690 770 - 895
Resolution (m)	2 (03-06-15: 1.6)	2	2	2
Acquisition dates	03-06-15 25-09-15 19-10-15	09-05-15 03-09-15	05-11-15	18-05-15 09-07-15

APPENDIX 6: ELEVATION AND BUILDUP STRATA SOUGOUMBA

(A) Elevation strata



(B) Buildup strata



APPENDIX 7: MONO- AND MULTI-TEMPORAL CLASSIFICATION RESULTS SAMANKO

(A) Single image classification results with different classification methods

Method	MAXL	MINDIST	SPECTANG	K-NN	REGT
Overall accuracy	0.85	0.72	0.72	0.83	0.82
Average user accuracy	0.86	0.74	0.74	0.84	0.83
Average prod. accuracy	0.85	0.72	0.72	0.83	0.82
Kappa	0.77	0.58	0.58	0.66	0.65

(B) Multi-temporal classification results with different classifiers using the NDVI

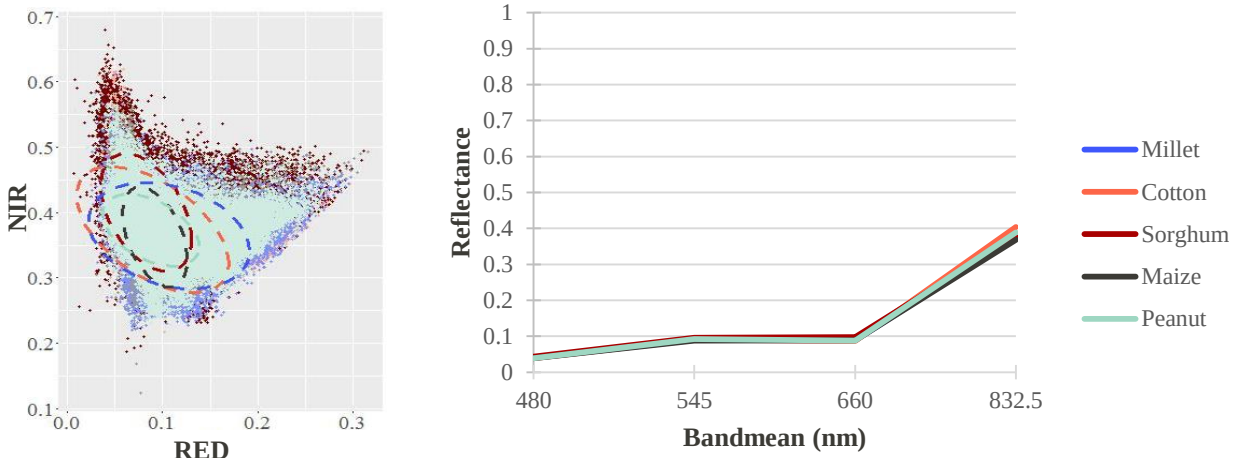
Method	MAXL	MINDIST	SPECTANG	K-NN	REGT
Overall accuracy	0.70	0.77	0.65	0.85	0.83
Average user accuracy	0.81	0.80	0.66	0.86	0.84
Average prod. accuracy	0.70	0.77	0.65	0.85	0.83
Kappa	0.55	0.66	0.48	0.70	0.65

(C) Multi-temporal classification results of different vegetation indices using the K-NN method

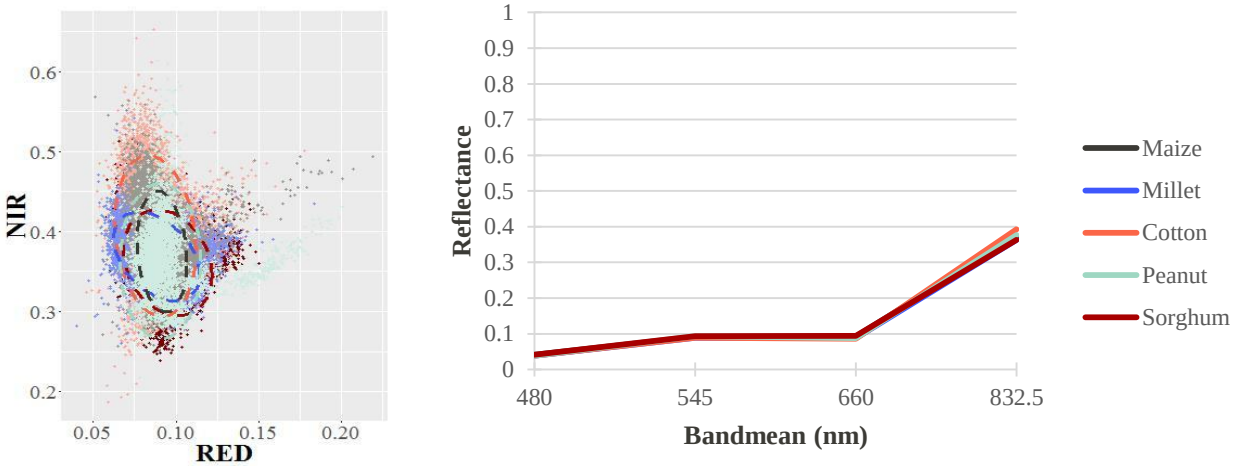
Index	NDVI	WDVI	SAVI	TSAVI	PVI
Overall accuracy	0.85	0.93	0.89	0.91	0.94
Average user accuracy	0.86	0.93	0.90	0.92	0.94
Average prod. accuracy	0.85	0.93	0.89	0.91	0.94
Kappa	0.70	0.86	0.78	0.82	0.87

APPENDIX 8: MONO AND MULTI-TEMPORAL SIGNATURES AND FEATURE SPACE PLOTS FOR SOUGOUMBA WITHOUT STRATIFICATION

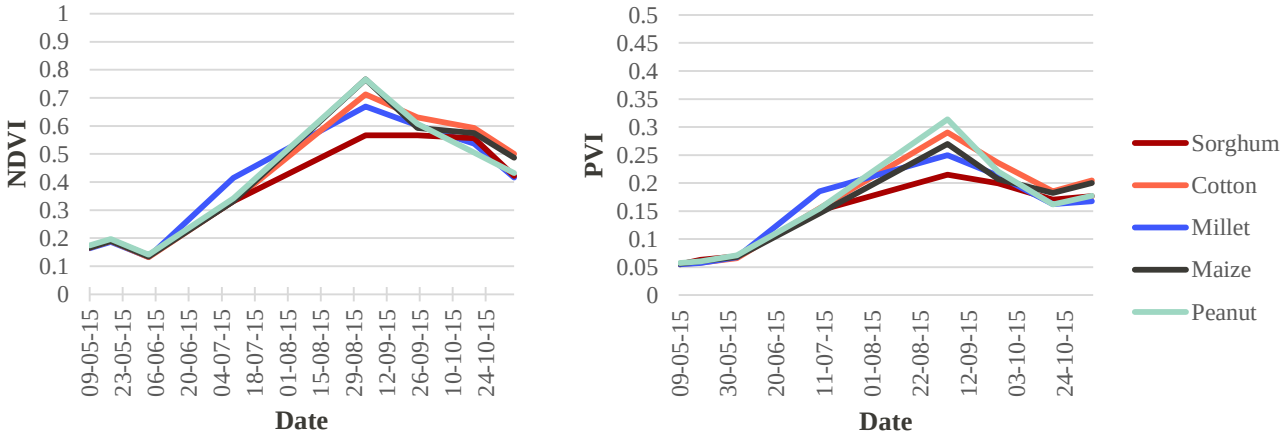
(A) Feature space plot of the NIR (832.5nm) and RED (660nm) band with the 95 percent confidence ellipse & mean spectral profiles. Based on WorldView-2 image: 25-09-2015. Training fields of de Schaetzen (2015) were used create the feature space plot and mean profiles.



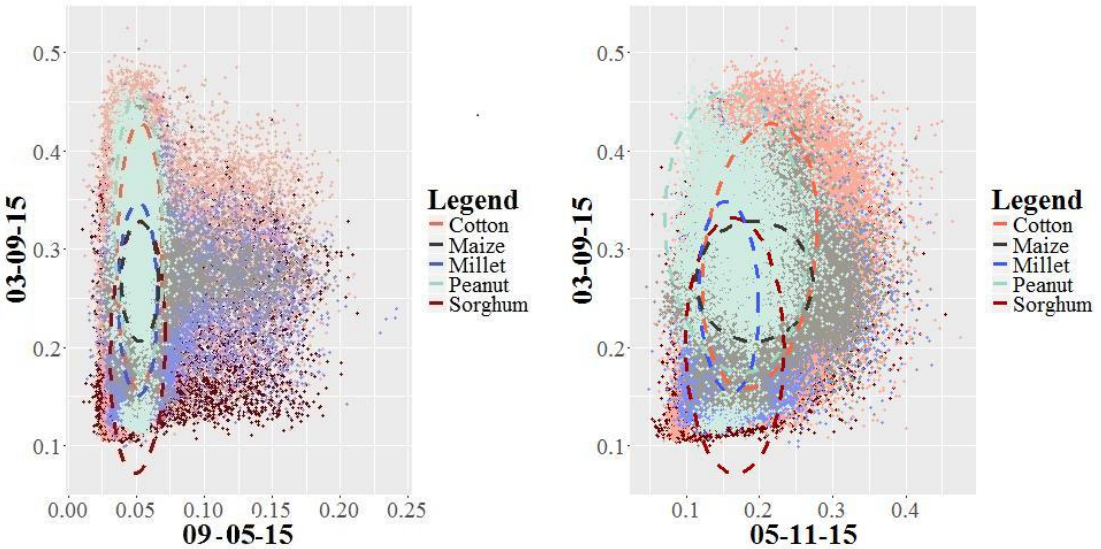
(B) Feature space plot of the NIR (832.5nm) and RED (660nm) band with the 95 percent confidence ellipse & mean spectral profiles. Based on WorldView-2 image: 25-09-2015. Training fields of STARS (2015) were used create the feature space plot and mean profiles.



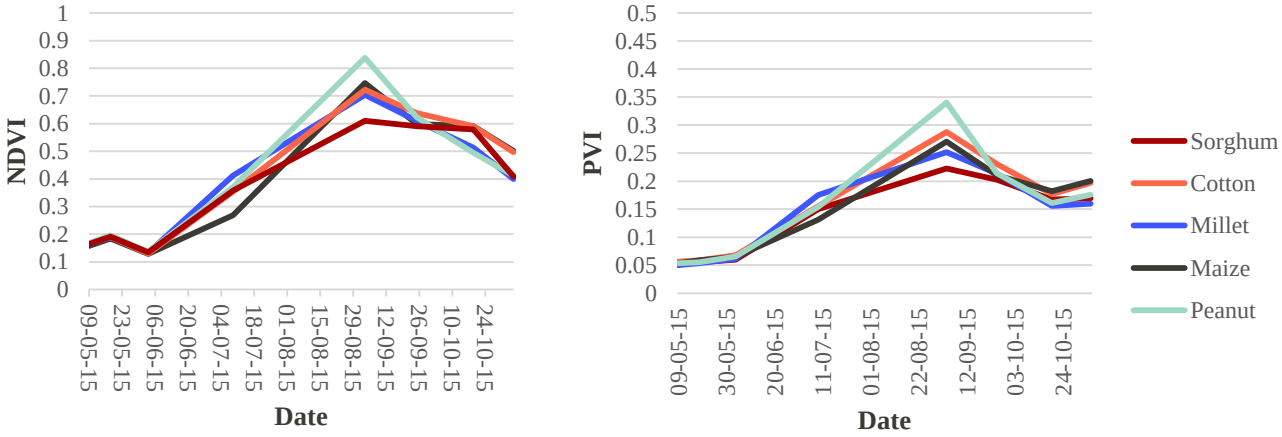
(C) Temporal mean NDVI and PVI profiles. Based on satellite images of Sougoumba 2015 (using fields of de Schaetzen (2015)).



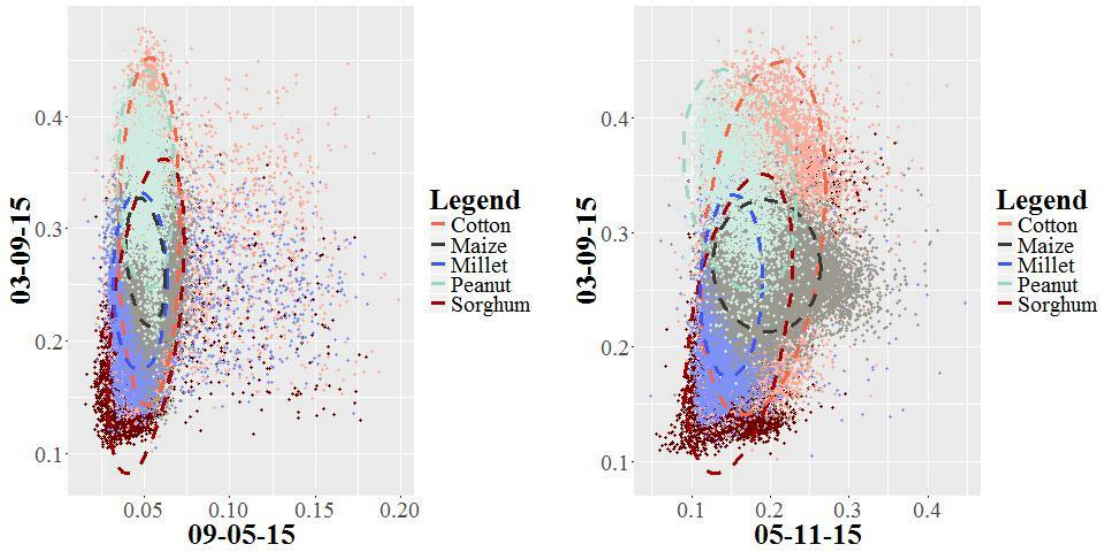
(D) Feature space plot of the early-middle and middle-end season PVI classes of 2015 with the 95 percent confidence ellipse (using fields of de Schaetzen (2015)).



(E) Temporal mean NDVI and PVI profiles. Based on satellite images of Sougoumba 2015 (using fields of farmers in the STARS project).



(F) Feature space plot of the early-middle and middle-end season PVI classes of 2015 with the 95 percent confidence ellipse (using fields of farmers in the STARS project).



APPENDIX 9: MONO- AND MULTI-TEMPORAL CLASSIFICATION RESULTS OF SOUGOUMBA WITHOUT STRATIFICATION

(A) Single image classification results of Sougoumba in 2014

	K-NN	REGT
Overall accuracy	0.54	0.52
Average user accuracy	0.56	0.54
Average prod. accuracy	0.54	0.52
Kappa	0.43	0.40

(B) Single image classification results of Sougoumba in 2015 (using fields of de Schaetzen (2015))

	K-NN	REGT
Overall accuracy	0.36	0.34
Average user accuracy	0.37	0.35
Average prod. accuracy	0.36	0.34
Kappa	0.20	0.18

(C) Single image classification results of Sougoumba in 2015 (using fields of farmers in the STARS project)

	K-NN	REGT
Overall accuracy	0.43	0.37
Average user accuracy	0.44	0.37
Average prod. accuracy	0.43	0.37
Kappa	0.29	0.22

(D) Multi-temporal classification results of Sougoumba in 2014

	K-NN PVI	REGT PVI	K-NN TSAVI	REGT TSAVI
Overall accuracy	0.65	0.55	0.61	0.55
Average user accuracy	0.68	0.57	0.68	0.57
Average prod. accuracy	0.65	0.55	0.61	0.55
Kappa	0.57	0.44	0.51	0.44

(E) Multi-temporal classification results of Sougoumba in 2015 (using fields of de Schaetzen (2015))

	K-NN PVI	REGT PVI	K-NN TSAVI	REGT TSAVI
Overall accuracy	0.58	0.51	0.56	0.51
Average user accuracy	0.61	0.54	0.61	0.54
Average prod. accuracy	0.58	0.52	0.56	0.52
Kappa	0.48	0.39	0.45	0.39

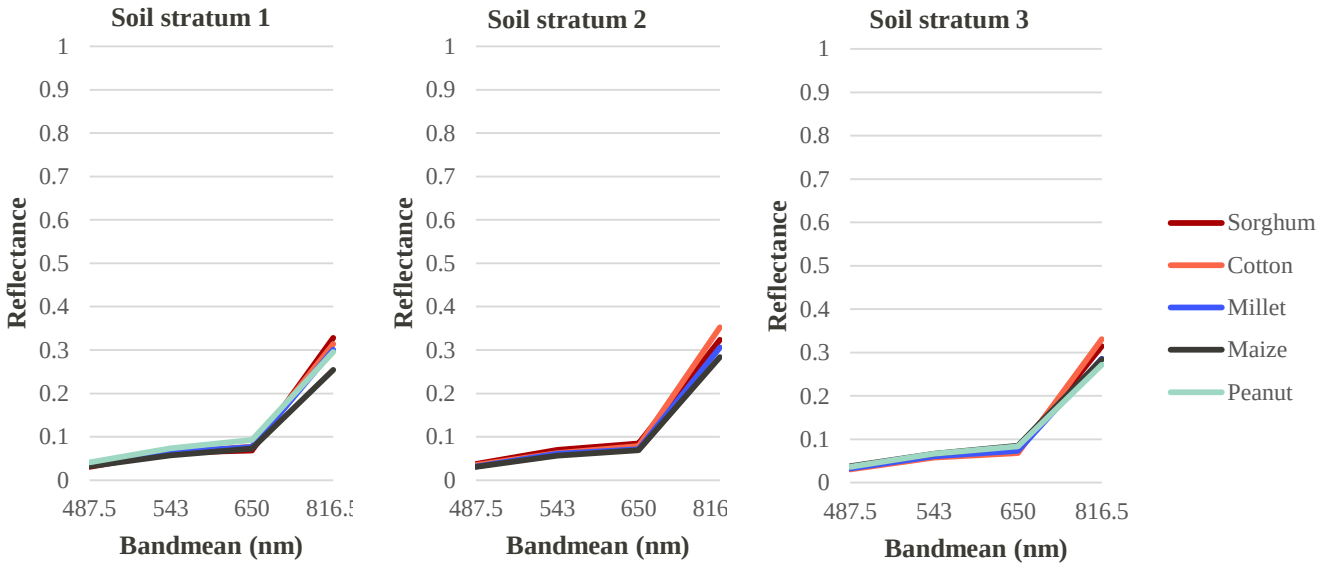
(F) Multi-temporal classification results of Sougoumba in 2015 (using fields of farmers in the STARS project)

	K-NN PVI	REGT PVI	K-NN TSAVI	REGT TSAVI
Overall accuracy	0.65	0.55	0.64	0.58
Average user accuracy	0.66	0.58	0.66	0.59
Average prod. accuracy	0.65	0.55	0.64	0.58
Kappa	0.57	0.44	0.56	0.47

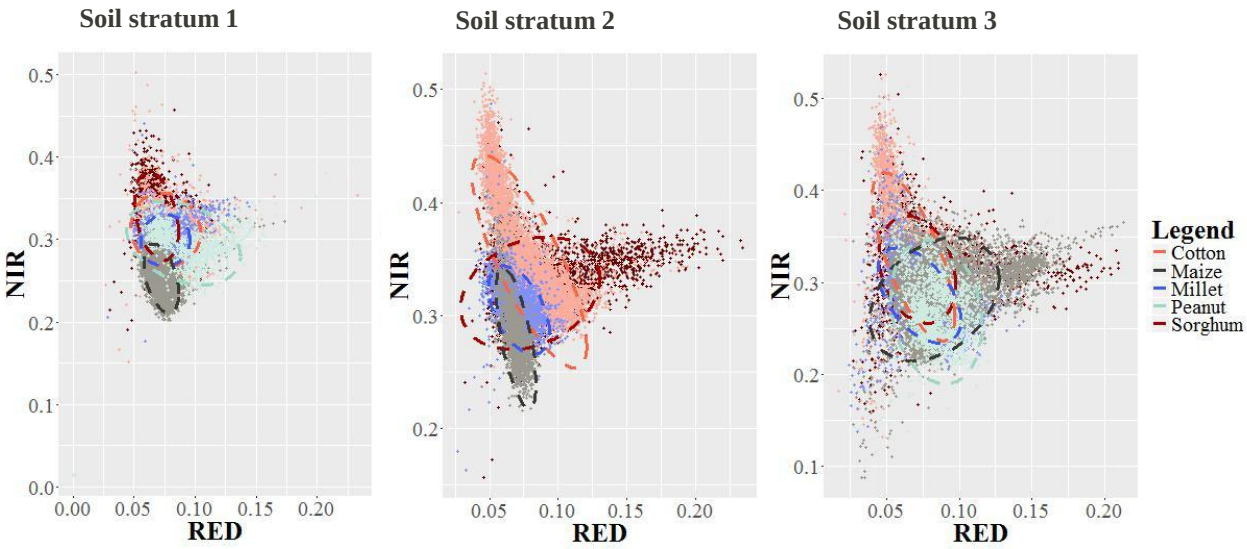
APPENDIX 10: MONO-TEMPORAL SIGNATURES AND FEATURE SPACE PLOTS FOR SOUGOUMBA WITH STRATIFICATION OF SOIL

Stratum 1: Plateau alluvium.
Stratum 2: Valley alluvium.
Stratum 3: Intermediate soils.

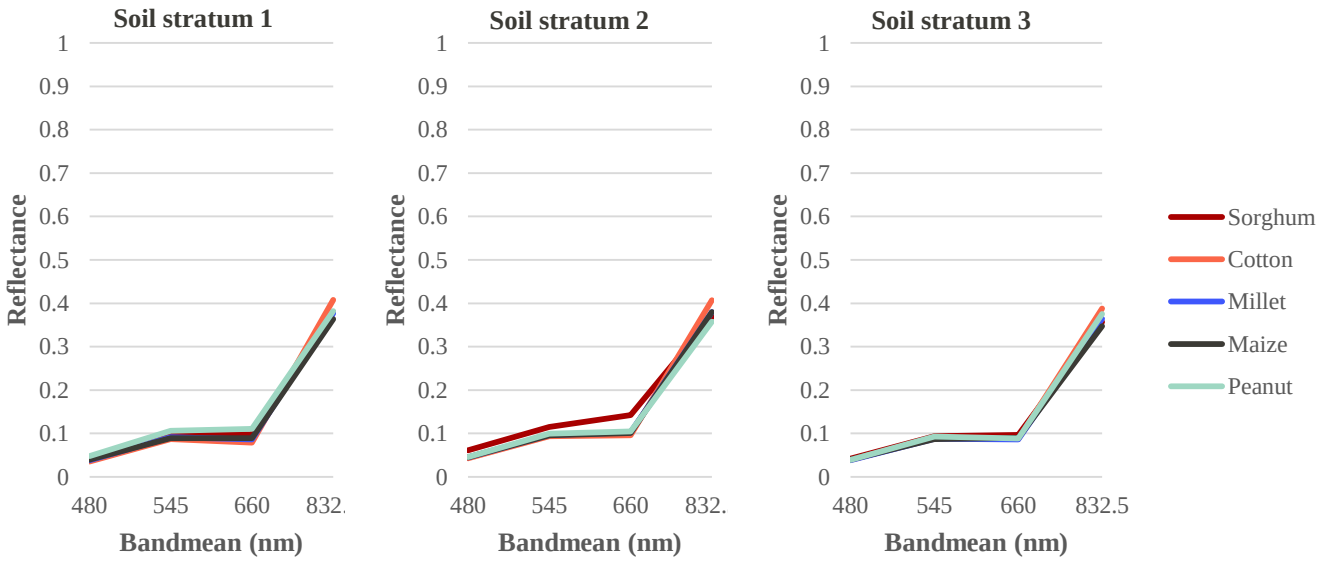
(A) Mean spectral profiles per soil stratum. Based on Quickbird image: 04-10-2014.



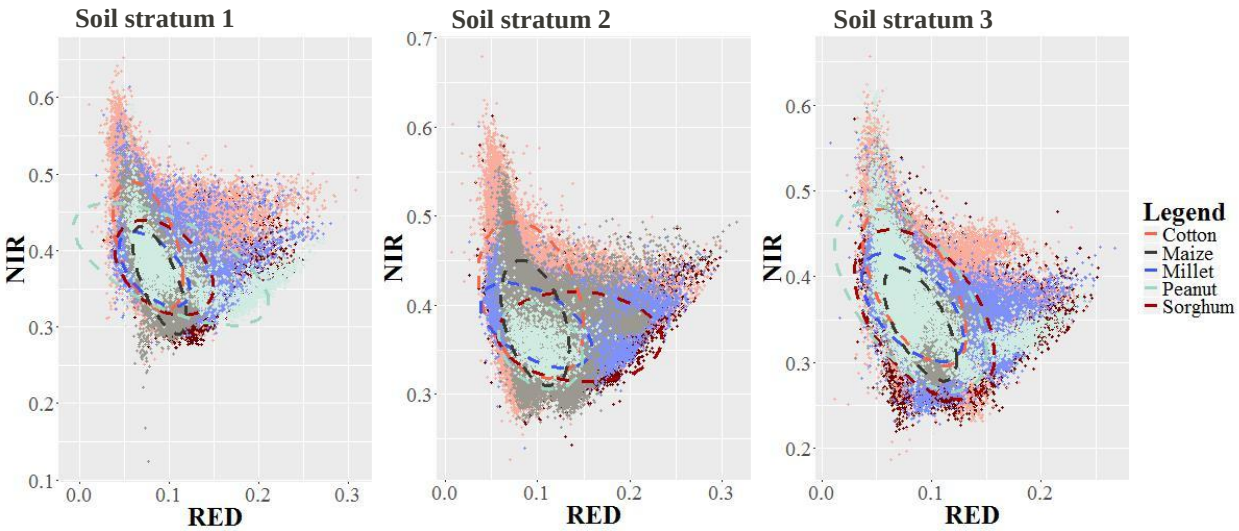
(B) Feature space plot per soil stratum of the NIR (816.5nm) and RED (650nm) band with the 95 percent confidence ellipse. Based on Quickbird image: 04-10-2014



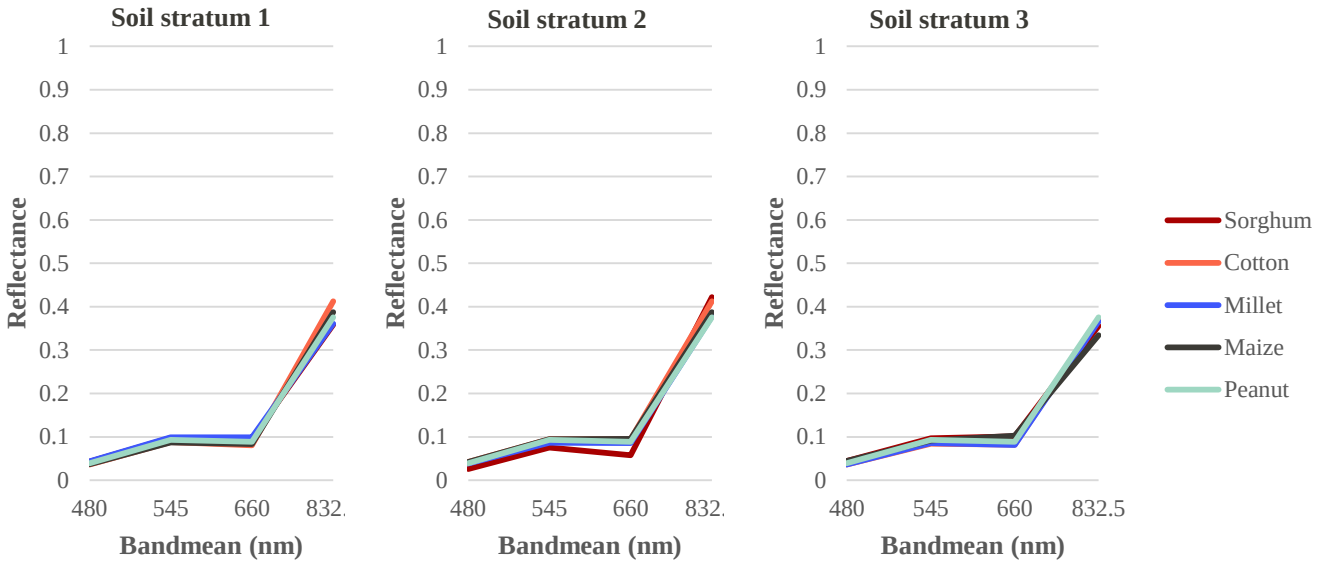
(C) Mean spectral profiles per soil stratum. Based on WorldView-2 image: 25-09-2015 (using fields of de Schaetzen (2015)).



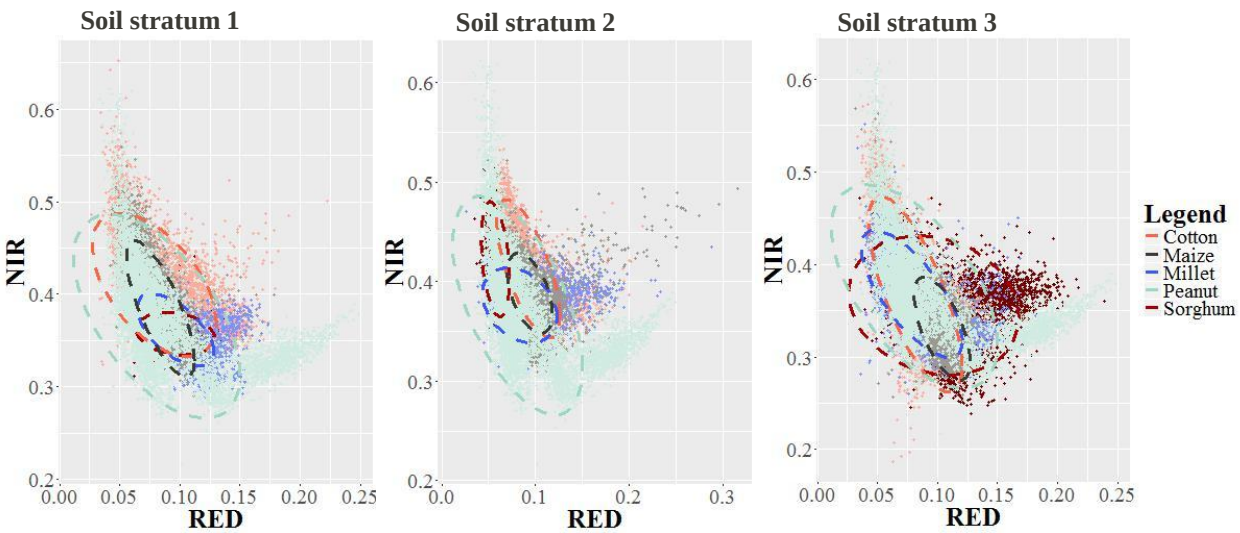
(D) Feature space plot per soil stratum of the NIR (832.5nm) and RED (660nm) band with the 95 percent confidence ellipse. Based on WorldView-2 image: 25-09-2015. Training fields of de Schaetzen (2015) were used create the feature space plot and mean profiles.



(E) Mean spectral profiles per soil stratum. Based on WorldView-2 image: 25-09-2015 (Using fields of farmers in the STARS project).



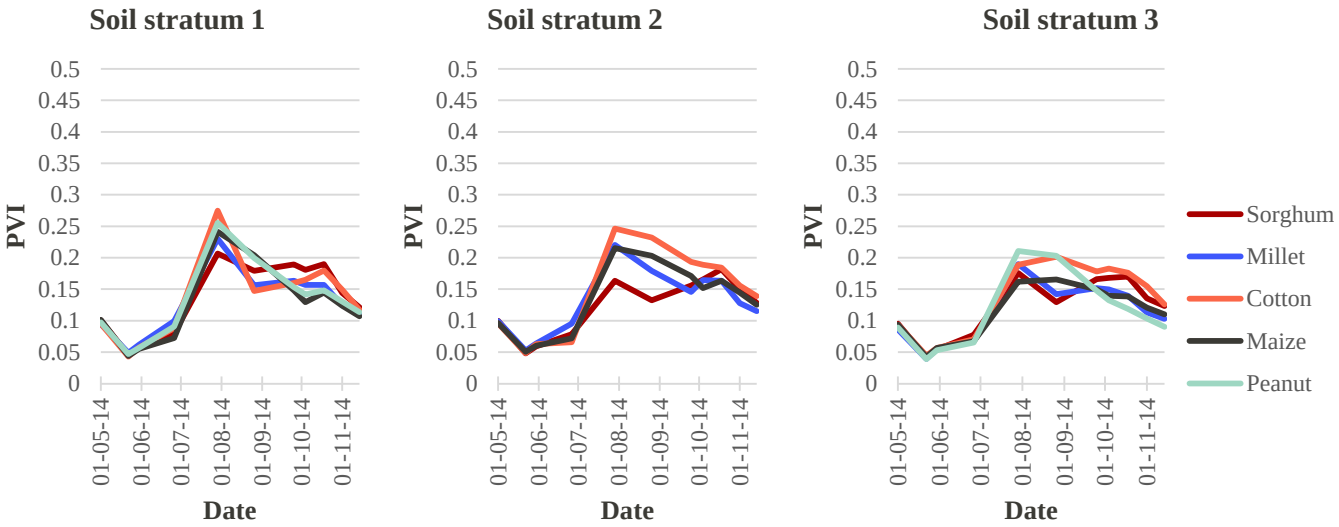
(F) Feature space plot per soil stratum of the NIR (832.5nm) and RED (660nm) band with the 95 percent confidence ellipse. Based on WorldView-2 image: 25-09-2015. Training fields of STARS (2015) were used create the feature space plot and mean profiles.



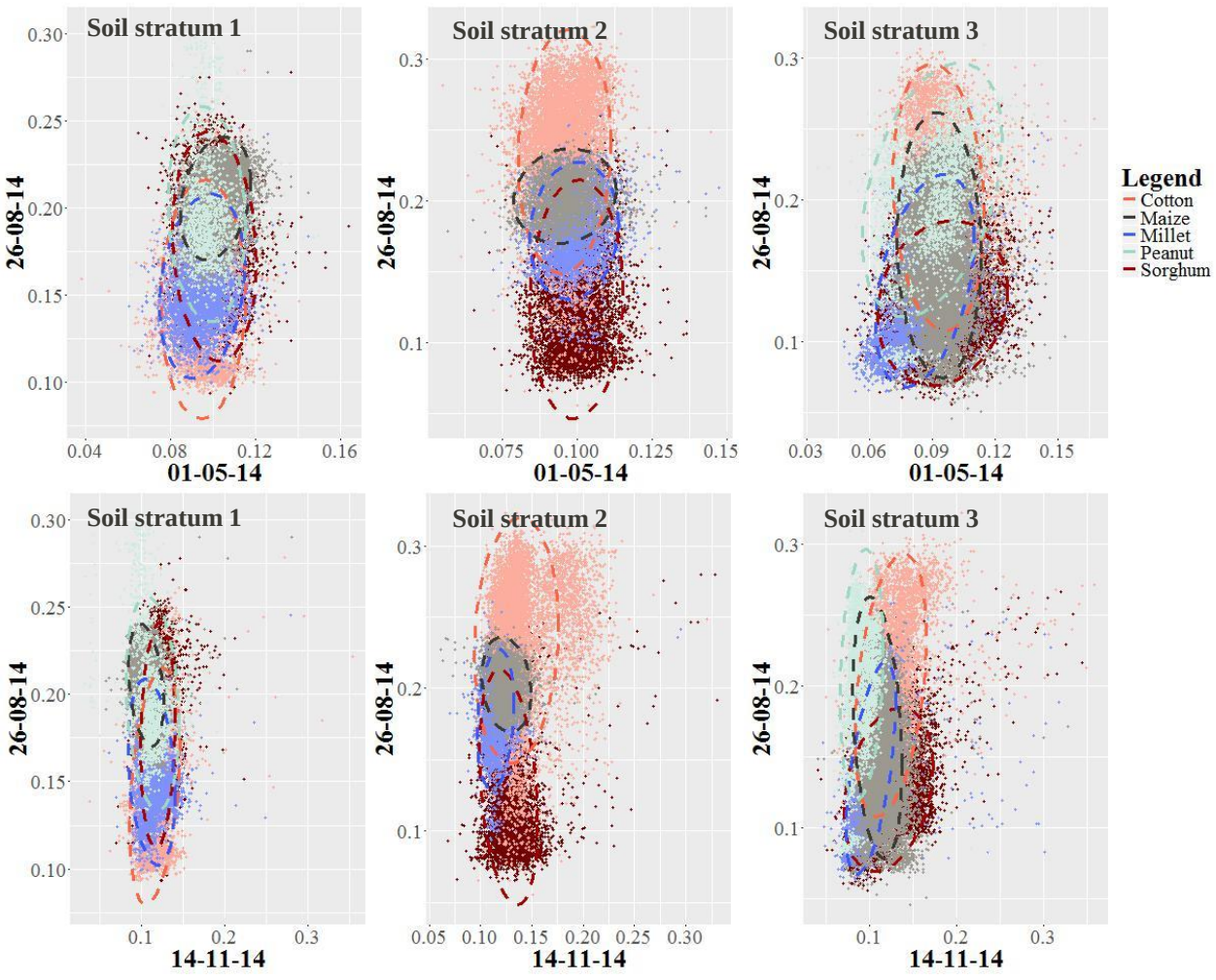
APPENDIX 11: TEMPORAL SIGNATURES AND FEATURE SPACE PLOTS FOR SOUGOUMBA WITH STRATIFICATION OF SOIL

Stratum 1: Plateau alluvium.
Stratum 2: Valley alluvium.
Stratum 3: Intermediate soils.

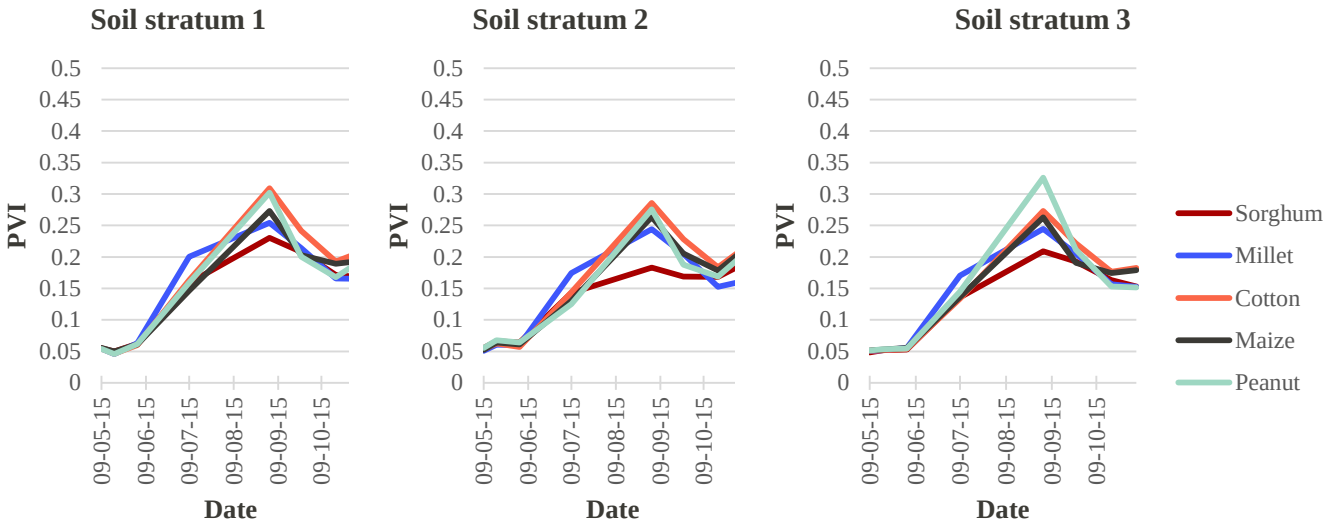
(A) Temporal mean PVI profiles per soil stratum. Based on satellite images of Sougoumba 2014.



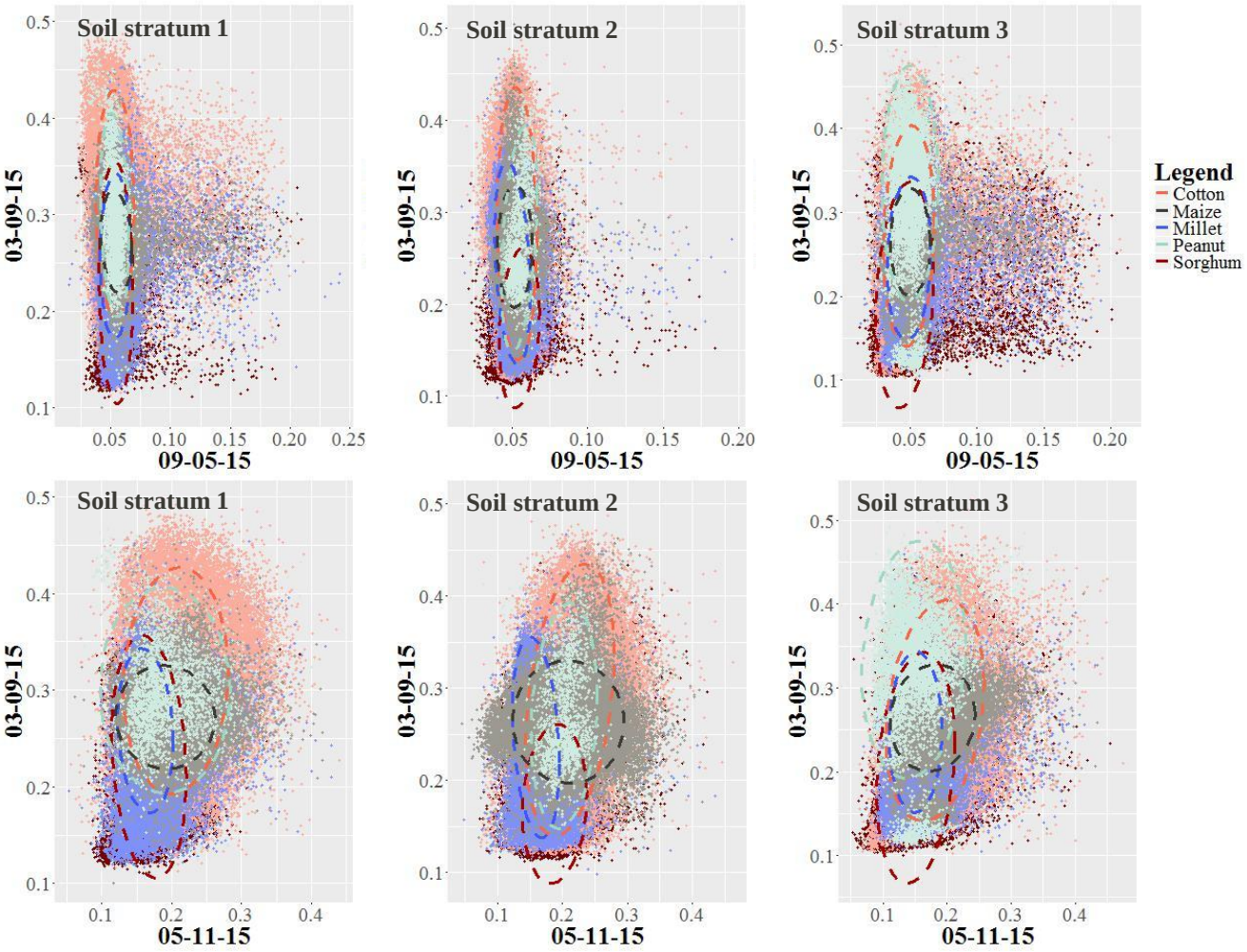
(B) Feature space plot of the early-middle and middle-end season PVI classes of 2014 with the 95 percent confidence ellipse.



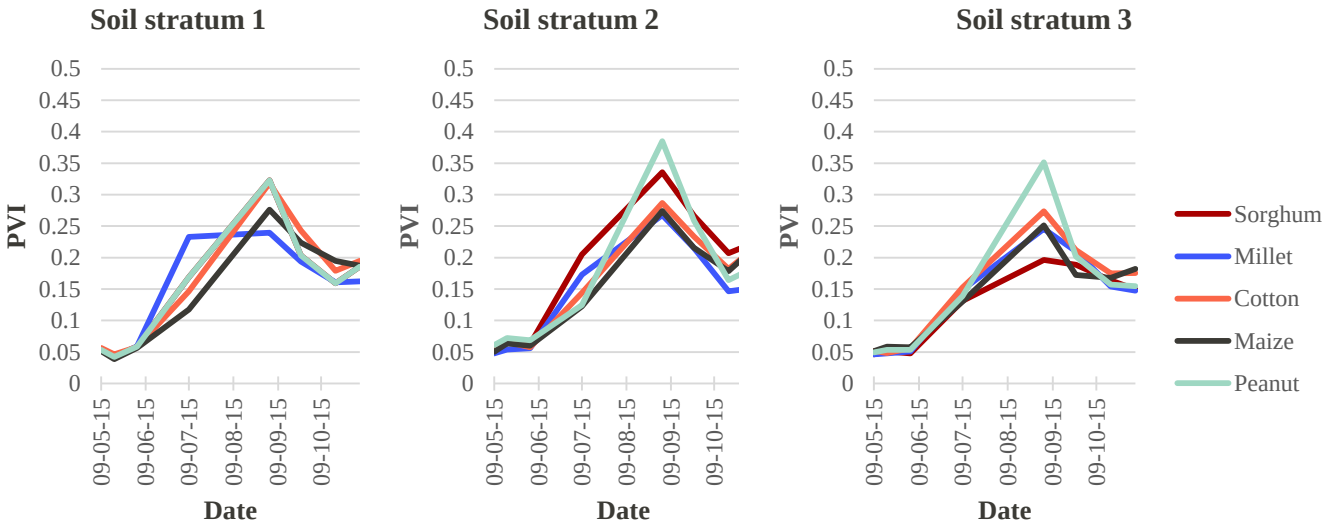
(C) Temporal mean PVI profiles per soil stratum. Based on satellite images of Sougoumba 2015 (using fields of de Schaetzen (2015)).



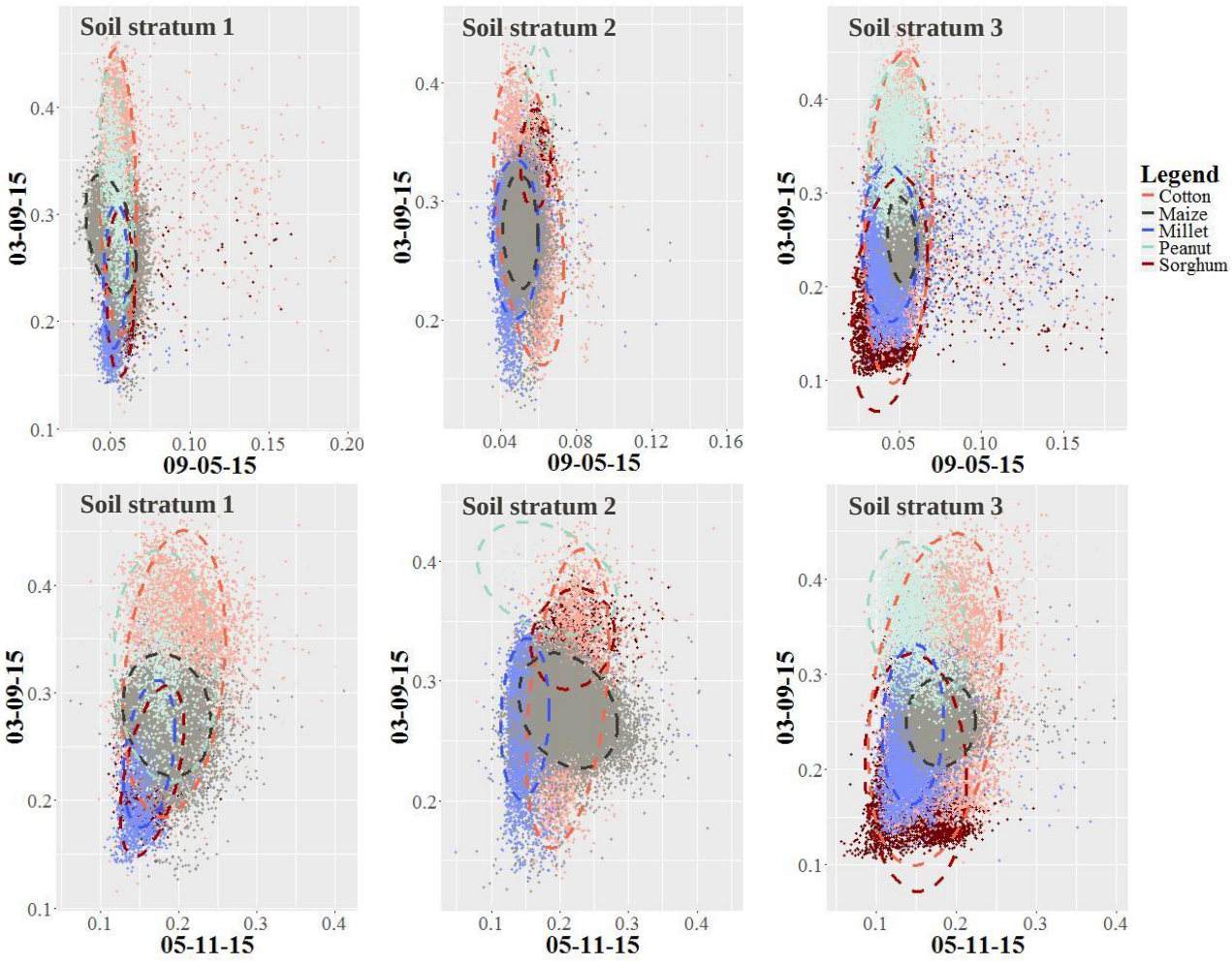
(D) Feature space plot of the early-middle and middle-end season PVI classes of 2015 with the 95 percent confidence ellipse (using fields of de Schaetzen (2015)).



(E) Temporal mean PVI profiles per soil stratum. Based on satellite images of Sougoumba 2015 (using fields of farmers in the STARS project).



(F) Feature space plot of the early-middle and middle-end season PVI classes of 2015 with the 95 percent confidence ellipse (using fields of farmers in the STARS project).



APPENDIX 12: MONO- AND MULTI-TEMPORAL CLASSIFICATION RESULTS OF SOUGOUMBA WITH STRATIFICATION

(A) Single image classification results per strata of Sougoumba in 2014

K-NN method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.56	0.53	0.53
Average user accuracy	0.55	0.54	0.54
Average prod. accuracy	0.55	0.53	0.53
Kappa	0.44	0.41	0.40

Regression tree method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.52	0.52	0.51
Average user accuracy	0.52	0.54	0.53
Average prod. accuracy	0.51	0.52	0.51
Kappa	0.39	0.40	0.38

(B) Single image classification results per strata of Sougoumba in 2015 (using fields of de Schaetzen (2015))

K-NN method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.42	0.40	0.40
Average user accuracy	0.42	0.42	0.41
Average prod. accuracy	0.42	0.40	0.40
Kappa	0.28	0.25	0.25

Regression tree method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.38	0.38	0.39
Average user accuracy	0.41	0.41	0.41
Average prod. accuracy	0.38	0.38	0.39
Kappa	0.23	0.23	0.24

(C) Single image classification results per strata of Sougoumba in 2015 (using fields of farmers in the STARS project)

K-NN method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.58	0.52	0.54
Average user accuracy	0.60	0.54	0.55
Average prod. accuracy	0.56	0.52	0.54
Kappa	0.48	0.41	0.43

Regression tree method

	Soil strata	Buildup strata	Elevation strata
Overall accuracy	0.55	0.48	0.49
Average user accuracy	0.56	0.50	0.50
Average prod. accuracy	0.53	0.48	0.49
Kappa	0.43	0.35	0.37

(D) Multi-temporal classification results per strata of Sougoumba in 2014

K-NN method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.81	0.80	0.81	0.80	0.79	0.78
Average user accuracy	0.81	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.81	0.61	0.62	0.57	0.57	0.63
Kappa	0.74	0.51	0.52	0.46	0.46	0.54

Regression tree method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.71	0.76	0.70	0.71	0.72	0.68
Average user accuracy	0.71	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.70	0.61	0.62	0.57	0.57	0.63
Kappa	0.61	0.51	0.52	0.46	0.46	0.54

(E) Multi-temporal classification results per strata of Sougoumba in 2015 (using fields of de Schaetzen (2015))

K-NN method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.61	0.61	0.62	0.57	0.57	0.63
Average user accuracy	0.63	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.61	0.61	0.62	0.57	0.57	0.63
Kappa	0.51	0.51	0.52	0.46	0.46	0.54

Regression tree method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.54	0.54	0.53	0.53	0.50	0.51
Average user accuracy	0.55	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.54	0.61	0.62	0.57	0.57	0.63
Kappa	0.42	0.51	0.52	0.46	0.46	0.54

(F) Multi-temporal classification results per strata of Sougoumba in 2015 (using fields of farmers in the STARS project)

K-NN method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.74	0.72	0.73	0.75	0.70	0.70
Average user accuracy	0.75	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.72	0.61	0.62	0.57	0.57	0.63
Kappa	0.68	0.51	0.52	0.46	0.46	0.54

Regression tree method

	PVI Soil	PVI Buildup	PVI Elevation	TSAVI Soil	TSAVI Buildup	TSAVI Elevation
Overall accuracy	0.71	0.70	0.68	0.71	0.65	0.66
Average user accuracy	0.72	0.63	0.63	0.59	0.58	0.65
Average prod. accuracy	0.69	0.61	0.62	0.57	0.57	0.63
Kappa	0.64	0.51	0.52	0.46	0.46	0.54

APPENDIX 13: STANDARD DEVIATION OF CLASSES IN SAMANKO

(A) Single image standard deviations per class. Based on 10 cm Tetracam image: 15-10-2015.

Band mean (nm)	Sorghum	Cotton	Millet
550	0.021	0.019	0.018
680	0.025	0.018	0.020
710	0.035	0.039	0.031
740	0.050	0.081	0.045
800	0.066	0.105	0.061

(B) Multi-temporal standard deviations per class. Based on Tetracam images of Samanko.

Date	Sorghum	Cotton	Millet
15-10-15	0.027	0.048	0.022
21-10-15	0.024	0.040	0.017
23-10-15	0.024	0.042	0.019
26-10-15	0.024	0.043	0.018
29-10-15	0.024	0.041	0.019
02-11-15	0.019	0.036	0.013
05-11-15	0.022	0.038	0.014
09-11-15	0.021	0.033	0.013
13-11-15	0.019	0.033	0.011
18-11-15	0.018	0.030	0.013
Average	**0.022**	**0.038**	**0.016**

APPENDIX 14: STANDARD DEVIATION OF CLASSES IN SOUGOUMBA WITHOUT STRATIFICATION

(A) Single image standard deviations per class. Based on Quickbird image: 04-10-2014.

Band mean (nm)	Sorghum	Millet	Cotton	Maize	Peanut
487.5	0.008	0.005	0.008	0.008	0.008
543	0.011	0.008	0.009	0.012	0.011
650	0.019	0.013	0.018	0.019	0.019
816.5	0.038	0.024	0.045	0.032	0.038

(B) Single image standard deviations per class. Based on WorldView-2 image: 25-09-2015 (using fields of de Schaetzen (2015)).

Band mean (nm)	Sorghum	Millet	Cotton	Maize	Peanut
480	0.016	0.012	0.011	0.008	0.018
545	0.019	0.014	0.013	0.010	0.018
660	0.040	0.028	0.026	0.019	0.043
832.5	0.039	0.028	0.043	0.036	0.047

(C) Single image standard deviations per class. Based on WorldView-2 image: 25-09-2015 (using fields of farmers in the STARS project)

Band mean (nm)	Sorghum	Millet	Cotton	Maize	Peanut
480	0.012	0.009	0.009	0.006	0.0150
545	0.013	0.012	0.012	0.008	0.0168
660	0.029	0.022	0.022	0.015	0.0367
832.5	0.034	0.027	0.047	0.036	0.0469

(D) Temporal standard deviations per class. Based on satellite images of 2014.

Date	Sorghum	Millet	Cotton	Maize	Peanut
01-05-14	0.012	0.011	0.009	0.010	0.014
22-05-14	0.006	0.007	0.005	0.005	0.006
30-05-14	0.007	0.006	0.006	0.005	0.005
26-06-14	0.020	0.024	0.016	0.009	0.025
29-07-14	0.054	0.049	0.075	0.061	0.057
26-08-14	0.038	0.033	0.051	0.038	0.039
25-09-14	0.029	0.026	0.040	0.026	0.036
04-10-14	0.024	0.021	0.038	0.022	0.029
18-10-14	0.029	0.023	0.041	0.024	0.025
01-11-14	0.027	0.020	0.034	0.020	0.027
14-11-14	0.021	0.016	0.024	0.014	0.018
Average	**0.024**	**0.021**	**0.031**	**0.021**	**0.026**

(E) Temporal standard deviations per class. Based on satellite images of 2015 (using fields of de Schaetzen (2015)).

Date	Sorghum	Millet	Cotton	Maize	Peanut
09-05-15	0.013	0.010	0.011	0.010	0.012
18-05-15	0.012	0.010	0.011	0.012	0.011
03-06-15	0.012	0.009	0.008	0.009	0.009
09-07-15	0.046	0.058	0.043	0.034	0.048
03-09-15	0.060	0.046	0.062	0.032	0.068
25-09-15	0.043	0.032	0.041	0.033	0.055
19-10-15	0.029	0.024	0.031	0.033	0.038
05-11-15	0.032	0.023	0.038	0.039	0.046
Average	**0.031**	**0.026**	**0.031**	**0.025**	**0.036**

(F) Temporal standard deviations per class. Based on satellite images of 2015 (using fields of farmers in the STARS project).

Date	Sorghum	Millet	Cotton	Maize	Peanut
09-05-15	0.014	0.011	0.014	0.006	0.009
18-05-15	0.011	0.008	0.013	0.014	0.010
03-06-15	0.011	0.008	0.007	0.008	0.008
09-07-15	0.049	0.056	0.040	0.015	0.042
03-09-15	0.062	0.038	0.069	0.028	0.046
25-09-15	0.037	0.030	0.043	0.033	0.049
19-10-15	0.031	0.024	0.033	0.026	0.032
05-11-15	0.034	0.021	0.037	0.034	0.036
Average	**0.031**	**0.024**	**0.032**	**0.020**	**0.029**

APPENDIX 15: AVERAGE STANDARD DEVIATION OF CLASSES IN SOUGOUMBA WITH A STRATIFICATION OF SOIL

(A) Average single image standard deviations per soil strata. Based on Quickbird image: 04-10-2014.

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,011	0,013	0,010	0,009	0,018
Stratum 2	0,021	0,021	0,011	0,011	0,018
Stratum 3	0,017	0,019	0,014	0,020	0,018

(B) Average single image standard deviations per soil strata. . Based on WorldView-2 image: 25-09-2015 (using fields of de Schaetzen (2015)).

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,021	0,021	0,017	0,015	0,034
Stratum 2	0,026	0,025	0,021	0,020	0,017
Stratum 3	0,026	0,022	0,020	0,015	0,031

(C) Average single image standard deviations per soil strata. Based on WorldView-2 image: 25-09-2015 (using fields of farmers in the STARS project)

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,010	0,022	0,013	0,015	0,031
Stratum 2	0,011	0,017	0,015	0,012	0,031
Stratum 3	0,024	0,021	0,018	0,013	0,031

(D) Average temporal standard deviations per soil strata. Based on PVI from satellite images of 2014.

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,020	0,025	0,017	0,015	0,022
Stratum 2	0,019	0,029	0,014	0,016	0,022
Stratum 3	0,026	0,028	0,022	0,021	0,024

(E) Average temporal standard deviations per soil strata. Based on PVI from satellite images of 2015 (using fields of de Schaetzen (2015)).

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,027	0,029	0,026	0,024	0,032
Stratum 2	0,024	0,027	0,024	0,023	0,019
Stratum 3	0,032	0,031	0,027	0,026	0,037

(F) Average temporal standard deviations per soil strata. Based on PVI from satellite images of 2015 (using fields of farmers in the STARS project)

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0,019	0,028	0,020	0,020	0,029
Stratum 2	0,018	0,025	0,019	0,016	0,020
Stratum 3	0,029	0,036	0,024	0,017	0,027

APPENDIX 16: PER-FIELD AGGREGATION RESULTS

(A) Confusion matrix of the per-field aggregation results of 2014. Based on satellite images of 2014 (using fields of farmers in the STARS project for pixel-based classification and per-field aggregation).

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	8	1	0	0	0	9	0.89
Cotton	0	8	0	0	0	8	1.00
Millet	1	0	10	1	1	13	0.77
Maize	0	1	0	8	1	10	0.80
Peanuts	0	0	0	1	6	7	0.86
Total	9	10	10	10	8	47	
Producer accuracy	0.89	0.80	1.00	0.80	0.75		
Overall accuracy	0.85						
Kappa	0.81						

(B) Confusion matrix of the per-field aggregation results of 2015. Based on satellite images of 2015 (using fields of farmers in the STARS project for pixel-based classification and per-field aggregation)

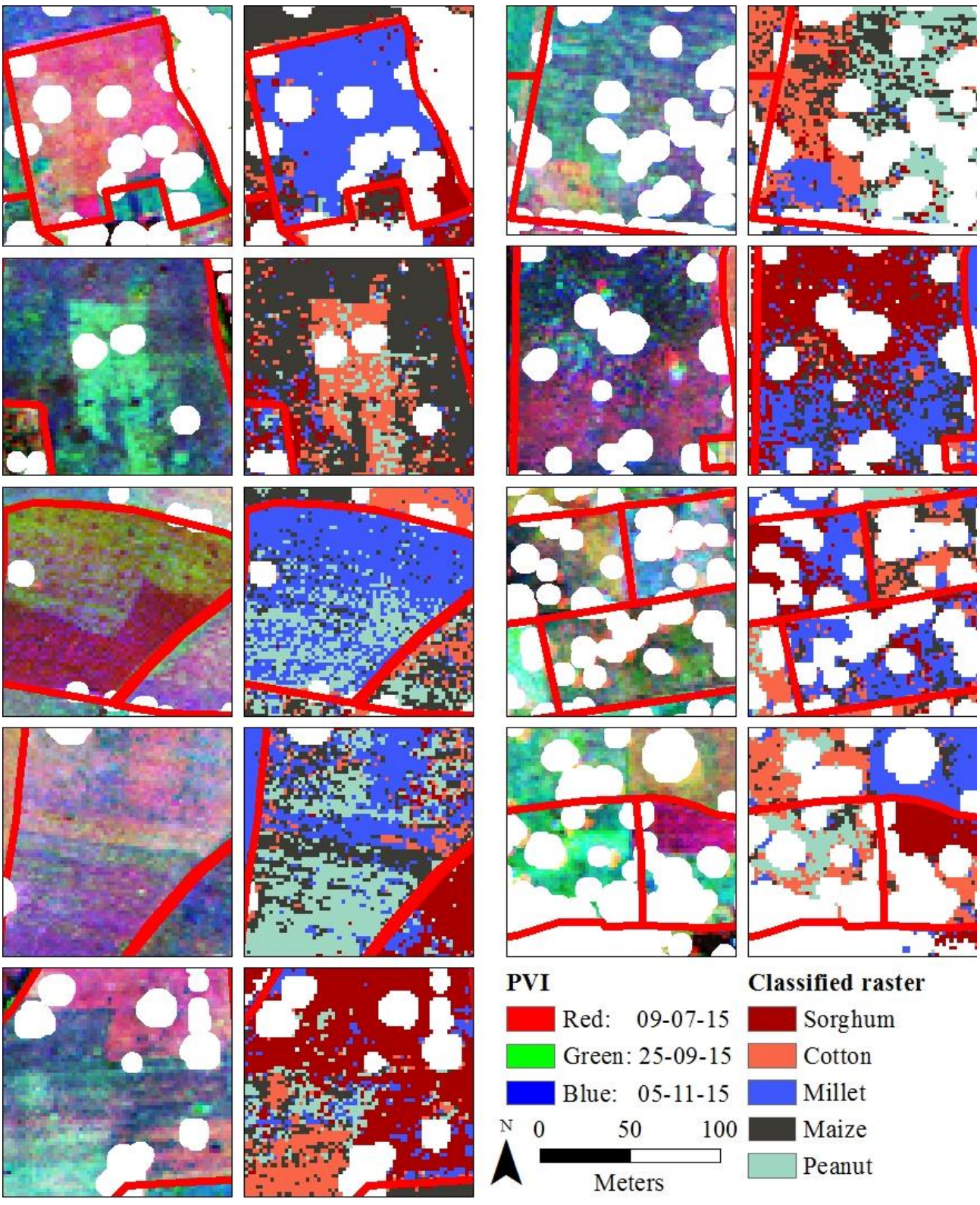
	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	**7**	1	1	0	0	9	0.78
Cotton	0	**5**	0	0	1	6	0.83
Millet	2	1	**9**	0	0	12	0.75
Maize	0	2	0	**10**	0	12	0.83
Peanuts	0	1	0	0	**9**	10	0.90
Total	9	10	10	10	10	49	
Producer accuracy	0.78	0.50	0.90	1.00	0.90		
Overall accuracy	0.82						
Kappa	0.77						

(C) Confusion matrix of the per-field aggregation results of 2015. Based on satellite images of 2015 (using fields of farmers in the STARS project for pixel-based classification and fields of de Schaetzen (2015) for per-field aggregation)

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	**46**	24	28	8	2	108	0.43
Cotton	27	**65**	10	7	7	116	0.56
Millet	24	22	**75**	7	3	131	0.57
Maize	11	44	5	**55**	3	118	0.47
Peanuts	2	9	1	4	**15**	31	0.48
Total	110	164	119	81	30	504	
Producer accuracy	0.42	0.40	0.63	0.68	0.50		
Overall accuracy	0.51						
Kappa	0.37						

APPENDIX 17: MISCLASSIFICATIONS DUE TO WITHIN-FIELD VARIABILITY OR MISPLACED BOUNDARIES

Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation). The left images contain a RGB image of the PVI index, each colour represents a different date. The right images contain the resulting classified raster (using K-NN with the PVI).



APPENDIX 18: TEMPORAL RESOLUTION RESULTS

A) Pixel-based accuracy results with different temporal resolutions. Based on satellite images of 2014.

	8 bands	7 bands	6 bands	5 bands	4 bands	3 bands	2 bands	1 band
Overall acc.	0.76	0.73	0.73	0.71	0.68	0.65	0.48	0.39
Avg. user acc.	0.78	0.74	0.74	0.72	0.70	0.66	0.49	0.38
Avg. prod. acc.	0.76	0.72	0.72	0.70	0.67	0.63	0.47	0.37
Kappa	0.70	0.66	0.65	0.63	0.60	0.56	0.34	0.22

(B) Pixel-based accuracy results with different temporal resolutions. Based on satellite images of 2015 (using fields of de Schaetzen (2015)).

	8 bands	7 bands	6 bands	5 bands	4 bands	3 bands	2 bands	1 band
Overall acc.	0.61	0.60	0.57	0.60	0.54	0.53	0.48	0.33
Avg. user acc.	0.63	0.62	0.58	0.63	0.56	0.54	0.47	0.32
Avg. prod. acc.	0.61	0.60	0.57	0.60	0.54	0.53	0.48	0.33
Kappa	0.51	0.50	0.46	0.50	0.43	0.41	0.35	0.16

(C) Pixel-based accuracy results with different temporal resolutions. Based on satellite images of 2015 (using fields of farmers in the STARS project)

	8 bands	7 bands	6 bands	5 bands	4 bands	3 bands	2 bands	1 band
Overall acc.	0.74	0.74	0.73	0.74	0.69	0.69	0.62	0.37
Avg. user acc.	0.75	0.75	0.73	0.75	0.70	0.70	0.64	0.34
Avg. prod. acc.	0.72	0.72	0.71	0.72	0.67	0.67	0.61	0.36
Kappa	0.68	0.67	0.66	0.67	0.61	0.62	0.53	0.20

(D) Pixel-based accuracy results with different temporal resolutions for all training datasets. Using 3 classes chosen in the green up, peak of the season and senescence phase.

	2014	2015 (de Schaetzen)	2015 (STARS)
Overall acc.	0.60	0.36	0.58
Avg. user acc.	0.62	0.36	0.58
Avg. prod. acc.	0.60	0.36	0.57
Kappa	0.49	0.20	0.48

APPENDIX 19: SPATIAL RESOLUTION RESULTS

(A) Pixel-based accuracy results with different spatial resolutions. Based on satellite images of 2014.

	4 m res.	6 m res.	8 m res.
Overall accuracy	0.80	0.83	0.82
Average user accuracy	0.80	0.84	0.83
Average prod. accuracy	0.80	0.83	0.79
Kappa	0.75	0.79	0.78

(B) Pixel-based accuracy results with different spatial resolutions. Based on satellite images of 2015 (using fields of de Schaetzen (2015)).

	4 m res.	6 m res.	8 m res.
Overall accuracy	0.62	0.64	0.65
Average user accuracy	0.60	0.66	0.66
Average prod. accuracy	0.62	0.64	0.59
Kappa	0.53	0.55	0.57

(C) Pixel-based accuracy results with different spatial resolutions. Based on satellite images of 2015 (using fields of farmers in the STARS project)

	4 m res.	6 m res.	8 m res.
Overall accuracy	0.80	0.80	0.82
Average user accuracy	0.79	0.81	0.82
Average prod. accuracy	0.78	0.78	0.79
Kappa	0.75	0.75	0.77

APPENDIX 20: PER-FIELD RESULTS WITH OPTIMAL TEMPORAL RESOLUTION (5 DATES)

(A) Confusion matrix of the per-field aggregation results with the 5 selected images. Based on satellite images of 2014.

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	**7**	0	1	0	0	8	0.88
Cotton	1	**8**	0	0	0	9	0.89
Millet	1	0	**9**	1	1	12	0.75
Maize	0	2	0	**7**	0	9	0.78
Peanuts	0	0	0	2	**7**	9	0.78
Total	9	10	10	10	8	47	
Producer accuracy	0.78	0.80	0.90	0.70	0.88		
Overall accuracy	0.81						
Kappa	0.76						

(B) Confusion matrix of the per-field aggregation results with the 5 selected images. Based on satellite images of 2015 (using fields of de Schaetzen (2015) for pixel-based classification and per-field aggregation)

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	**69**	25	24	0	3	121	0.57
Cotton	12	**97**	2	4	5	120	0.81
Millet	19	8	**85**	9	2	123	0.69
Maize	7	28	7	**63**	1	106	0.59
Peanuts	3	6	1	5	**19**	34	0.56
Total	110	164	119	81	30	504	
Producer accuracy	0.63	0.59	0.71	0.78	0.63		
Overall accuracy	0.66						
Kappa	0.64						

(C) Confusion matrix of the per-field aggregation results with the 5 selected images. Based on satellite images of 2015 (using fields of farmers in the STARS project)

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	**6**	1	1	0	0	8	0.75
Cotton	0	**6**	0	0	1	7	0.86
Millet	3	1	**9**	0	0	13	0.69
Maize	0	1	0	**10**	1	12	0.83
Peanuts	0	1	0	0	**8**	9	0.89
Total	9	10	10	10	10	49	
Producer accuracy	0.67	0.60	0.90	1.00	0.80		
Overall accuracy	0.80						
Kappa	0.74						

APPENDIX 21: PER-FIELD RESULTS WITH OPTIMAL SPATIAL RESOLUTION (8 METER)

(A) Confusion matrix of the per-field aggregation results with an 8 meter spatial resolution. Based on satellite images of 2014.

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	8	2	2	0	0	12	0.67
Cotton	0	8	0	0	0	8	1.00
Millet	1	0	8	1	1	11	0.73
Maize	0	0	0	7	2	9	0.78
Peanuts	0	0	0	2	5	7	0.71
Total	9	10	10	10	8	47	
Producer accuracy	0.89	0.80	0.80	0.70	0.63		
Overall accuracy	0.77						
Kappa	0.71						

(B) Confusion matrix of the per-field aggregation results with an 8 meter spatial resolution. Based on satellite images of 2015 (using fields of de Schaetzen (2015))

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	72	23	23	1	2	121	0.60
Cotton	14	104	2	8	5	133	0.78
Millet	16	8	89	7	3	123	0.72
Maize	6	20	4	59	1	90	0.66
Peanuts	2	9	1	6	19	37	0.51
Total	110	164	119	81	30	504	
Producer accuracy	0.65	0.63	0.75	0.73	0.63		
Overall accuracy	0.68						
Kappa	0.59						

(C) Confusion matrix of the per-field aggregation results with an 8 meter spatial resolution. Based on satellite images of 2015 (using fields of farmers in the STARS project)

	Sorghum	Cotton	Millet	Maize	Peanuts	Total	User accuracy
Sorghum	6	0	1	0	0	7	0.86
Cotton	1	6	0	0	0	7	0.86
Millet	2	1	9	0	0	12	0.75
Maize	0	3	0	9	0	12	0.75
Peanuts	0	0	0	1	10	11	0.91
Total	9	10	10	10	10	49	
Producer accuracy	0.67	0.60	0.90	0.90	1.00		
Overall accuracy	0.82						
Kappa	0.77						

APPENDIX 22: AVERAGE STAND. DEV. OF CLASSES IN SOUGOUMBA WITH A SOIL STRATIFICATION AND AN 8 METER RESOLUTION

(A) Average temporal standard deviations per soil strata. Based on PVI derived from satellite images of 2014.

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0.022	0.027	0.019	0.022	0.031
Stratum 2	0.013	0.028	0.020	0.016	0.015
Stratum 3	0.029	0.033	0.027	0.024	0.025

(B) Average temporal standard deviations per soil strata. Based on PVI derived from satellite images of 2015 (using fields of de Schaetzen (2015)).

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0.023	0.022	0.019	0.021	0.024
Stratum 2	0.022	0.027	0.019	0.016	0.024
Stratum 3	0.021	0.029	0.024	0.026	0.028

(C) Average temporal standard deviations per soil strata. Based on PVI derived from satellite images of 2015 (using fields of farmers in the STARS project)

Strata	Sorghum	Millet	Cotton	Maize	Peanut
Stratum 1	0.022	0.027	0.019	0.022	0.031
Stratum 2	0.013	0.028	0.020	0.016	0.015
Stratum 3	0.029	0.033	0.027	0.024	0.025

APPENDIX 23: REMOVAL OF THE TEMPORAL BANDS

(A) Removal based on standard deviations

Order of removal	2014	2015
1	22-05-14	18-05-15
2	30-05-14	09-05-15
3	01-05-14	03-06-15
4	26-06-14	25-09-15
5	25-09-14	19-10-15
6	26-08-14	03-09-15
7	18-10-14	19-10-15

(B) The dates representing the green up, peak of season and senescence phase

2014	2015
26-07-14	09-07-15
25-09-14	03-09-15
18-10-14	19-10-15

APPENDIX 24: CONFUSION MATRICES OF THE PER-FIELD AGGREGATIONS WITH THE OPTIMAL TEMPORAL AND SPATIAL RESOLUTION.

(A) Confusion matrix of the per-field aggregation results with the 5 selected images an 8 meter spatial resolution. Based on satellite images of 2014.

	Sorghum	**Cotton**	**Millet**	**Maize**	**Peanut**	**Field Count**	**User accuracy**
Sorghum	**7**	0	1	1	0	9	0.78
Cotton	1	**8**	0	0	0	9	0.89
Millet	1	0	**9**	1	2	13	0.69
Maize	0	2	0	**7**	0	9	0.78
Peanut	0	0	0	1	**6**	7	0.86
Field count	9	10	10	10	8	47	
Producer accuracy	0.78	0.80	0.90	0.70	0.75		
Overall accuracy	0.79						
Kappa	0.73						

(B) Confusion matrix of the per-field aggregation results with the 5 selected images an 8 meter spatial resolution. Based on satellite images of 2015 (using fields of farmers in the STARS project).

	Sorghum	**Cotton**	**Millet**	**Maize**	**Peanut**	**Field Count**	**User accuracy**
Sorghum	**7**	0	1	0	0	8	0.88
Cotton	0	**8**	0	0	2	10	0.80
Millet	2	0	**9**	0	0	11	0.82
Maize	0	2	0	**10**	0	12	0.83
Peanut	0	0	0	0	**8**	8	1.00
Field count	9	10	10	10	10	49	
Producer accuracy	0.78	0.80	0.90	1.00	0.80		
Overall accuracy	0.86						
Kappa	0.82						

APPENDIX 25: SOWING DATES PER CROP IN SOUGOUMBA

(A) STARS fields in 2014 (some dates are missing)

Cotton	Maize	Sorghum	Millet	Peanut
15-05-14	05-06-14	11-05-14	15-05-14	03-06-14
25-05-14	11-06-14	17-05-14	20-05-14	10-06-14
03-06-14	15-06-14	15-06-14	22-05-14	16-06-14
07-06-14	15-06-14	19-06-14	22-05-14	17-06-14
08-06-14	16-06-14	20-06-14	28-05-14	19-06-14
10-06-14	22-06-14	22-06-14	15-06-14	21-06-14
10-06-14	23-06-14	18-07-14	17-06-14	25-06-14
11-06-14	26-06-14		20-06-14	01-07-14
23-06-14	30-06-14		23-06-14	04-07-14
	03-07-14			

(B) STARS fields in 2015

Cotton	Maize	Sorghum	Millet	Peanut
23-05-15	11-06-15	16-05-15	15-05-15	02-06-15
24-05-15	16-06-15	19-05-15	16-05-15	05-06-15
27-05-15	17-06-15	20-05-15	19-05-15	14-06-15
28-05-15	18-06-15	23-05-15	21-05-15	16-06-15
01-06-15	20-06-15	28-05-15	21-05-15	19-06-15
02-06-15	22-06-15	03-06-15	21-05-15	20-06-15
04-06-15	27-06-15	08-06-15	21-05-15	24-06-15
04-06-15	27-06-15	19-06-15	23-05-15	27-06-15
05-06-15	28-06-15	20-06-15	23-05-15	27-07-15
17-06-15	03-07-15	08-07-15	25-05-15	27-07-15