

**DETECTION, IDENTIFICATION, AND MAPPING OF MAIZE STREAK VIRUS AND
GREY LEAF SPOT DISEASES OF MAIZE USING DIFFERENT REMOTE
SENSING TECHNIQUES**

By

INOS DHAU

THESIS

Submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

GEOGRAPHY

in the

FACULTY OF SCIENCE AND AGRICULTURE

(School of Agricultural and Environmental Sciences)

At the

UNIVERSITY OF LIMPOPO SOUTH AFRICA

Supervisor:	Dr E. Adam	(University of the Witwatersrand)
Co-Supervisors:	Prof K.K Ayisi	(University of Limpopo)
	Prof O. Mutanga	(University of KwaZulu-Natal)

2019

Table of Contents

Table of Contents	i
DECLARATION.....	viii
DEDICATION	ix
ACKNOWLEDGEMENTS.....	x
LIST OF TABLES	xi
Publications and Manuscripts	xv
Abstract.....	xvii
Chapter 1	1
GENERAL INTRODUCTION.....	1
1.1 Introduction.....	1
1.2 Problem statement	6
1.3 Aim and objectives.....	7
1.4 Reliability, validity and objectivity	7
1.5 Bias.....	8
1.6 Significance of the study	8
1.7 Format of the thesis	9
1.7.1 Chapter One	9
1.7.2 Chapter Two	10
1.7.3 Chapter Three	10
1.7.4 Chapter Four	10
1.7.5 Chapter Five	11
1.7.6 Chapter Six	11
1.7.7 Chapter Seven	11
1.7.8 Chapter Eight	12
1.7.9 Chapter Nine	12

1.8 References.....	13
Chapter 2.....	17
LITERATURE REVIEW.....	17
Abstract.....	18
2.0 Introduction.....	19
2.1 Applications of remote sensing in monitoring crops.....	21
2.3 Multi- and hyperspectral sensors.....	24
2.4 Remote sensing of maize diseases.....	26
2.5 Challenges associated with remote sensing of maize diseases.....	26
2.6 Conclusions.....	28
2.7 References.....	29
Chapter 3.....	36
Dimensionality reduction using variable selection methods.....	36
Abstract.....	37
3.0 Introduction.....	38
3.1 Variable Ranking.....	41
3.2 Random forest and regularized random forest.....	42
3.3 Resampling field spectra.....	44
3.4 Optimizing using the random forest.....	45
3.5 Variable Selection.....	45
3.6 Accuracy assessment.....	47
3.7 Results.....	47
3.7.1 Variables importance using the random forest algorithm.....	47
3.7.2 Random Forest's Forward Variable selection.....	49
3.7.3 Variable selection using GRRF.....	51
3.7.4 Accuracy Assessment.....	52

3.8 Discussion.....	53
3.9 Conclusion	55
3.10 References	56
Chapter 4	64
TESTING THE UTILITY OF IN-SITU HYPERSPECTRAL DATA IN DETECTING THE SEVERITY OF MAIZE STREAK VIRUS.....	64
Abstract.....	65
4.0 Introduction.....	66
4.1 Methodology	69
4.1.1 Study Area	69
4.2 Maize leaves sampling	70
4.3 Remote sensing of different levels of MSV infections	71
4.5 Discriminating against different levels of MSV infection using remotely sensed data.....	73
4.6 Accuracy assessment	73
4.7 Results.....	73
4.7.1 Spectral separability of MSV disease infestation levels on maize.....	73
4.7.2 Selection of optimal variables for discriminating different levels of MSV infections.	74
4.7.3 Discrimination accuracy assessment.....	75
4.8 Discussion.....	77
4.8.1 Remote sensing different levels of MSV infections on maize	77
4.9 Conclusions	79
Chapter 5	85
UNDERSTANDING THE SUBTLE SPATIAL DISTRIBUTION OF MSV DISEASE USING HIGH MULTISPECTRAL RESOLUTION DATA.....	85
Abstract.....	86
5.1 Materials and Methods	89

5.1.1 Description of the study area	89
5.2 Field data collection.....	90
5.3 Satellite data acquisition and pre-processing	91
5.4 Spectral analysis and characterization	91
5.5 Image classification and accuracy assessment	92
5.6 Improving classification using vegetation indices.....	93
5.7 Results.....	94
5.7.1 Spectral separability of land cover classes	94
5.7.2 Image classification	95
5.7.3 Variable importance ranking	96
5.7.4 Spatial distribution of MSV infected maize and other land cover classes.....	98
5.8 Discussion.....	99
5.9 Conclusion	101
5.10 References	102
Chapter 6	107
MAPPING MAIZE STREAK VIRUS DISEASE USING LANDSAT 8 DATA	107
Abstract.....	108
6.0 Introduction.....	109
6.1 Materials and Methods	111
6.1.1 Description of the study area	111
6.1.2 Field data collection and pre-processing	112
6.2 Image classification and accuracy assessment	113
6.3 Results.....	115
6.3.1 Spectral response of maize infected with MSV and other land cover types.....	115
6.4 Image classification	115

6.4.1 Classification results using Landsat 8 spectral bands	115
6.4.2 Classification results using computed vegetation indices	116
6.5 Spatial distribution of MSV infected maize and other land cover classes	117
6.6 Discussion	119
6.7 Implications of the study on food security in Sub Saharan Africa	120
6.8 Conclusion	121
6.9 References	122
Chapter 7	125
SPATIAL MODELLING OF MAIZE STREAK VIRUS DISEASE USING ENVIRONMENTAL VARIABLES IN SEMI-ARID ENVIRONMENTS	125
Abstract	126
7.0 Introduction	127
7.1 Materials and Methods	129
7.1.1 Study Area	129
7.1.2 Field data collection	131
7.2 Modelling	132
7.3 Results	133
7.3.1 Pearson's correlation analysis	134
7.3.2 Selection of optimally MSV disease predictor variables	136
7.3.3 Probability of MSV disease occurrence as a function of the selected variables	136
7.4 Probability mapping	139
7.5 Discussion	140
7.5.1 Implications of the study on maize crop disease control	141
7.6 Conclusion	142
7.8 References	143

Chapter 8.....	145
TESTING THE POTENTIAL OF MULTISPECTRAL SENSORS IN DETECTING GREY LEAF SPOT DISEASE OF MAIZE	145
Abstract.....	146
8.0 Introduction.....	147
8.1 Material and methods.....	151
8.1.1 Study Area	151
8.1.2 Field spectral measurements and processing.....	152
8.2 Random forest classifier and accuracy assessment	157
8.3 Results	159
8.3.1 Optimizing random forest parameters	159
8.3.2 Classification results	160
8.3.3 Measuring the importance of variables (bands) in detecting the GLS stages in maize.....	162
8.4 Discussion.....	163
8.5 Conclusions	165
8.6 References.....	166
Chapter 9.....	173
IDENTIFICATION, DETECTION, AND MAPPING OF MAIZE STREAK VIRUS AND GREY LEAF SPOT DISEASES:	173
A SYNTHESIS	173
9.0 Introduction.....	173
9.1 To compare Random forest’s forward variable selection and guided regularized random forest methods for optimum variable selection.....	176
9.2 Detect the severity of maize streak virus infestations in maize hybrid lines using in-situ hyperspectral data	177
9.3 Detect and map maize streak virus using RapidEye satellite imagery	178

9.4 Investigating the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations	179
9.5 Spatial modelling of maize streak virus disease using environmental variables	180
9.6 Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot infection in maize hybrid line	180
9.7 Conclusions and future perspectives	181

DECLARATION

I declare that the thesis hereby submitted to the University of Limpopo, for the degree of Doctor of Philosophy in Geography has not been submitted previously by me or anybody for a degree at this or any other University. Also, this is my work in design and in execution, and related materials contained herein have been duly acknowledged.

05/04/2019

Candidate: INOS DHAU

Date



05/04/2019

Supervisor: Dr E. Adam

Date



05/04/2019

Co-supervisor: Prof O. Mutanga

Date



05/04/2019

Co-supervisor: Prof K.K Ayisi

Date

DEDICATION

To my parents, (Mr. G. and the Late Mrs. D. Dhau Mawere), as well as the rest of the family.

Heatherwell, Tinotenda, Makanaka, Akatendeka Dhau

I love you

ACKNOWLEDGEMENTS

This entire work would not have been possible without the guidance, critical review and constructive comments from my supervisors, Dr. E. Adam, Prof O. Mutanga and Prof K.K Ayisi. My PhD. was financially supported by the University of Limpopo under the staff development programme and VLIR-IUC programme. I would like to thank these institutions for giving me the opportunity to study for my PhD. Their financial support for procuring a very expensive remote sensing instrument (Spectroradiometer, ASD Spec 4 (R800 000), field work, satellite images, and data analyses is highly appreciated. Without this sponsorship, my dream to pursue a PhD. would not have been possible. My appreciation also goes to Dr. M.R Ramudzuli, and Dr T. Dube for their immense words of advice and encouragement during the write up of this thesis. I would like to also thank Clodean Mothapo and Thabang Maphanga who assisted me during the field work with technical support on some aspects of the field data collection and analysis. To my wife, *Heatherwell*, our children *Tino, Maka, Aka*, and our parents and relatives, I appreciate your moral support during the period of my studies at the University of Limpopo. Above all, I thank Jehovah, The Almighty God, the Father of my Lord and Saviour Jesus Christ, for His unconditional love in granting me life, good health, divine protection, wisdom, and understanding during this project. Without Jehovah God, what had been accomplished would not have been possible. Heavenly Father may the glory, the honour be unto you always forever and ever, Amen.

LIST OF TABLES

Table 3.1: Shows bands selected by GRRF from AISA Eagle (n= 6) yielded a higher accuracy (89.17%) while those selected using FVS for AISA Eagle (n=10) yielded an overall accuracy of 85.85%	52
Table 3.2: Confusion matrix for AISA Eagle and Hymap showing the classification accuracy for the different levels of maize streak virus severity (H- Healthy, E- Early stage, M-Moderate and S-Severe)	53
Table 4.1: Confusion matrix derived based on the optimal variables selected by the GRRF algorithm in discriminating different levels of maize infections by MSV.....	76
Table 5.1: Summary of RapidEye derived indices used in this study	93
Table 5.2: RapidEye transformed divergence indices showing interclass separability of land cover classes.....	94
Table 5.3: Classification accuracies (%) for RapidEye derived spectral bands and computed vegetation indices	96
Table 6.1: Landsat 8 OLI spectral bands (http://landsat.usgs.gov)	113
Table 6.2: Vegetation indices used in the classification process.....	114
Table 7.1: Climatic, environmental and remotely sensed variables.....	131
Table 7.2: Descriptive statistics.....	133
Table 7.3: Correlation between variables and area affected with MSV disease	135
Table 7.4: Optimally selected remotely sensed and climatic variables in predicting MSV disease prevalence.....	136
Table 8.1: Multispectral sensors and their spectral properties.	156
Table 8.2: Overall accuracy (OA) and Kappa index of agreement achieved by field spectroradiometer and four multispectral sensors calculated using the 2013 training and an independent 2014 testing datasets.....	160

Table 8.3: Confusion matrix used to develop producer's and user's accuracy (%) for Sentinel-2 for the different levels of maize grey leaf spot (H- Healthy, M-Moderate and S-Severe). OOB method was used to evaluate the accuracy on the independent test dataset acquired in 2014.....162

LIST OF FIGURES

Figure 1.1 Disease Cycle	4
Figure 3.1 The variable importance as measured by traditional random forest for resampled AISA Eagle (a) and resampled Hymap (b). The errors were calculated using the mean decrease in the Gini index and the default settings. The black arrow indicates the most important wavelengths.....	48
Figure 3.2: Forward variable selection method AISA dataset (a) and HyMap dataset (b).....	50
Figure 4.1 Location of the study area.....	70
Figure 4.2 (a) Health, (b) Early, (c) Moderate and (d) Severe stages of maize streak virus infections on maize plants (courtesy Dhau, 2015).	71
Figure 4.3 Field spectra for healthy, early, medium and severely infected maize leaves by the MSV	72
Figure 4.4 Spectral profiles of healthy, early, medium and severely infested maize leaves.....	74
Figure 4.5 Important bands for detecting MSV on maize leaves.	75
Figure 4.6 Producer and User accuracies derive based on the optimal variables in discriminating different levels of MSV infections.	76
Figure 5.1 Location of the study area.....	90
Figure 5.2 Spectral characteristics of infected and non-infected maize	92
Figure 5.3 a and b: The importance of RapidEye spectral bands and derived vegetation indices in detection and mapping MSV.....	97
Figure 5.4 Land cover classification map of infected maize and other classes	98
Figure 6.1 Location of the study area.....	112
Figure 6.2 Spectral reflectance of maize infected with MSV and other landcover types.....	115
Figure 6.3 Classification accuracy for the infected and health maize and other classes using spectral bands values	116
Figure 6.4 Classification accuracy for the infected and health maize and other classes using spectral bands values and vegetation indices.....	117
Figure 6.5 Land cover classification map of infected maize and other classes	118

Figure 6.6 Estimated Area of each land-cover class in Ofcolaco farm, 2016.	119
Figure 7.1 Location of the study area	130
Figure 7.2 Environmental variables that were used in this study.....	132
Figure 7.3: Relationship between disease occurrence and remotely sensed and climatic variables.....	138
Figure 7.4: Maize streak virus probability of occurrence map	139
Figure 8.1: Boxplots showing the median, quartile data and the spread of data for healthy, moderate and severely infected maize with grey leaf spot for the two seasons. NDVIs were calculated from the field spectroradiometer measurements.....	153
Figure 8.2: Disease status, mean reflectance and number of samples for healthy, moderate and severe maize grey leaf spot infestation for the spectroradiometer measurements and multispectral sensors. The arrows show where major differences between the three disease stages are located.....	155
Figure 8.3: Comparison between the mean reflectance of the three GLS infestation stages (healthy, moderate and severely).....	156
Figure 8.4: Optimization of the random forest parameters (mtry and ntree) using the grid search procedure. The OOB method was used to determine the error rates for the different combinations; 60 combinations for Quickbird (A), 240 combinations for Sentinel-2 (B), 160 combinations for Worldview-2 (C) and 80 combinations for RapidEye (D).....	160
Figure 8.5: The importance of each band for different sensors used in this study for distinguishing the three stages of GLS infestation in maize using random forest (RF).....	163

Publications and Manuscripts

Inos Dhau, Elhadi Adam, Onesimo Mutanga, Kwabena Ayisi, Elfatih Mohamed Abdel-Rahman, John Odindi & Mhosisi Masocha (2017) Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop, Geocarto International, DOI: [10.1080/10106049.2017.1343391](https://doi.org/10.1080/10106049.2017.1343391) Pages 1-14

Inos Dhau, Elhadi Adam, Onesimo Mutanga & Kingsley K. Ayisi (2017) Detecting the severity of maize streak virus infestations in maize crop using *in situ* hyperspectral data, Transactions of the Royal Society of South Africa, 73:1, 8-15, DOI: [10.1080/0035919X.2017.1370034](https://doi.org/10.1080/0035919X.2017.1370034), Pages 8-15

Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga (2018) Detection and mapping of maize streak virus using RapidEye satellite imagery, Geocarto International, DOI: [10.1080/10106049.2018.1450448](https://doi.org/10.1080/10106049.2018.1450448)

Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga (Under Review) Comparative analysis of Random forest forward variable selection and guided regularized random forest techniques for optimum band selection. *Transactions of the Royal Society of South Africa*

Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga (Under Review) Investigating the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations. *African Journal of Ecology*

Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga (Under Review) Spatial modelling of maize streak virus disease using environmental variables in semi-arid environments. *African Journal of Ecology*,

Poster Conference presentations

- 1. Comparative analysis of Random forest forward variable selection and guided regularized random forest techniques for optimum band selection.** *Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga, The Society of South African Geographers 10th Biennial conference Stellenbosch 25-29 September 2016*
- 2. Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop**
Inos Dhau, Elhadi Adam, Kingsley K Ayisi, Onesimo Mutanga, SA-GEO Symposium 2013 University of Fort Hare

All the conceptual, field work, analysis of data, and preparation of above publications and manuscripts were accomplished by the candidate, Inos Dhau, under the supervision of Dr E. Adam, Prof O. Mutanga and Prof K.K Ayisi.

Abstract

Of late climate change and consequently, the spread of crop diseases has been identified as one of the major threat to crop production and food security in sub-Saharan Africa. This research, therefore, aims to evaluate the role of in situ hyperspectral and new generation multispectral data in detecting maize crop viral and fungal diseases, that is maize streak virus and grey leaf spot respectively. To accomplish this objective; a comparison of two variable selection techniques (Random Forest's Forward Variable, (FVS) and Guided Regularized Random Forest: (GRRF) was done in selecting the optimal variables that can be used in detecting maize streak virus disease using in-situ resampled hyperspectral data. The findings indicated that the GRRF model produced high classification accuracy (91.67%) whereas the FVS had a slightly lower accuracy (87.60%) based on Hymap when compared to the AISA. The results have shown that the GRRF algorithm has the potential to select compact feature sub sets, and the accuracy performance is better than that of RF's variable selection method. Secondly, the utility of remote sensing techniques in detecting the geminivirus infected maize was evaluated in this study based on experiments in Ofcolaco, Tzaneen in South Africa. Specifically, the potential of hyperspectral data in detecting different levels of maize infected by maize streak virus (MSV) was tested based on Guided Regularized Random Forest (GRRF). The findings illustrate the strength of hyperspectral data in detecting different levels of MSV infections. Specifically, the GRRF model was able to identify the optimal bands for detecting different levels of maize streak disease in maize. These bands were allocated at 552 nm, 603 nm, 683 nm, 881 nm, and 2338 nm. This study underscores the potential of using remotely sensed data in the accurate detection of maize crop diseases such as MSV and its severity which is critical in crop monitoring to foster food security, especially in the resource-limited sub-Saharan Africa. The study then investigated the possibility to upscale the previous findings to space borne sensor. RapidEye data and derived vegetation indices were tested in detecting and mapping the maize streak virus. The results revealed that the use of RapidEye spectral bands in detection and mapping of maize streak virus disease yielded good classification results with an overall accuracy of 82.75%. The inclusion of RapidEye derived vegetation indices improved the classification accuracies by 3.4%. Due to the cost involved in acquiring commercial images, like

RapidEye, a freely available Landsat-8 data can offer a new data source that is useful for maize diseases estimation, in environments which have limited resources. This study investigated the use of Landsat 8 and vegetation indices in estimating and predicting maize infected with maize streak virus. Landsat 8 data produced an overall accuracy of 50.32%. The inclusion of vegetation indices computed from Landsat 8 sensor improved the classification accuracies by 1.29%. Overall, the findings of this study provide the necessary insight and motivation to the remote sensing community, particularly in resource-constrained regions, to shift towards embracing various indices obtained from the readily-available and affordable multispectral Landsat-8 OLI sensor. The results of the study show that the medium-resolution multispectral Landsat 8-OLI data set can be used to detect and map maize streak virus disease. This study demonstrates the invaluable potential and strength of applying the readily-available medium-resolution, Landsat-8 OLI data set, with a large swath width (185 km) in precisely detecting and mapping maize streak virus disease. The study then examined the influence of climatic, environmental and remotely sensed variables on the spread of MSV disease on the Ofcolaco maize farms in Tzaneen, South Africa. Environmental and climatic variables were integrated together with Landsat 8 derived vegetation indices to predict the probability of MSV occurrence within the Ofcolaco maize farms in Limpopo, South Africa. Correlation analysis was used to relate vegetation indices, environmental and climatic variables to incidences of maize streak virus disease. The variables used to predict the distribution of MSV were elevation, rainfall, slope, temperature, and vegetation indices. It was found that MSV disease infestation is more likely to occur on low-lying altitudes and areas with high Normalised Difference Vegetation Index (NDVI) located at an altitude ranging of 350 and 450 m.a.s.l. The suitable areas are characterized by temperatures ranging from 24°C to 25°C. The results indicate the potential of integrating Landsat 8 derived vegetation indices, environmental and climatic variables to improve the prediction of areas that are likely to be affected by MSV disease outbreaks in maize fields in semi-arid environments. After realizing the potential of remote sensing in detecting and predicting the occurrence of maize streak virus disease, the study further examined its potential in mapping the most complex disease; Grey Leaf Spot (GLS) in maize fields using WorldView-2, Quickbird, RapidEye, and Sentinel-2 resampled from hyperspectral data. To accomplish this objective, field spectra were acquired from healthy, moderate and

severely infected maize leaves during the 2013 and 2014 growing seasons. The spectra were then resampled to four sensor spectral resolutions – namely WorldView-2, Quickbird, RapidEye, and Sentinel-2. In each case, the Random Forest algorithm was used to classify the 2013 resampled spectra to represent the three identified disease severity categories. Classification accuracy was evaluated using an independent test dataset obtained during the 2014 growing season. Results showed that Sentinel-2 achieved the highest overall accuracy (84%) and kappa value (0.76), while the WorldView-2, produced slightly lower accuracies. The 608 nm and 705nm were selected as the most valuable bands in detecting the GLS for Worldview 2, and Sentinel-2. Overall, the results imply that opportunities exist for developing operational remote sensing systems for detection of maize disease. Adoption of such remote sensing techniques is particularly valuable for minimizing crop damage, improving yield and ensuring food security.

Chapter 1

GENERAL INTRODUCTION

1.1 Introduction

A wide range of crops in both commercial and smallholders' farms are being affected by biotic and abiotic factors and resulting in significant yield losses. Plant diseases account for at least a 10% loss of the world's food production, (Christou and Twyman, 2004, Strange and Scott, 2005). There is an increase in the cost of production and toxic residues in agricultural products and pollutants in the environment due to excessive use of pesticides and fungicides. Patches of disease within fields needs to be timely and treated locally in order to effectively control these diseases. This requires one to obtain information of infected areas in the field as early and accurately as possible. Most people use the conventional method of data collection. The most common and conservative method is manual field survey. The old-fashioned ground-based survey method requires high labour costs and produces low efficiency. Thus, it is impracticable for large areas. Fortunately, remote sensing technology can provide spatial distribution information of diseases over a large area with relatively low costs. The presence of diseases on plants or canopy surface causes changes in pigment, chemical concentrations, cell structure, nutrient, water uptake, and gas exchange. These changes result in differences in colour and temperature of the canopy and affect canopy reflectance characteristics, which can be detectable by remote sensing (Raikes and Burpee, 1998). Therefore, remote sensing provides a harmless, rapid, and cost-effective means of identifying and quantifying crop stress from differences in the spectral characteristics of canopy surfaces affected by biotic and abiotic stress agents.

Maize (*Zea mays L.*) contributes 15–50% of the energy in human diets in sub-Saharan Africa (Kagoda et al., 2011, Archetti et al., 2009). In South Africa, maize is the main diet for the majority of the African population as well as being the major livestock feed grain (Langyintuo et al., 2010, Walker and Schulze, 2006). Furthermore, 50% of maize in Southern African Development Community (SADC) is

produced in South Africa (Benhin, 2008). Therefore, maize is considered to be the second largest crop produced in South Africa after sugar cane. Maize (*Zea mays L.*) is the most essential grain crop in South Africa and is produced throughout the country under different environments. Approximately 8,0 million tons of maize grain is produced in South Africa annually on about 3,1 million ha of land. Half of the production consists of white maize, for human food consumption.

Maize is grown in two main zones in South Africa, a marginal western belt, and a reliable and higher productivity eastern core. In these areas, maize is highly dependent on climate variables such as temperature and precipitation (Benhin, 2008). Only less than 10% of the maize is produced under irrigation (Benhin, 2008). Therefore, climate change and variability is one of the key factors influencing year to year crop maize production and thus information on climate has to be deliberated during the planning of agricultural activities and related decision making (Kang et al., 2009, Moeletsi and Walker, 2012).

However the productivity of maize and quality of maize products is impacted negatively by pests and diseases worldwide, despite many years of research and development on improved methods for their control. Research has shown that some of these diseases are causing serious yield losses on maize farms. The yields of maize have greatly reduced due to disease infection. The reduction in the crop's yield, affects the returns obtained by the maize farmers. The cost of controlling the diseases increases the cost of production, thus decreasing the gross margin. The most prevalent of these diseases that threaten the stability of maize production in the southern African region, particularly in South Africa, are Grey leaf spot (GLS) (Derera et al., 2008), northern corn leaf blight (NCLB) (Degefu et al., 2004, Welz and Geiger, 2000) *Phaeophaeria* leaf spot (PLS) (Sibiya et al., 2011) and Maize Streak virus (MSV).

Maize streak disease is a prevalent virus disease of maize in sub-Saharan Africa and several Indian Ocean Islands (Fuller, 1901, Redinbaugh and Zambrano-Mendoza, 2014, Lapierre and Signoret, 2004, Agriculture and Storey, 1925, Thottappilly et al., 1993). The disease continues to cause major losses and food insecurity across sub-Saharan Africa (Shepherd et al., 2010). The disease is caused

mainly by Maize streak virus-strain A (MSV-A; family Geminiviridae; genus Mastrevirus), which is spread in a tenacious manner by leafhoppers in the genus *Cicadulina*, especially *C. mbila* (Naude´). Distinctive symptoms include longitudinal chlorotic streaks along the leaf veins and a reduction in plant growth and yield. Early planted maize crops serve as pools of both virus and vectors in regions with staggered, overlapping growing seasons, and a number of wild grasses can also serve as virus pools (Shepherd et al., 2010, Konate and Traore, 1992).

Maize streak virus is controlled using virus-resistant maize hybrids and cultivars, by eliminating virus pools, and by decreasing vector populations. MSV can be spread using viruliferous vectors, and this approach is often used in genetic studies and in the development of resistant hybrids and cultivars. In the laboratory, MSV and virus clones can be spread using vascular puncture inoculation (Redinbaugh, 2003) and clones can be transmitted using agroinfiltration (Boulton et al., 1993). The importance of MSV disease as a constraint to maize production has been reported from many African countries including Kenya (Bock et al., 1977), South Africa (Van Rensburg, 1981), Zimbabwe (Mzira, 1984), Zaire (Vogel et al., 1993), Nigeria and other West African countries (Fajemisin et al., 1984).

Grey leaf spot of maize is now known as one of the most important yield-limiting disease of maize worldwide and in the province of KwaZulu-Natal. It is not only a threat to maize production in the commercial farming sector; it also reduces yields of maize on small-scale farms. Grey leaf spot (GLS) is a foliar fungal disease that affects, maize crop. There are two fungal pathogens that cause GLS, which are *Cercospora zea-maydis* and *Cercospora zeina*. GLS caused by *Cercospora zea-maydis* was first identified from specimens collected in 1924 by Tehon and Daniels in Alexander County, in southern Illinois near the Mississippi River (Ward et al., 1999). In South Africa, GLS was first noted in KwaZulu-Natal during 1988 and has since spread rapidly to neighbouring provinces (Ward et al., 1997). Currently, GLS is recognised as the most significant yield-limiting disease of maize crop (*Zea mays L.*) in South Africa especially in areas with warm temperatures and extended relative humidity (Wegary et al., 2003, Derera et al., 2008, Ward et al., 1999, Lyimo et al., 2012).

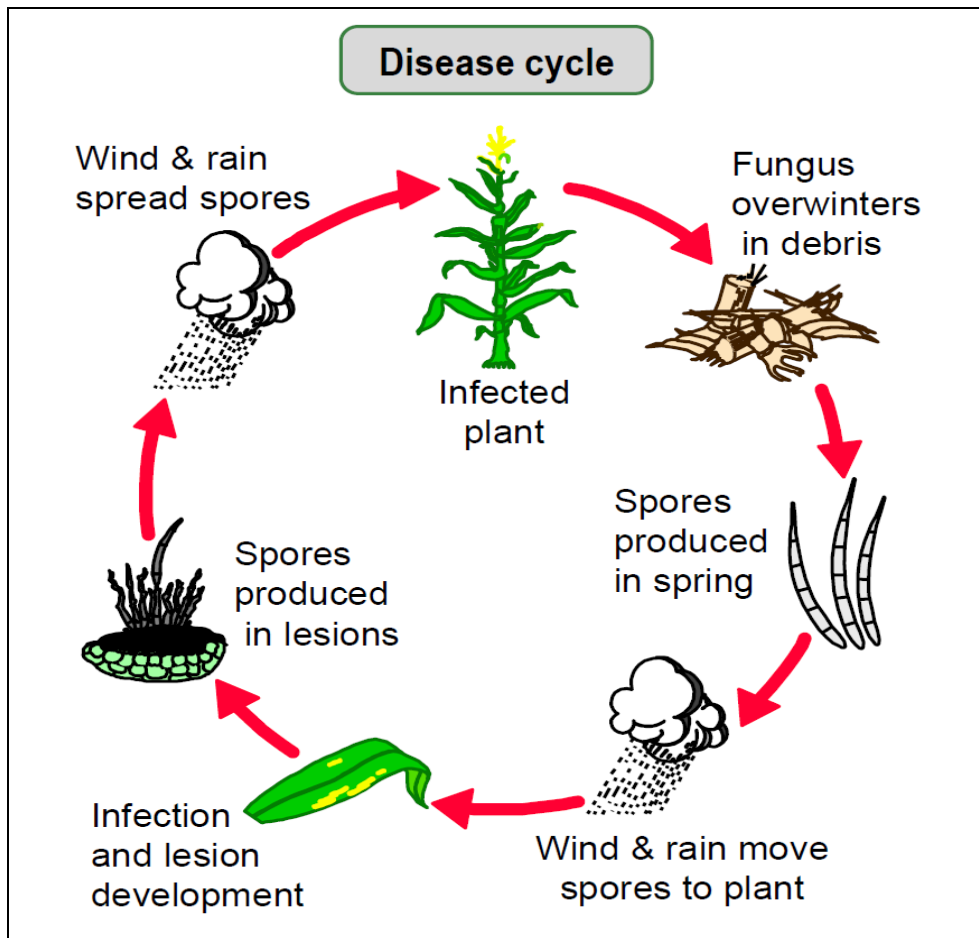


Figure 1.1 Disease Cycle

Grey leaf spot reduces the maize yield by damaging photosynthetic tissue and increasing stem and root lodging, Figure 1.1 (Derera et al., 2008). High yield reduction of 69% is attributable to GLS in South Africa (Derera et al., 2008). Grey leaf spot is dependent on recurrent and extended periods of overcast, high relative humidity and warm temperatures to complete spore germination and the infection process (Beckman and Payne, 1982). Spores (conidia) are produced from infested residues of previous maize crops in spring under conditions of high humidity and these are windblown to infect the newly planted maize crop. The fungus lives in the debris of topsoil and infects healthy crop via asexual spores called conidia. GLS impact on crop yield and quality, and also reduce resource-use efficiency. Management techniques include crop resistance, crop rotation, residue management, use of fungicides, and weed control. Value-added crop protection

approaches to prevent such damage and loss can increase production and make a substantial contribution to food security.

Currently, no studies have been conducted in South Africa to evaluate the effect of remote sensing on GLS and MSV diseases detection or to compare different methods of detecting GLS and MSV on maize under field conditions. Remote sensing technologies are one basic tool of precision agricultural practice which can provide an alternative to visual disease assessment (Nutter Jr and Schultz, 1995). Discovery and control of maize diseases are a vital task in agricultural management. Presently, the South African government spends several million rands annually to detect crop diseases. Farmers spend millions trying to control maize diseases every year. Traditional ways of field observations and ground surveys are used to collect the information about diseases in the fields. However, these approaches give results which often differ from what is on the ground.

Remote sensing has proved to be reliable tool and techniques that agronomists and plant pathologists can use to detect and record crop conditions. Detection of crop diseases is one of the potential applications of remote sensing technology. These techniques may offer useful and timely information about the health of crops and also present the information which has previously been unavailable.

1.2 Problem statement

Grey leaf spot and Maize streak virus diseases of maize are being recognised as the most significant diseases that limit yields of maize globally and in South Africa. These diseases are not the only risk to maize crop production in commercial farms but also reduces yields in smallholder farms.

In Africa, maize diseases are rapidly spreading to most maize growing countries of sub-Saharan Africa. Being rapidly spreading diseases in the region, there is a need for a quick but effective control strategy. Since pathogen and viral populations cause widespread damages, it is logical that control strategies should target populations rather than individuals. Remote Sensing can target these populations and is cost effective. Although methods exist to detect these GLS and MSV diseases in maize, current field-based techniques (carried out by plant pathologists) are not effective in the early detection and quantification of the fungi and virus. Identifying plant diseases visually is expensive, inefficient, and difficult. Sustainable management of diseases helps the farmers to reduce the number of ancillary disease cycles and protect leaf area from damage before grain development. Early detection of GLS and MSV will reduce the widespread use of fungicides and chemical insecticides which might have adverse effects on the environment.

1.3 Aim and objectives

The purpose of this study was to examine the potential of field spectroscopy and new generation satellite images for detecting, identification, and mapping of MSV and GLS diseases of maize while the objectives were to:

- compare Random forest's forward variable selection and guided regularized random forest methods for optimum variable selection;
- detect the severity of maize streak virus infestations in maize hybrid lines using in-situ hyperspectral data;
- detect and map maize streak virus using RapidEye satellite imagery;
- Investigate the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations
- To determine environmental factors that explain the probability of maize streak virus disease occurrence; and
- Test the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot infection in maize hybrid line

1.4 Reliability, validity and objectivity

Joppe, (2000) defines reliability as: ...

The extent to which results are consistent over time and an accurate representation of the total population under study is referred to as reliability and if the results of a study can be reproduced under a similar methodology, then the research instrument is considered to be reliable.

Demonstrated in this research is the notion of repeatability of results or observations. Kirk and Miller, (1986) acknowledged three types of reliability referred to in quantitative research, which relate to (1) the degree to which a measurement, given repeatedly, remains the same (2) the stability of a measurement over time; and (3) the similarity of measurements within a given time period. In this study, reliability in several tests was ensured by using applicable levels of statistical significance for

discrimination and when calculating the agreement measured using the Kappa statistic.

Joppe, (2000) explains what validity is in research: Validity determines whether the research truly measures that which it was intended to measure or how truthful the research results are. In other words, does the research instrument allow you to hit "the bull's eye" of your research project?. Braun, (2009) defines the validity in research as "construct validity". The idea is the initial concept, notion, question or hypothesis that defines which data is to be collected and how it is to be collected. They also emphasise that quantitative researchers actively cause or affect the interplay between ideas and data in order to validate their investigation, usually by the use of a test or other process. Validity was maintained by conducting the experiment at the same location over two seasons. Objectivity is described as striving, as far as possible or practicable, to reduce or eliminate biases, prejudices or subjective evaluations by relying on verifiable data (Leedy and Ormrod, 2005). Objectivity is attained by deliberating the findings on the basis of pragmatic evidence as shown by statistical analyses, with findings compared and contrasted with findings in other studies (Hills, 1984).

1.5 Bias

Bias is defined as any influence, environments or set of situations that separately or completely distort the data collected (Leedy and Ormrod, 2005). In this study, bias was minimised by making sure that the error in each experiment was minimized through increased repetitions and random sampling in data collection (Hills, 1984).

1.6 Significance of the study

The outcomes of this study will contribute to a better understanding of maize reflectance properties during disease development. Methods will further be applicable in precision crop protection, to realize the early detection, mapping, and quantification of GLS and MSV. Knowledge generated by the project concerning the detection of GLS and MSV, together with existing literature, will be used to design,

develop and promote integrated disease management strategies for GLS and MSV. Currently, there is minimal work done on the detection, mapping, and quantification of GLS and MSV.

1.7 Format of the thesis

This thesis is consisting of six stand-alone articles, which have been submitted to internationally recognised peer-reviewed journals. Three of the manuscripts have already been published online, three under review. In this thesis, each article has been presented as a stand-alone chapter that can be read and reflected independently, from the entire dissertation, but it contributes to the overall introduction (Chapter One) and synthesis (Chapter Nine). It is also critical to note that the content of most of the manuscripts submitted to peer-reviewed journals has been recollected. This means that each of the stand-alone chapters consists of its abstract, conclusion, and references, which relate it to the following chapter, hence the presence of duplications and overlaps, particularly in the 'introduction' and 'methods' sections, of the various chapters in the Thesis. This duplication is assumed to be of little concern when considering that these are peer-reviewed scientific articles, which are stand-alone chapters that can be read separately, without losing the overall context. The entire thesis is made up of nine chapters.

1.7.1 Chapter One

This chapter serves as an introduction and a contextualization of the study. It highlights the importance of detection, identification, and mapping of maize streak virus and grey leaf spot diseases of maize using different remote sensing techniques. The chapter further highlights the importance of remote sensing as reliable tools and techniques that agronomists and plant pathologists can use to observe and inventory crop conditions. These techniques may provide useful and timely information about the health of crops and also makes available information which has previously been unavailable. In addition, the detailed research problem, aim, and objectives are outlined in this chapter.

1.7.2 Chapter Two

This chapter provides a detailed review of the literature. The review attempts to highlight the challenges and opportunities that are geared towards the detection and mapping of maize diseases. How remote sensing data was used and is currently being used in maize crop disease detection and mapping was reviewed, and gaps in literature were identified in order to assess the use of remote sensing data in maize crop diseases. The review focuses on how remote sensing is used in maize crop disease mapping and detection

1.7.3 Chapter Three

The chapter focuses on identifying techniques for the ‘curse of dimensionality’ reduction without sacrificing significant information which is critical in hyperspectral data processing and analysis. A comparison of two variable selection techniques Random Forest’s Forward Variable, (FVS) and Guided Regularized Random Forest: (GRRF) was done in selecting the optimal variables that can be used in detecting maize streak virus disease using in-situ resampled hyperspectral data.

1.7.4 Chapter Four

In this chapter, the utility of remote sensing in detecting the geminivirus infected maize was assessed based on experiments in Ofcolaco, Tzaneen in South Africa. Specifically, the potential of hyperspectral data in detecting different levels of MSV in maize was tested based on Guided Regularized Random Forest (GRRF). This study underscores the potential of using remotely sensed data in the accurate detection of maize crop diseases such as MSV and its severity which is critical in crop monitoring to foster food security especially in the resource-limited sub-Saharan Africa

1.7.5 Chapter Five

Tests were then conducted on the RapidEye data and derived vegetation indices in detecting and mapping the maize streak virus. RapidEye image was classified using the robust Random Forest algorithm to detect and map maize streak virus in Ofcolaco farm. Variable importance technique was used to determine the influence of each spectral band and indices on the mapping accuracy. For better performance of image data, commonly used vegetation indices were tested if they can significantly improve the classification accuracy. The study recommends future studies to evaluate the importance of multi-temporal remote sensing applications in detecting and monitoring the spread of maize streak virus

1.7.6 Chapter Six

Due to the cost involved in acquiring commercial images, like RapidEye, a freely available Landsat-8 OLI data can offer a new data source that is useful for maize diseases estimation, in environments which have limited resources. The study investigated the use of Landsat-8 OLI data and vegetation indices in estimating and predicting maize infected with maize streak virus. The purpose of this chapter was to assess the utility of the medium-resolution multispectral Landsat-8 OLI data set, in detecting and mapping maize streak virus disease at Ofcolaco farms in Tzaneen, South Africa. More specifically, the study sought to spectrally discriminate and map maize infected with maize streak virus from other land-cover classes using Landsat-8 OLI data.

1.7.7 Chapter Seven

In this study, the researcher aimed at examining the influence of climatic, environmental and remotely sensed variables on the spread of MSV disease on maize farms at the Ofcolaco farms in Tzaneen, South Africa. Environmental and climatic variables together with Landsat-8 OLI derived vegetation indices were integrated to predict the probability of MSV occurrence at Ofcolaco maize farms in Limpopo, South Africa. Correlation analysis was used to relate vegetation indices, environmental and climatic variables to incidences of maize streak virus disease. The variables used to predict the distribution of MSV were elevation, rainfall, slope,

temperature, and vegetation indices. The results indicated the potential of integrating vegetation indices derived from Landsat-8 OLI, environmental and climatic variables to improve the prediction of areas that are likely to be affected by MSV disease outbreaks in maize fields in semi-arid environments.

1.7.8 Chapter Eight

After realizing the potential of remote sensing in detecting and predicting the occurrence of maize streak virus disease, the study further examined its potential in mapping the most complex disease; Grey Leaf Spot (GLS) in maize fields using WorldView-2, Quickbird, RapidEye, and Sentinel-2 resampled from hyperspectral data. To accomplish this objective, field spectra were acquired from healthy, moderate and severely infected maize leaves during the 2013 and 2014 growing seasons. The spectra were then resampled to four sensor spectral resolutions – namely WorldView-2, Quickbird, RapidEye, and Sentinel-2. In each case, classification of the 2013 resampled spectra was done using Random forest algorithm to represent the three identified disease severity categories. Classification accuracy was evaluated using an independent test dataset obtained during the 2014 growing season.

1.7.9 Chapter Nine

This chapter shows a combination of the findings and conclusions drawn, based on the preceding chapters. It makes further recommendations for future research, based on the highlighted limitations of this study. Overall, the results imply that opportunities exist for developing operational remote sensing systems for detection of maize disease. Adoption of such remote sensing techniques is particularly valuable for minimizing crop damage, improving yield and ensuring food security.

1.8 References

- AGRICULTURE, S. A. D. O. & STOREY, H. 1925. *Streak disease of sugar-cane*.
- ARCHETTI, M., D RING, T. F., HAGEN, S. B., HUGHES, N. M., LEATHER, S. R., LEE, D. W., LEV-YADUN, S., MANETAS, Y., OUGHAM, H. J. & SCHABERG, P. G. 2009. Unravelling the evolution of autumn colours: an interdisciplinary approach. *Trends in ecology & evolution*, 24, 166-173.
- BECKMAN, P. M. & PAYNE, G. A. 1982. External growth, penetration, and development of *Cercospora zeae-maydis* in corn leaves. *Phytopathology*, 72, 810-815.
- BENHIN, J. K. A. 2008. South African crop farming and climate change: An economic assessment of impacts. *Global Environmental Change*, 18, 666-678.
- BOCK, K., GUTHRIE, E., MEREDITH, G. & BARKER, H. 1977. RNA and protein components of maize streak and cassava latent viruses. *Annals of applied Biology*, 85, 305-308.
- BOULTON, M. I., PALLAGHY, C. K., CHATANI, M., MACFARLANE, S. & DAVIES, J. W. 1993. Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology*, 192, 85-93.
- BRAUN, J. H. 2009. *A study of the reliability and validity of the standardized exams used in grade 10 science in a Winnipeg school division*. University of Manitoba.
- CHRISTOU, P. & TWYMAN, R. M. 2004. The potential of genetically enhanced plants to address food insecurity. *Nutrition research reviews*, 17, 23-42.
- DEGEFU, Y., LOHTANDER, K. & PAULIN, L. 2004. Expression patterns and phylogenetic analysis of two xylanase genes (htxyl1 and htlyl2) from *Helminthosporium turcicum*, the cause of northern leaf blight of maize. *Biochimie*, 86, 83-90.
- DERERA, J., TONGOONA, P., PIXLEY, K. V., VIVEK, B., LAING, M. D. & VAN RIJ, N. C. 2008. Gene action controlling gray leaf spot resistance in Southern African maize germplasm. *Crop Science*, 48, 93-98.
- FAJEMISIN, J., KIM, S., EFRON, Y. & ALAM, M. 1984. Breeding for durable disease resistance in tropical maize with special reference to maize streak virus. *FAO Plant Production and Protection Paper*, 55, 49-71.

- FULLER, C. 1901. Mealic variegation. *Mealic variegation.*, 17-19.
- HILLS, J. R. 1984. Quantitative Methods Used in the Study of Item Bias.
- JOPPE, M. 2000. The Research Process. Retrieved February 25, 1998.
- KAGODA, F., DERERA, J., TONGOONA, P., COYNE, D. L. & TALWANA, H. L. 2011. Grain yield and heterosis of maize hybrids under nematode infested and nematicide treated conditions. *Journal of nematology*, 43, 209.
- KANG, Y., KHAN, S. & MA, X. 2009. Climate change impacts on crop yield, crop water productivity and food security – A review. *Progress in Natural Science*, 19, 1665-1674.
- KIRK, J. & MILLER, M. L. 1986. *Reliability and validity in qualitative research*, Sage.
- KONATE, G. & TRAORE, O. 1992. Reservoir hosts of maize streak virus (MSV) in the Sudan-Sahel zone: Identification and spatio-temporal distribution. *Phytoprotection*, 73, 111-117.
- LANGYINTUO, A. S., MWANGI, W., DIALLO, A. O., MACROBERT, J., DIXON, J. & BÄNZIGER, M. 2010. Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private–public intervention to promote growth. *Food Policy*, 35, 323-331.
- LAPIERRE, H. & SIGNORET, P.-A. 2004. *Viruses and virus diseases of Poaceae (Gramineae)*, Editions Quae.
- LEEDY, P. D. & ORMROD, J. E. 2005. *Practical research*, Pearson Custom.
- LYIMO, H. J. F., PRATT, R. C. & MNYUKU, R. S. O. W. 2012. Composted cattle and poultry manures provide excellent fertility and improved management of gray leaf spot in maize. *Field Crops Research*, 126, 97-103.
- MOELETSI, M. E. & WALKER, S. 2012. A simple agroclimatic index to delineate suitable growing areas for rainfed maize production in the Free State Province of South Africa. *Agricultural and Forest Meteorology*, 162–163, 63-70.
- MZIRA, C. 1984. Assessment of effects of maize streak virus on yield of maize. *Zimbabwe Journal of Agricultural Research*, 22, 141-149.
- NUTTER JR, F. W. & SCHULTZ, P. M. 1995. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Canadian Journal of Plant Pathology*, 17, 174-184.
- RAIKES, C. & BURPEE, L. 1998. Use of multispectral radiometry for assessment of Rhizoctonia blight in creeping bentgrass. *Phytopathology*, 88, 446-449.

- REDINBAUGH, M. 2003. Transmission of Maize streak virus by vascular puncture inoculation with unit-length genomic DNA. *Journal of virological methods*, 109, 95-98.
- REDINBAUGH, M. G. & ZAMBRANO-MENDOZA, J. 2014. Control of virus diseases in maize. *Adv. Virus Res*, 90, 391-429.
- SHEPHERD, D. N., MARTIN, D. P., VAN DER WALT, E., DENT, K., VARSANI, A. & RYBICKI, E. P. 2010. Maize streak virus: an old and complex 'emerging' pathogen. *Molecular plant pathology*, 11, 1-12.
- SIBIYA, J., TONGOONA, P., DERERA, J., VAN RIJ, N. & MAKANDA, I. 2011. Combining ability analysis for Phaeosphaeria leaf spot resistance and grain yield in tropical advanced maize inbred lines. *Field Crops Research*, 120, 86-93.
- STRANGE, R. N. & SCOTT, P. R. 2005. Plant disease: a threat to global food security. *Annual review of phytopathology*, 43.
- THOTTAPPILLY, G., BOSQUE-PÉREZ, N. & ROSSEL, H. 1993. Viruses and virus diseases of maize in tropical Africa. *Plant Pathology*, 42, 494-509.
- VAN RENSBURG, G. 1981. Effect of plant age at the time of infection with maize streak virus on yield of maize. *Phytophylactica*.
- VOGEL, W., HENNESSEY, R., BERHE, T. & MATUNGULU, K. 1993. Yield losses to maize streak disease and *Busseola fusca* (Lepidoptera: Noctuidae), and economic benefits of streak-resistant maize to small farmers in Zaïre. *International journal of pest management*, 39, 229-238.
- WALKER, N. J. & SCHULZE, R. E. 2006. An assessment of sustainable maize production under different management and climate scenarios for smallholder agro-ecosystems in KwaZulu-Natal, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, 31, 995-1002.
- WARD, J. M. J., LAING, M. D. & NOWELL, D. C. 1997. Chemical control of maize grey leaf spot. *Crop Protection*, 16, 265-271.
- WARD, J. M. J., STROMBERG, E. L., NOWELL, D. C. & NUTTER JR, F. W. 1999. Gray leaf spot: a disease of global importance in maize production. *Plant disease*, 83, 884-895.
- WEGARY, D., HABTAMU, Z., SINGH, H. & HUSIEN, T. 2003. Inheritance of grey leaf spot resistance in selected maize inbred lines. *Afr. Plant Prot*, 9, 53-54.

WELZ, H. & GEIGER, H. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant breeding*, 119, 1-14.

Chapter 2

LITERATURE REVIEW

Remote sensing of maize diseases

Progress in the application of remote sensing in detecting and mapping of maize diseases

Abstract

Timely, and up to date information which is accurate for the detection and mapping of maize crop disease can be obtained by remote sensing techniques. In this study, the use of remote sensing techniques to maize crop disease detection and mapping was reviewed. How remote sensing data was used and is currently being used in maize crop disease detection and mapping was reviewed, and gaps in literature were identified in order to assess how remote sensing data can be used in maize crop diseases. Remote sensing can provide information about how maize leaves reflect at different stages of infection. Most of the studies that have been undertaken focus on disease classification, areal extent mapping, and crop health monitoring. Few studies have focused on the application of remote sensing on maize crop diseases in different environments. However remote sensing can be extremely useful in maize crop disease monitoring only if appropriate spatial and spectral resolution data is identified.

2.0 Introduction

Maize is an essential food for many people around the world and mainly in sub-Saharan Africa (FAO 2004). In Europe and USA, maize is mainly used as animal feed. Maize (*Zea mays L.*) is considered to be the most vital grain crop in South Africa. Nearly 8.0 million tonnes of maize grain is produced in South Africa every year on a land which is approximately 3.1 million hectares. Almost half of the maize is white maize and is used for human consumption. Furthermore, 50% of maize grain in the Southern African Development Community (SADC) is produced in South Africa (Benhin, 2008). Therefore, maize is considered to be the subsequent largest crop that is produced in South Africa after sugar cane.

It is of paramount importance to detect and map crop diseases in order to plan effective control measures so as to increase crop yields and strengthen the country's food security issues. Remote sensing can be used to detect and map crop diseases that are found in fields since crops infested by a disease, can show variations in the biochemical constituents of the plant and also the spectral signatures. Most of the prevalent maize crop diseases includes leaf blight caused by *Exserohilum turcicum*, common rust-induced by *Puccinia sorghi*, grey leaf spot (GLS) and maize streak virus caused by *Cercospora zea maydis*, ear rots caused by *Fusarium* and *Diplodia*, head smut caused by *Sporisorium reilianum* and *Phaeosphaeria* leaf spot (PLS) caused by *Phaeosphaeria maydis* have negative effect on maize crop production.

Turcicum leaf blight which is distributed worldwide, has also caused yield losses of more than 60% in vulnerable germplasm (Raymundo et al., 1981). Another disease which was reported to cause severe economic losses on approximately 34% of maize in subtropical and highland tropical maize fields is Grey leaf spot. Grey leaf spot has been reported as a disease of serious concern in the greater parts of the world and can account to almost 60% yield losses in South Africa (Ward et al., 1999) with projected losses going up to 100% under severe epidemics (Latterell and Rossi, 1983, McGee, 1988)

Other diseases of maize include maize streak virus (MSV) which was initially observed in East Africa and it is now found in many countries in sub-Saharan Africa, India, South Pacific, South east Asia and some of the islands in the Indian Ocean (Bonga, 1992). Losses of about 1.5% and total loss during severe MSV epidemics were also reported. *Phaeosphaeria* leaf spot (PLS) is a severe problem for maize in Brazil (Pegoraro et al., 2002) and has the potential to cause substantial yield losses in maize (Carson, 2005) and although it has not been reported in epidemic proportions in Africa. Incidences of PLS have been detected by researchers in Africa and it is becoming an important disease affecting maize yields.

Maize diseases are normally assessed using direct field survey methods like scouting and checking the plants for any damage symptoms. These methods are tedious, time-consuming, expensive and subjective as only a few sites within the fields are sampled (Al-Hiary et al., 2011, Pinter Jr et al., 2003). Therefore, corresponding disease monitoring methods that allow the implementation of site-specific practices are essential. In this setting, remote sensing is capable of offering synoptic, timely, accurate and relatively inexpensive data that can be utilized to provide an explicit overview of maize disease infestation. Also, remote sensing allows a wide-area monitoring approach thus less spotty and spatially more effective and comprehensible compared to point-specific field-based survey methods (Moran et al., 1997, Mulla, 2013). The objective of this research is to provide an overview of the progress of remote sensing in crop disease detection and mapping particularly focusing on maize crop diseases.

Significant progress has been made in remote sensing of crop diseases mainly using multispectral and machine learning systems. Classification of multispectral data were done using simple classes such as bare soil, water, and green vegetation. Handheld optical devices have also been used in the rapid detection of pest damage or disease (Mirik et al., 2007, Moshou et al., 2004, Xu et al., 2007) and airborne sensors (Sudbrink Jr et al., 2003). Since healthy and diseased plants reflect differently, researches tried to use reflectance measurements to distinguish between healthy and diseased plants in fields (Moshou et al., 2004). Cellular structure changes affect reflectance in the 0.75-to 1.35-micrometer wavelengths (Mirik et al., 2007). Changes in pigmentation affect reflectance in the 0.5-to 0.75-micrometer

range, and wavelengths between 1.35 and 2.5 micrometers are intensely influenced by the amount of water in the leaf (Carter and Spiering, 2002). The optical properties of leaves are categorised by light transmission through a leaf, light that is absorbed by leaf chemicals (pigments, lignin, and amino acids), and light that has been reflected from internal leaf structures or directly reflected from the leaf surface (Carter and Spiering, 2002). The reflectance of light from plants is a multifaceted issue that depends on several biophysical and biochemical interactions. The visible range (VIS 400 to 700 nm) is influenced by leaf pigment content, the near-infrared reflectance (NIR 700 to 1,100 nm) depends on the leaf structure, internal scattering processes, and on the absorption by leaf water, and the short-wave infrared (1,100 to 2,500nm) is influenced by the composition of leaf chemicals and water (Jacquemoud and Ustin, 2001). Therefore, changes in reflectance due to crop diseases can be attributed to damages in the structure of the leaf and chemical composition of the tissue during origin and development of the disease, like a succession of chlorotic tissue or the appearance of typical fungal and viral structures, such as powdery mildew hyphae and maize streak virus (Jacquemoud and Ustin, 2001).

2.1 Applications of remote sensing in monitoring crops

Remote sensing has been providing information about the status of vegetation looking at visible wavelengths (Price, 1992), active or passive microwaves (Prevot et al., 1993a, Prevot et al., 1993b) and emitted thermal wavelengths (Moran et al., 1994, Seguin et al., 1994, Seguin et al., 1991). Most researchers have so far concentrated on the use of the blue, green, red and near-infrared regions of the spectrum on agriculture (Gates et al., 1965, Woolley, 1971). A hand-held radiometer was used to collect the red (0.65-0.70 μm) and infrared (0.775-0.825 μm) spectral reflectance and were correlated with total aboveground winter wheat (*Triticum aestivum* L.) biomass accumulation over the growing season (Tucker et al., 1981). A high correlation between the spectral data and the vigour and state of the plant canopy was reported by the authors.

Remote sensing has been used widely in agricultural production in many aspects of crops (Adamchuk et al., 2004, Moran et al., 1997, Pinter Jr et al., 2003). Remote sensing has been applied in crop yield and biomass estimation (Shanahan et al., 2001, Yang et al., 2000), water stress and crop nutrient assessment (Clay et al., 2012, Cohen et al., 2005, Möller et al., 2006, Mulla, 2013), insects and plant diseases (Seelan et al., 2003), infestations of weeds (Lamb and Brown, 2001, Thorp and Tian, 2004), identification of crops (Foody et al., 1989, Saha and Jonna, 1994), detection of crop stresses (Carter et al., 1996, Carter and Spiering, 2002, Karimi et al., 2005, Lelong et al., 1998), detection of crop diseases (Lorenzen and Jensen, 1989, Peñuelas and Filella, 1998), weed detection (Brown et al., 1994, Everitt et al., 1996, Goel et al., 2003, Vrindts et al., 2002), yield estimation (Ferencz et al., 2004, Hamar et al., 1996) and precision farming (Moran et al., 1997, Pearson et al., 1994, Seelan et al., 2003, Wallace, 1994). From these studies, one can conclude that remote sensing has been employed widely in monitoring crop production.

Different crop species were classified using remote sensing data (Rao, 2008). Plant water stress, pests, crop weeds, and crop diseases were also monitored using remote sensing techniques. Crops such as corn, cotton, sorghum, wheat, and canola, have been monitored using different remote sensing techniques (Lelong et al., 1998, Yang, 2010, Zhao et al., 2007). Different remote sensing data like Aerial photography (Godwin and Miller, 2003), satellite hyperspectral (Rao, 2008), airborne hyperspectral (DeTar et al., 2008, Lelong et al., 1998), satellite multi-spectral (Clevers, 1997, Gomez et al., 2008) and close range hyperspectral techniques (Gomez et al., 2008, Rao, 2008, Zhao et al., 2007) have also been used to examine different field crop spectral responses. (Gallego, 1999) used satellite data to successfully estimate crop acreage, (Chen et al., 2007) successfully identified the infestations of the take-all disease (*Gaeumannomyces graminis*) in wheat using Landsat multispectral imagery. Franke and Menz (2007) assessed a QuickBird satellite multispectral imagery in detecting powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat. Results revealed that multispectral images are mostly suitable for detecting infield heterogeneities in wheat vigour, especially at later stages of fungal infections.

Different crop parameters such as crop productivity, yield, plant nutrient, type of plant species, and plant pathological status have been assessed using the blue, green, red and near infra-red regions of the electromagnetic spectrum. Reflectance measurements from different crops were used successfully to distinguish between healthy and diseased crops (Deering, 1989). Reflectance data was also used to detect *Magnaporthe grisea* on rice (Kobayashi et al., 2001), *Phytophthora infestans* on tomato (Zhang et al., 2003), and yellow rust in wheat (Huang et al., 2007). Genc et al., (2008) assessed the sunn pest damage to wheat using NDVI and structure insensitive index derived from the handheld radiometer. Furthermore, field and laboratory spectroscopy reflectance of healthy and infested canopies of mustard were compared. Vegetation indices (Normalised Difference Vegetation index (*NDVI*), and Structure Insensitive Pigment Index (*SIPi*)) were meaningfully correlated with aphid infestation and these indices can be used in identifying aphid infestation in mustard. Rumpf et al., (2010) assessed crops damaged by the virus using sensors. Most of these studies employed airborne spectral data for discrimination of mature disease symptoms and healthy leaves and hyperspectral data is also increasingly being used often in agricultural areas. Some studies have used hyperspectral imaging successfully for quality assessment of maize kernels, and pickling cucumbers, (Ariana and Lu, 2010, Larsen et al., 2009).

Riedell and Blackmer, (1999) utilized the leaf reflectance spectra obtained from a handheld radiometer in the greenhouse to portray wheat stressed by Russian wheat aphid. The specialists discovered that leaf reflectance in the 625– 635 nm and 680– 695 nm ranges, just as the standardized complete color to chlorophyll a, ratio index, were great indicators of chlorophyll loss brought about by aphid feeding. Yang et al., (2005) completed an examination on greenbug (*Schizaphis graminum* Rondani) stress in wheat that was developed in a greenhouse. Their outcomes uncovered a waveband focused at 694 nm and vegetation indices inferred utilizing 800 nm and 694 nm were most delicate to greenbug-harmed wheat. Ashourloo et al., (2014), created two indices: Leaf Rust Disease Severity Index 1 and 2 (LRDSI1 and LRDSI2) in view of the reflectance in the 605, 695 and 455 nm wavelengths. These vegetation indices were obtained from information acquired with a hyperspectral radiometer for identifying diseases of wheat leaf Rust (*Puccinia triticina*). The two indices had high R^2 with the disease severity (0.94 and 0.95, respectively).

2.3 Multi- and hyperspectral sensors.

Remote sensors are characterised by the spectral, spatial resolution and the type of detector utilized (imaging and non-imaging sensor frameworks). Multispectral sensors measure spectral data of features in several broad bands. Multispectral imaging cameras measures in the RGB wavebands and Near-Infrared band. Hyperspectral sensors estimates information in the 350 to 2500 nm spectral range with a narrow spectral resolution of 1 nm and below (Rumpf et al., 2010). Hyperspectral information are typically observed as huge matrices with spatial x-and y-axes, and the spectral data as reflectance intensity per waveband in the third measurement, (z). The spatial resolution relies upon the distance between the sensor and the feature under investigation. In this manner, airborne or spaceborne, far range systems have lower spatial resolution than close range or microscopic systems. The spatial resolution of a sensor has an impact on the detection of crop diseases (Mahlein et al., 2012, West et al., 2003). Airborne sensors proved to be valuable for the detection of field patches that are affected with soil-borne pathogens or in later stages of the diseases (Hillnhütter et al., 2011, Mahlein et al., 2012).

When utilizing airborne images to distinguish infected crops, it is imperative to choose a sensor with a suitable spatial and spectral resolution. For example, Mewes et al., (2011) looked at the viability of the identifying wheat plants infected with brown rust (*Puccinia recondita f. sp. tritici*) with two hyperspectral cameras (AISA-DUAL, Specim LTD, Oulu, Finland) which recorded the reflected radiation in the 498 channels in the range of 400 - 2500 nm with a spectral resolution of 2.5 - 5.8 nm and (ROSIS, German Space Agency, DLR) in the 115 directs in the range of 383 - 839 nm with a spectral resolution of 5 nm.

The accuracy with which healthy and unhealthy plants were recognized in the AISA-DUAL images was higher than in the ROSIS images (respectively 84.32% and 80.33%) and was related with stronger correlations at longer NIR wavelengths. AISA images were recorded from a lower elevation (2300m) with a higher spatial resolution of (1.5m) than ROSIS images (2880m) with a spatial resolution of 2.0m

and more grounded AISA signal intensity because of lower atmospheric absorption and scattering of the signal reflected from the field surface. Garcia-Ruiz et al., (2013) compared the value of citrus greening disease, (brought about by motile microbes) (*Candidatus Liberibacter spp*) detection utilising a UAV based sensor with a comparative imaging system mounted on a piloted airplane with spatial resolutions of 5.45 cm/pixel and 0.5 m/pixel, respectively. In general classification accuracy of 67–85% were accomplished using UAV.

Mahlein et al., (2013) showed the distinction of foliar pathogens of sugar beet using leaf reflectance. Rumpf et al., (2010) had the capacity to identify early *Cercospora* leaf spot, powdery mildew, and rust-infected sugar beets before the presence of visible symptoms. Non-invasive spectral data was also used to detect *Fusarium graminearum* in wheat (Bauriegel et al., 2011a, Bauriegel et al., 2011b), *Venturia inaequalis* in apple (Delalieux et al., 2007), and *Phytophthora infestans* in tomato (Wang et al., 2008). In proximal detecting, hyperspectral imaging also appeared to be helpful for the appraisal of mycotoxin-producing pathogens in maize (Del Fiore et al., 2010). Bravo et al., (2003) utilised in-field spectral images for the early detection of yellow rust infected wheat. Soil-borne infections were effectively separated by (Hillnhütter et al., 2011). Apan et al., (2004) detected orange rust using EO-1 Hyperion hyperspectral imaging. Huang et al., (2007) identified yellow rust in wheat by ground-based spectral measurements and airborne hyperspectral imaging. In addition to the detection of plant diseases, hyperspectral imaging is broadly used for monitoring fruit health and quality. Canker lesions of citrus fruits (Qin et al., 2009), apple surface imperfections (Mehl et al., 2004), or rot of strawberries (ElMasry et al., 2007) were detected by hyperspectral imaging sensors. These techniques are vital in screening fruits and crops to avoid diseases. Different types of crops have been monitored using different techniques of remote sensing, but only a couple of studies have used remote sensed data on maize crop in different environments (Adam et al., 2017, Dhau et al., 2017, Dhau et al., 2018).

2.4 Remote sensing of maize diseases

There are various studies that have utilised remote sensing data in detecting maize diseases (Adam et al., 2017, Dhau et al., 2017, Dhau et al., 2018, Bauer et al 1971, Carroll et al 2008). Schaafsma et al., (1993) precisely identified maize plots infested by corn rootworm (*Diabrotica virgifera*) using hyperspectral images. The overall classification accuracies for identification of insect infected plots were up to 99% and were more noteworthy on account of images recorded later in the season.

The maximum separability between healthy and infected maize was determined using Simple Ratio index determined as the proportion of two bands in VIS (648 nm) and NIR (747 nm) wavelengths. Aerial photography has been used to greatly in the investigation of mould and dwarf mosaic virus infected corn (Berger et al., 1989, Curran, 1985). Williams et al., (2012) assessed the capability of the hyperspectral near infrared (NIR) imaging to assess fungal contamination in maize kernels. Jingcheng Zhang, (2015) used satellite multispectral data for mapping of damage caused armyworm (*Spodoptera frugiperda*) in maize at a regional scale. The objectives of their study were to determine suitable spectral features for armyworm detection and to build up a mapping technique at a regional scale based on satellite remote sensing data. Their outcomes exhibited the credibility of the strategy and its promising potential for implementation in practice.

2.5 Challenges associated with remote sensing of maize diseases

In order to improve agricultural management in precision agriculture, there is a need to combine historical remote sensing data with real-time data, (Thenkabail, 2003). A noteworthy challenge of remote sensing applications in crop diseases is the capacity to separate spectral signals originating with a plant response to a specific disease from signals associated with normal plant biomass or the background “noise” that is introduced by exogenous non-plant factors. Results from several crops across a number of different locations indicate that general relationships between spectral properties and plant response are achievable (Wiegand et al., 1990, Wiegand et al., 1992, Richardson and Everitt, 1992). One of the significant

research challenges is to develop disease detection algorithms that perform reliably and precisely across space and time. Procedures ought to be independent of location, soils, and the board management factors. They ought to likewise work well all through the season, from planting through maturity. Current satellite-based sensors which are openly accessible or less expensive have fixed spectral bands that might be unsuitable for a given application, spatial resolutions too coarse for within field investigation, inadequate repeat coverage for intensive crop disease monitoring, and long time periods between image acquisition and delivery to the user.

A variety of studies have used hyperspectral data to model crop diseases. However, hyperspectral data are high dimensional, complex, expensive and their analysis is associated with a high computational cost. In addition, the narrow and contiguous wavebands of hyperspectral data are highly correlated and cause a collinearity problem when they are integrated into an empirical projecting model. Techniques for 'curse of dimensionality' reduction without sacrificing significant information are critical in hyperspectral data processing and analysis (Borges et al., 2007, Pal, 2005, Shaw and Manolakis, 2002). For the past few decades, significant improvements have been achieved in the adoption of efficient remote sensing approaches to detect and map maize diseases. However, there are still more issues to be explored as far as remote sensing of maize diseases is concerned. The challenge of disease infestation on maize production warrants more investigation as very little was found during this literature survey. Forthcoming research on remote sensing of maize diseases should focus more on the utilization of new generation satellite images at local and regional level.

2.6 Conclusions

One of the potential applications of remote sensing in agriculture is the detection of plant diseases in broad areas before the indications obviously show up on the plant leaves. This is profitable on the grounds that remote sensing distinguishes biophysical changes before physiological changes are visible. Early detection and delineation of maize infested areas, particularly in some of the high profitable zones that are prone to infections (e.g. Maize streak virus and grey leaf spot) using hyperspectral and new generation multispectral remotely data could be attempted. Therefore, spectroscopy analysis could be considered as an efficient technique for non-destructive, rapid, and accurate measurement which is widely applied in agricultural fields for crop discrimination and monitoring of diseases (Sankaran et al., 2010). The current study has reviewed previous studies on the prospective of remote sensing applications in maize diseases detection, identification, and mapping. Research has shown that the use of field surveys in detecting and mapping maize diseases, their spread, its life cycle remains a challenge in most parts of the world. Remote sensing technology hereby offers better assessments in detection and mapping of maize diseases.

The use of medium spatial resolution in the detection, identification, and mapping of maize diseases has been hampered by satellite sensor's limited spatial, spectral and radiometric resolution. In spite of the fact that the application of high spatial and spectral resolution sensors has accurately detected and mapped crop diseases at local scales, the application of these data sets is compromised by smaller swath width, low temporal resolution, and high acquisition costs. Challenges encountered *in* remote sensing of maize include the problem of similarity and multi-collinearity in spectral signatures of maize diseases leading to low classification accuracy. In future, some of these challenges can be minimized by the use of robust algorithms for classification such as Random Forest.

2.7 References

- ADAM, E., DENG, H., ODINDI, J., ABDEL-RAHMAN, E. M. & MUTANGA, O. 2017. Detecting the Early Stage of Phaeosphaeria Leaf Spot Infestations in Maize Crop Using In Situ Hyperspectral Data and Guided Regularized Random Forest Algorithm. *Journal of Spectroscopy*, 2017.
- ADAMCHUK, V. I., HUMMEL, J., MORGAN, M. & UPADHYAYA, S. 2004. On-the-go soil sensors for precision agriculture. *Computers and electronics in agriculture*, 44, 71-91.
- AL-HIARY, H., BANI-AHMAD, S., REYALAT, M., BRAIK, M. & ALRAHAMNEH, Z. 2011. Fast and accurate detection and classification of plant diseases. *Machine learning*, 14.
- APAN, A., HELD, A., PHINN, S. & MARKLEY, J. 2004. Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. *International journal of remote sensing*, 25, 489-498.
- ARIANA, D. P. & LU, R. 2010. Hyperspectral waveband selection for internal defect detection of pickling cucumbers and whole pickles. *Computers and Electronics in Agriculture*, 74, 137-144.
- ASHOORLOO, D., MOBASHERI, M. R. & HUETE, A. 2014. Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements. *Remote Sensing*, 6, 5107-5123.
- BAURIEGEL, E., GIEBEL, A., GEYER, M., SCHMIDT, U. & HERPPICH, W. 2011a. Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture*, 75, 304-312.
- BAURIEGEL, E., GIEBEL, A. & HERPPICH, W. B. 2011b. Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of Fusarium culmorum on the photosynthetic integrity of infected wheat ears. *Sensors*, 11, 3765-3779.
- BENHIN, J. K. A. 2008. South African crop farming and climate change: An economic assessment of impacts. *Global Environmental Change*, 18, 666-678.
- BERGER, P., LUCIANO, C., THORNBURY, D., BENNER, H., HILL, J. & ZEYEN, R. 1989. Properties and in vitro translation of maize dwarf mosaic virus RNA. *Journal of general virology*, 70, 1845-1851.
- BOCK, C., POOLE, G., PARKER, P. & GOTTWALD, T. 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29, 59-107.
- BONGA, J. 1992. *Serological Differentiation of Maize-infecting Maize Streak Virus (MSV) Isolates from Different Locations in Africa and Epitope Characterization of the MSV Coat Protein*. Ohio State University.
- BORGES, J. S., MARCAL, A. R. S. & DIAS, J. M. B. Evaluation of feature extraction and reduction methods for hyperspectral images. *In: BOCHENEK, Z., ed. New Developments and Challenges in Remote Sensing*, 29 May- 2 June 2006 2007 Poland. 255-264.
- BRAVO, C., MOSHOU, D., WEST, J., MCCARTNEY, A. & RAMON, H. 2003. Early disease detection in wheat fields using spectral reflectance. *Biosystems Engineering*, 84, 137-145.
- BROWN, R., STECKLER, J.-P. & ANDERSON, G. 1994. Remote sensing for identification of weeds in no-till corn. *Transactions of the ASAE*, 37, 297-302.

- CARSON, M. 2005. Yield loss potential of Phaeosphaeria leaf spot of maize caused by Phaeosphaeria maydis in the United States. *Plant disease*, 89, 986-988.
- CARTER, G. A., CIBULA, W. G. & MILLER, R. L. 1996. Narrow-band Reflectance Imagery Compared with Thermal Imagery for Early Detection of Plant Stress. *Journal of Plant Physiology*, 148, 515-522.
- CARTER, G. A. & SPIERING, B. A. 2002. Optical properties of intact leaves for estimating chlorophyll concentration. *Journal of environmental quality*, 31, 1424-1432.
- CHEN, X., MA, J., QIAO, H., CHENG, D., XU, Y. & ZHAO, Y. 2007. Detecting infestation of take-all disease in wheat using Landsat Thematic Mapper imagery. *International Journal of Remote Sensing*, 28, 5183-5189.
- CHRISTY, C. D. 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and electronics in agriculture*, 61, 10-19.
- CLAY, D. E., KHAREL, T. P., REESE, C., BECK, D., CARLSON, C. G., CLAY, S. A. & REICKS, G. 2012. Winter wheat crop reflectance and nitrogen sufficiency index values are influenced by nitrogen and water stress. *Agronomy Journal*, 104, 1612-1617.
- CLEVERS, J. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment*, 61, 221-228.
- COHEN, Y., ALCHANATIS, V., MERON, M., SARANGA, Y. & TSIPRIS, J. 2005. Estimation of leaf water potential by thermal imagery and spatial analysis. *Journal of experimental botany*, 56, 1843-1852.
- CURRAN, P. J. 1985. Aerial photography for the assessment of crop condition: a review. *Applied Geography*, 5, 347-360.
- DEERING, D. W. 1989. Field measurements of bidirectional reflectance. *Theory and applications of optical remote sensing*, 14, 65.
- DEL FIORE, A., REVERBERI, M., RICELLI, A., PINZARI, F., SERRANTI, S., FABBRI, A., BONIFAZI, G. & FANELLI, C. 2010. Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *International journal of food microbiology*, 144, 64-71.
- DELALIEUX, S., VAN AARDT, J., KEULEMANS, W., SCHREVEENS, E. & COPPIN, P. 2007. Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. *European Journal of Agronomy*, 27, 130-143.
- DETAR, W. R., CHESSON, J. H., PENNER, J. V. & OJALA, J. C. 2008. Detection of soil properties with airborne hyperspectral measurements of bare fields. *Transactions of the ASABE*, 51, 463-470.
- DHAU, I., ADAM, E., MUTANGA, O., AYISI, K., ABDEL-RAHMAN, E. M., ODINDI, J. & MASOCHA, M. 2017. Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop. *Geocarto International*, 1-28.
- DHAU, I., ADAM, E., MUTANGA, O. & AYISI, K. K. 2018. Detecting the severity of maize streak virus infestations in maize crop using in situ hyperspectral data. *Transactions of the Royal Society of South Africa*, 73, 8-15.
- ELMASRY, G., WANG, N., ELSAYED, A. & NGADI, M. 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81, 98-107.

- EVERITT, J. H., ESCOBAR, D. E., ALANIZ, M. A., DAVIS, M. R. & RICHERSON, J. V. 1996. Using spatial information technologies to map Chinese tamarisk (*Tamarix chinensis*) infestations. *Weed Science*, 194-201.
- FERENCZ, C., BOGNAR, P., LICHTENBERGER, J., HAMAR, D., TARCSAI, G., TIMAR, G., MOLNÁR, G., PÁSZTOR, S., STEINBACH, P. & SZEKELY, B. 2004. Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 25, 4113-4149.
- FOODY, G., CURRAN, P., GROOM, G. & MUNRO, D. 1989. Multi-temporal airborne synthetic aperture radar data for crop classification. *Geocarto International*, 4, 19-29.
- FRANKE, J. & MENZ, G. 2007. Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture*, 8, 161-172.
- GALLEGO, F. Crop area estimation in the MARS project. Conference on ten years of the MARS Project, 1999.
- GARCIA-RUIZ, F., SANKARAN, S., MAJA, J. M., LEE, W. S., RASMUSSEN, J. & EHSANI, R. 2013. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, 91, 106-115.
- GATES, D. M., KEEGAN, H. J., SCHLETER, J. C. & WEIDNER, V. R. 1965. Spectral properties of plants. *Applied optics*, 4, 11-20.
- GENC, H., GENC, L., TURHAN, H., SMITH, S. & NATION, J. 2008. Vegetation indices as indicators of damage by the sunn pest (Hemiptera: Scutelleridae) to field grown wheat. *African Journal of Biotechnology*, 7.
- GODWIN, R. & MILLER, P. 2003. A review of the technologies for mapping within-field variability. *Biosystems engineering*, 84, 393-407.
- GOEL, P. K., PRASHER, S. O., LANDRY, J.-A., PATEL, R. M., BONNELL, R., VIAU, A. A. & MILLER, J. 2003. Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn. *Computers and electronics in agriculture*, 38, 99-124.
- GOMEZ, C., ROSSEL, R. A. V. & MCBRATNEY, A. B. 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, 146, 403-411.
- HAMAR, D., FERENCZ, C., LICHTENBERGER, J., TARCSAI, G. & FERENCZ-ÁRKOS, I. 1996. Yield estimation for corn and wheat in the Hungarian Great Plain using Landsat MSS data. *International Journal of Remote Sensing*, 17, 1689-1699.
- HILLNHÜTTER, C., MAHLEIN, A.-K., SIKORA, R. & OERKE, E.-C. 2011. Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields. *Field Crops Research*, 122, 70-77.
- HUANG, W., LAMB, D. W., NIU, Z., ZHANG, Y., LIU, L. & WANG, J. 2007. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agriculture*, 8, 187-197.
- JACQUEMOUD, S. & USTIN, S. L. Leaf optical properties: A state of the art. 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, 2001. 223-332.
- KARIMI, Y., PRASHER, S., MCNAIRN, H., BONNELL, R., DUTILLEUL, P. & GOEL, P. 2005. Classification accuracy of discriminant analysis, artificial neural networks, and decision trees for weed and nitrogen stress detection in corn. *Transactions of the ASAE*, 48, 1261-1268.

- KOBAYASHI, T., KANDA, E., KITADA, K., ISHIGURO, K. & TORIGOE, Y. 2001. Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. *Phytopathology*, 91, 316-323.
- LAMB, D. & BROWN, R. B. 2001. Pa—precision agriculture: Remote-sensing and mapping of weeds in crops. *Journal of Agricultural Engineering Research*, 78, 117-125.
- LARSEN, R., ARNGREN, M., HANSEN, P. W. & NIELSEN, A. A. Kernel based subspace projection of near infrared hyperspectral images of maize kernels. Scandinavian Conference on Image Analysis, 2009. Springer, 560-569.
- LATTERELL, F. M. & ROSSI, A. E. 1983. Gray leaf spot of corn: a disease on the move. *Plant Disease*, 67, 842-847.
- LELONG, C. C., PINET, P. C. & POILVÉ, H. 1998. Hyperspectral imaging and stress mapping in agriculture: a case study on wheat in Beauce (France). *Remote sensing of environment*, 66, 179-191.
- LOBELL, D., LESCH, S., CORWIN, D., ULMER, M., ANDERSON, K., POTTS, D., DOOLITTLE, J., MATOS, M. & BALTES, M. 2010. Regional-scale assessment of soil salinity in the Red River Valley using multi-year MODIS EVI and NDVI. *Journal of environmental quality*, 39, 35.
- LORENZEN, B. & JENSEN, A. 1989. Changes in leaf spectral properties induced in barley by cereal powdery mildew. *Remote Sensing of Environment*, 27, 201-209.
- MAHLEIN, A.-K., OERKE, E.-C., STEINER, U. & DEHNE, H.-W. 2012. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, 133, 197-209.
- MAHLEIN, A.-K., RUMPF, T., WELKE, P., DEHNE, H.-W., PLÜMER, L., STEINER, U. & OERKE, E.-C. 2013. Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, 128, 21-30.
- MCGEE, D. Seedborne and seed-transmitted diseases of maize in rice-based cropping systems. Rice Seed Health: Proceedings of the International Workshop on Rice Seed Health, 16-20 March 1987, 1988. Int. Rice Res. Inst., 203.
- MEHL, P. M., CHEN, Y.-R., KIM, M. S. & CHAN, D. E. 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*, 61, 67-81.
- MEWES, T., FRANKE, J. & MENZ, G. 2011. Spectral requirements on airborne hyperspectral remote sensing data for wheat disease detection. *Precision Agriculture*, 12, 795.
- MIRIK, M., MICHELS, G., KASSYMZHANOVA-MIRIK, S. & ELLIOTT, N. 2007. Reflectance characteristics of Russian wheat aphid (Hemiptera: Aphididae) stress and abundance in winter wheat. *Computers and Electronics in Agriculture*, 57, 123-134.
- MÖLLER, M., ALCHANATIS, V., COHEN, Y., MERON, M., TSIPRIS, J., NAOR, A., OSTROVSKY, V., SPRINTSIN, M. & COHEN, S. 2006. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *Journal of experimental botany*, 58, 827-838.
- MORAN, M., CLARKE, T., INOUE, Y. & VIDAL, A. 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote sensing of environment*, 49, 246-263.

- MORAN, M. S., INOUE, Y. & BARNES, E. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote sensing of Environment*, 61, 319-346.
- MOSHOU, D., BRAVO, C., WEST, J., WAHLEN, S., MCCARTNEY, A. & RAMON, H. 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and electronics in agriculture*, 44, 173-188.
- MULLA, D. J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering*, 114, 358-371.
- PAL, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222.
- PEARSON, R., GRACE, J. & MAY, G. 1994. Real-time airborne agricultural monitoring. *Remote Sensing of Environment*, 49, 304-310.
- PEGORARO, D. G., BARBOSA NETO, J. F., SOGLIO, D., KESSLER, F., VACARO, E., NUSS, C. N. & CONCEIÇÃO, L. D. H. 2002. Inheritance of the resistance to phaeosphaeria leaf spot in maize. *Pesquisa Agropecuária Brasileira*, 37, 329-336.
- PEÑUELAS, J. & FILELLA, I. 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in plant science*, 3, 151-156.
- PINTER JR, P. J., HATFIELD, J. L., SCHEPERS, J. S., BARNES, E. M., MORAN, M. S., DAUGHTRY, C. S. & UPCHURCH, D. R. 2003. Remote sensing for crop management. *Photogrammetric Engineering & Remote Sensing*, 69, 647-664.
- PREVOT, L., CHAMPION, I. & GUYOT, G. 1993a. Estimating surface soil moisture and leaf area index of a wheat canopy using a dual-frequency (C and X bands) scatterometer. *Remote Sensing of Environment*, 46, 331-339.
- PREVOT, L., DECHAMBRE, M., TACONET, O., VIDAL-MADJAR, D., NORMAND, M. & GALLEJ, S. 1993b. Estimating the characteristics of vegetation canopies with airborne radar measurements. *International Journal of Remote Sensing*, 14, 2803-2818.
- PRICE, J. C. 1992. Estimating vegetation amount from visible and near infrared reflectances. *Remote Sensing of Environment*, 41, 29-34.
- QIN, J., BURKS, T. F., RITENOUR, M. A. & BONN, W. G. 2009. Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. *Journal of food engineering*, 93, 183-191.
- RAO, N. R. 2008. Development of a crop-specific spectral library and discrimination of various agricultural crop varieties using hyperspectral imagery. *International Journal of Remote Sensing*, 29, 131-144.
- RAYMUNDO, A., HOOKER, A. & PERKINS, J. 1981. Effect of gene HtN on the development of northern corn leaf blight epidemics. *Plant disease*.
- RICHARDSON, A. J. & EVERITT, J. H. 1992. Using spectral vegetation indices to estimate rangeland productivity. *Geocarto International*, 7, 63-69.
- RIEDEL, W. E. & BLACKMER, T. M. 1999. Leaf reflectance spectra of cereal aphid-damaged wheat. *Crop Science*, 39, 1835-1840.
- RUMPF, T., MAHLEIN, A.-K., STEINER, U., OERKE, E.-C., DEHNE, H.-W. & PLÜMER, L. 2010. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74, 91-99.

- SAHA, S. & JONNA, S. 1994. Paddy acreage and yield estimation and irrigated crop land inventory using satellite and agrometeorological data. *Asian-Pacific Journal of Remote Sensing*, 6, 79-87.
- SANKARAN, S., MISHRA, A., EHSANI, R. & DAVIS, C. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, 72, 1-13.
- SCHAAFSMA, A., WHITFIELD, G., GILLESPIE, T. & ELLIS, C. 1993. Evaluation of infrared thermometry as a non-destructive method to detect feeding on corn roots by the western corn rootworm (Coleoptera: Chrysomelidae). *The Canadian Entomologist*, 125, 643-655.
- SEELAN, S. K., LAGUETTE, S., CASADY, G. M. & SEIELSTAD, G. A. 2003. Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment*, 88, 157-169.
- SEGUIN, B., COURAULT, D. & GUERIF, M. 1994. Surface temperature and evapotranspiration: Application of local scale methods to regional scales using satellite data. *Remote Sensing of Environment*, 49, 287-295.
- SEGUIN, B., LAGOUARDE, J.-P. & SAVANE, M. 1991. The assessment of regional crop water conditions from meteorological satellite thermal infrared data. *Remote sensing of environment*, 35, 141-148.
- SHANAHAN, J. F., SCHEPERS, J. S., FRANCIS, D. D., VARVEL, G. E., WILHELM, W. W., TRINGE, J. M., SCHLEMMER, M. R. & MAJOR, D. J. 2001. Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, 93, 583-589.
- SHAW, G. & MANOLAKIS, D. 2002. Signal processing for hyperspectral image exploitation. *IEEE Signal Processing Magazine*, 19, 12-16.
- SUDBRINK JR, D., HARRIS, F., ROBBINS, J., ENGLISH, P. & WILLERS, J. 2003. Evaluation of remote sensing to identify variability in cotton plant growth and correlation with larval densities of beet armyworm and cabbage looper (Lepidoptera: Noctuidae). *Florida Entomologist*, 86, 290-294.
- THENKABAIL, P. 2003. Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. *International Journal of Remote Sensing*, 24, 2879-2904.
- THORP, K. & TIAN, L. 2004. A review on remote sensing of weeds in agriculture. *Precision Agriculture*, 5, 477-508.
- TUCKER, C. J., HOLBEN, B. N., ELGIN, J. H. & MCMURTREY, J. E. 1981. Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sensing of Environment*, 11, 171-189.
- VRINDTS, E., DE BAERDEMAEKER, J. & RAMON, H. 2002. Weed detection using canopy reflection. *Precision Agriculture*, 3, 63-80.
- WALLACE, A. 1994. High-precision agriculture is an excellent tool for conservation of natural resources. *Communications in Soil Science & Plant Analysis*, 25, 45-49.
- WANG, X., ZHANG, M., ZHU, J. & GENG, S. 2008. Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN). *International Journal of Remote Sensing*, 29, 1693-1706.
- WARD, J. M., STROMBERG, E. L., NOWELL, D. C. & NUTTER JR, F. W. 1999. Gray leaf spot: a disease of global importance in maize production. *Plant disease*, 83, 884-895.
- WEST, J. S., BRAVO, C., OBERTI, R., LEMAIRE, D., MOSHOU, D. & MCCARTNEY, H. A. 2003. The potential of optical canopy measurement for

- targeted control of field crop diseases. *Annual review of Phytopathology*, 41, 593-614.
- WIEGAND, C., GERBERMANN, A., GALLO, K., BLAD, B. & DUSEK, D. 1990. Multisite analyses of spectral-biophysical data for corn. *Remote Sensing of Environment*, 33, 1-16.
- WIEGAND, C., MAAS, S., AASE, J., HATFIELD, J., PINTER, P. J., JACKSON, R., KANEMASU, E. & LAPITAN, R. 1992. Multisite analyses of spectral-biophysical data for wheat. *Remote Sensing of Environment*, 42, 1-21.
- WILLIAMS, P. J., GELADI, P., BRITZ, T. J. & MANLEY, M. 2012. Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, 55, 272-278.
- WOOLLEY, J. T. 1971. Reflectance and transmittance of light by leaves. *Plant physiology*, 47, 656-662.
- XU, H., YING, Y., FU, X. & ZHU, S. 2007. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosystems Engineering*, 96, 447-454.
- YANG, C.-M. 2010. Assessment of the severity of bacterial leaf blight in rice using canopy hyperspectral reflectance. *Precision Agriculture*, 11, 61-81.
- YANG, C.-M., CHENG, C.-H. & CHEN, R.-K. 2007. Changes in spectral characteristics of rice canopy infested with brown planthopper and leaf folder. *Crop science*, 47, 329-335.
- YANG, C., EVERITT, J., BRADFORD, J. & ESCOBAR, D. 2000. Mapping grain sorghum growth and yield variations using airborne multispectral digital imagery. *Transactions of the ASAE*, 43, 1927.
- YANG, Z., RAO, M., ELLIOTT, N., KINDLER, S. & POPHAM, T. 2005. Using ground-based multispectral radiometry to detect stress in wheat caused by greenbug (Homoptera: Aphididae) infestation. *Computers and electronics in agriculture*, 47, 121-135.
- ZHANG, M., QIN, Z., LIU, X. & USTIN, S. L. 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 4, 295-310.
- ZHAO, D., HUANG, L., LI, J. & QI, J. 2007. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 25-33.

Chapter 3

Dimensionality reduction using variable selection methods

This chapter is based on:

Inos Dhau, Elhadi Adam, Onesimo Mutanga & Kingsley K. Ayisi (Under Review)
Comparative analysis of Random forest forward variable selection and guided regularized random forest techniques for optimum band selection. *Transactions of the Royal Society of South Africa*,

Abstract

Random forest's forward variable selection method was compared with guided regularized random forest in selecting the optimum variables using maize streak virus dataset. In order to provide a robust indicator of comparative error, the OOB error was reported. The sensitivity of regularized random forest to specific hyper-parameters was examined. To examine the impact of the hyper-parameters, the OOB error rate and the number of bands that are selected by the algorithm were examined. The effect of *ntree* and *mtry* on band selection and error rates were examined. The best-selected bands (502.4, 636.3, 669.7, 683, 729, 8 and 850.4) from maize streak virus dataset to resampled AISA eagle using GRRF yielded an OOB error of 8.4 %. Best bands (480.4, 571.9, 633.5 and 708,9) selected from maize streak virus dataset resampled to HyMap yielded an OOB error of 7.25 %. The best subsets of bands were then used as input variables in random forest classifier. On a resampled Hymap, bands selected by GRRF (n= 4) yielded an overall accuracy of 91.67% and a Kappa value of 0.89 and bands selected by FVS (n=8) yielded a relatively lower overall accuracy of 87.60% and a Kappa value of 0.83. On a resampled AISA, bands selected by GRRF (n= 6) yielded an accuracy of 89.17% and a Kappa value of 0.86 and bands selected by FVS (n=10) yielded a relatively lower overall accuracy of 85.85% and a Kappa value of 0.81. The results have shown that the GRRF algorithm has the potential to select compact feature subsets, and the accuracy performance is better than that of RF's variable selection method. The GRRF was considered to be a vigorous model for dealing with redundancy in the complexity of hyperspectral data.

Keywords: Random forest, Forward variable selection, guided regularized random forest, hyperspectral remote sensing.

3.0 Introduction

Recent developments in hyperspectral remote sensing technology (imaging spectrometers) have seen a dramatic improvement in the characterisation of terrestrial features (Artigas and Yang, 2005, Bassani et al., 2009, Cho et al., 2010, Choe et al., 2008, Fava et al., 2009, Adam and Mutanga, 2009b, Ismail and Mutanga, 2011a). The value of imaging spectrometers lies in combining imaging and spectroscopy in a unique system that is able to detect subtle variations in surface features in many contiguous and narrow spectral bands (between 380 nm and 2500 nm with bandwidths of less than 2 nm) (Demarchi et al., 2014).

Processing of hyperspectral data is principally hampered by the hyper-dimensionality of the data which requires adequate training samples to simplify the multifaceted nature of classification and prediction processes (Li et al., 2011, Pal, 2009, Bajcsy and Groves, 2004, Hsu, 2007). Practically, in most of the hyperspectral applications, the number of bands limits the number of training samples (Hsu, 2007).

The 'curse of dimensionality' (small n large p problem) causes the 'peaking phenomenon' or 'Hughes phenomenon' (Hsu, 2007). The 'Peaking phenomenon' introduces multi-collinearity in the input data which makes the estimation of parameters inaccurate and unreliable (Hsu, 2007, Kavzoglu and Mather, 2002). Furthermore, the use of a large number of features is time-consuming and may increase the complexity of the modelling task (Pal, 2009). Therefore, techniques for 'curse of dimensionality' reduction without compromising significant information are critical in hyperspectral data processing and analysis (Borges et al., 2007, Pal, 2005, Shaw and Manolakis, 2002).

The commonly-used methods to reduce dimension can be grouped into two categories: feature selection and feature extraction. Feature selection methods reduce the dimensionality by selecting a subgroup of features capturing the relevant properties of the entire data set. On the other hand, feature extraction methods provide innovative features based on a linear or nonlinear transformation of the original feature sets (Pal, 2009). Given the fact that the hyperspectral band has its

own corresponding image, this conversion could not keep the original physical interpretation of the image. Thus feature extraction methods are not preferable for the dimensionality reduction of hyperspectral images (Bajwa et al., 2004, Bruzzone and Serpico, 2000, Li et al., 2011).

Several hyperspectral techniques for feature or band selection have been proposed to reduce the 'curse of dimensionality' and to identify the optimal bands required for different applications (Daughtry and Walthall, 1998, Schmidt and Skidmore, 2003, Thenkabail et al., 2004, Thenkabail et al., 2002, Vaiphasa et al., 2005, Vaiphasa et al., 2007). These methods can be classified into the wrapper or filter approaches, based on whether or not they use classification algorithms as part of the evaluation process (Kavzoglu and Mather, 2002). The wrapper approach is a feature selection algorithm that looks for the best subset of bands using the classification algorithm as part of the evaluation process. On the other hand, the filter approach evaluates subsets of bands using the training data and without direct reference to the classification algorithm (Kavzoglu and Mather, 2002, Kohavi and John, 1997). The filter approach is computationally more efficient and has been more commonly used than the wrapper approach (Ismail et al., 2007, Schmidt and Skidmore, 2003, Vaiphasa et al., 2005). In the application of high dimensionality data such as hyperspectral data, it is suggested that the classification algorithm should be a part of the variable selection process (Granitto et al., 2006, Adam et al., 2012b). It is therefore desirable to have an algorithm that offers direct measurement of the importance of variables at the same time of the classification process of hyperspectral data (Ismail, 2009). Such calculations are able to reach a solution faster by avoiding retraining a predictor from scratch for every variable subset investigated (Guyon and Elisseeff, 2003, Adam et al., 2012b, Mutanga et al., 2012).

In recent years Random Forest (Breiman, 2001b) has emerged as a powerful algorithm for the classification of remotely sensed data (Ismail and Mutanga, 2011b, Adam et al., 2012b, Adjorlolo et al., 2012a, Lawrence et al., 2006, Rodriguez-Galiano et al., 2012). The effectiveness and efficiency of the RF algorithm are clarified by the capability of the algorithm to deal with categorical and numerical data, missing values, different scales between variables, interactions and non-linearity's present in the dataset (Deng and Runger, 2012a). Within hyperspectral applications,

the algorithm has been used for classification purposes in addition to providing a measure of variable importance (Adam et al., 2012b, Mansour et al., 2012b, Lawrence et al., 2006). Later research has focussed on using the variable importance to improve model interpretation and classification accuracies (Archer and Kimes, 2008, Díaz-Uriarte and de Andrés, 2006, Granitto et al., 2006, Svetnik et al., 2003, Han et al., 2007, Hamza and Larocque, 2005). However, since the algorithm offers a ranking and does not eliminate redundant bands automatically, researchers have combined the algorithm with various feature selection methods such as recursive feature elimination (RFE) framework (Ismail and Mutanga, 2011b) and forward variable selection (Adam et al., 2012b, Mansour et al., 2012b) as a wrapper in order to obtain the optimum subset of bands that best explain the phenomena of interest. However, these methods are computationally intensive and require building multiple models after which the model with the lowest error rate is selected. Moreover, it has been reported that the integrations between variable importance measurements by traditional random forests and variable selection processes show significant preference on highly correlated predictor variables (Strobl et al., 2008, Nicodemus et al., 2010, Adjorlolo et al., In press). Consequently, experts have suggested the use of the kind of supervise approach to mitigate the problems associated with random forest variable selection process (Adjorlolo et al., In press). Moreover, it would be valuable if the feature selection algorithm required building only one model to identify the best subset of the variable that best explains the phenomena of interest (Guyon et al., 2010).

Consequently, a regularization framework which was applied to random forest and boosting tree was suggested by Deng and Runger, (2012b). The regularized framework builds a model once and only requires training a single model for variable selection which may significantly reduce the training time (Guyon et al., 2010). Furthermore, the framework avoids selecting the new feature for splitting the data in a tree node when that feature produces similar information to the feature already selected (Deng and Runger, 2012a). Thus, comparative research have shown that the framework is effective in selecting high-quality feature subsets while maintaining predictive accuracies (Deng and Runger, 2012a). It is within this context that the researcher compared the ability of the guided regularized Random Forest and traditional RF's variable selection methods to select a subset of optimal bands for the

classification of four stages of MSV infection on maize using resampled Hymap and AISA Eagle bands. Additionally, the behaviour of the regularized random forest is assessed by considering multiple criteria related to variations in the classifier parameters values. The performance of the guided regularized random forest is also evaluated in comparison to traditional Random Forest as a classification and a variable selection method. Random Forest is frequently applied as they attain a high prediction accuracy and have the ability to identify informative and important variables.

Random forest's forward variable selection method was compared with guided regularized random forest in selecting the optimum variables. The *mtry* value was accepted as the square root of the total variables ($mtry = \sqrt{p}$) as recommended by (Breiman, 2001b), the *ntree* value was set to 1000 trees and the coefficient of regularization for the regularized random forest was set to 0.8 (Deng and Runger, 2012b). In order to provide a robust indicator of comparative error, the OOB error was reported. Moreover, the sensitivity of regularized random forest to specific hyper-parameters was examined. To examine the effect of the hyper-parameters, the OOB error rate and the number of bands that are selected by the algorithm were examined. The effect of *ntree* and *mtry* on band selection and error rates was examined.

3.1 Variable Ranking

Random forest algorithm can return three methods of variable importance (Breiman, 2001b). Such variable importance measures depends on the selection rate of each candidate variables; the Gini index, based on the principle of impurity reduction (Pal, 2005) and permutation of predictor variables as an estimate of variable importance (Strobl et al., 2008). Amongst these variable importance measures, the Gini index has been found to show a bias when predictor variables vary in their number of categories or scale of measurement (Strobl et al., 2008). This is because the fundamental Gini gain splitting criterion is a biased estimator and can be affected by multiple testing effects (Strobl et al., 2008).

The other variable importance concept that is based on the impact of a predictor variable commonly termed "mean decrease in accuracy", compares each candidate predictor variable with respect to its effect in predicting the response or its causal effect using the OOB error selection rates of each ensemble of trees (Breiman, 2001b). The later variable importance measure has been shown, calculated successfully by means of randomly permuting each predictor variable's association with response variables (Strobl et al., 2008).

The variable importance follows the logic that a random permutation of the values of predictor variables represents the absence of a variable from the model. Thus, the variance in prediction accuracy prior and after permuting a variable (i.e. the class membership of a permuted variable, together with the remaining non-permuted predictor variables) is used to predict the response for the OOB observations as the measure of importance (Strobl et al., 2008). In this context, the number of observations classified correctly decreases substantially if the permuted band was associated with the multivariate response variables (Breiman, 2001b). That is, by randomly permuting the maize streak virus reflectance band values, its original association with the response variable is broken. The Gini index -based variable importance was implemented in this study.

3.2 Random forest and regularized random forest

In order to determine the split at each node of tree, Random forest has the additional modification of selecting only a random subset of candidate features (*mtry*). When the tree is maximally grown, it makes estimates using an out of bag (OOB) sample for that particular tree. The prediction error will then provide an unbiased assessment of the accuracy, since the OOB sample is not used in the training process. Additionally, the random forest provides an internal measure of variable importance using the OOB sample. In this study, the *Gini* index was examined as an importance measure. The index selects the best band based on an information gain score.

The *Gini* index at node v defined as:

$$Gini(v) = \sum_{k=1}^k p_k^v (1 - p_k^v)$$

Where p_k^v represents the ratio of class k observations at node v . The *Gini* information gain (X_i, v) is the variance between the impurity at node v and the weighted average of impurities at each child node of v . The weights are proportional to the number of samples assigned to each child from the split at node v so that

$$Gain(X_i, v) = Gain(v) - w_L Gini(v_L) - w_R Gini(v_R)$$

Where $Gini(v_L)$ and $Gini(v_R)$ are the impurity scores and w_L and w_R are the weights for the left and right child nodes. The principle of the regularization framework is to utilize a regularized version of gain at each node v as follows

$$Gain(X_i, v) = \begin{cases} \lambda_i, Gain(X_i, v) & X_i \notin F \\ Gain(X_i, v) & X_i \in F \end{cases}$$

Where F is the feature set selected in the previous nodes and $\lambda_i \in (0,1)$ is called the coefficient of regularization for (X_i). This coefficient is then used to penalise using a feature $X_i \notin F$ for splitting. A smaller λ_i leads to a larger penalty. The regularized random forest uses the $Gain_R(X_i, v)$ at each node v and adds new features to F if those features provide new predictive information. According to Deng and Runger, (2012a), and Deng and Runger (2012b) the procedure allows for the most significant bands to have an advantage to enter the feature set (F). The coefficient of regularization (λ_i) is calculated as

$$\lambda_i = (1 - \gamma) * \lambda_0 + \gamma * Imp_i$$

Where $Imp_i \in (0,1)$ is the normalized variable importance score of X_i from an original random forest $\lambda_0 \in (0,1)$ that controls the degree of regularization and is called the base coefficient and $\gamma \in (0,1)$ controls the weight of the normalized importance score and is called the importance coefficient. The regularized random forest was done in R (R Development Core Team, 2007) using the RRF package that is freely distributed and is available from <http://cran.r-project.org/>

3.3 Resampling field spectra

Processing hyperspectral data requires dimensionality of data to be reduced without losing the spectral separability of a considered feature space (Thenkabail et al., 2000, Schmidt and Skidmore, 2003). To reduce data dimensionality, a number of techniques have been developed. The techniques include among others the resampling of spectra to wider bandwidths around a few band-centers or to the spectral configuration of existing sensors, using their respective spectral response functions or spectral resolution (i.e., FWHMA, full width at half maximum,). In this study, the spectral measurements were resampled to the Hymap, and AISA Eagle spectral resolutions using the ENVI 4.7 image processing software (ENVI, 2009). Then variable selection and classification was done using the resampled spectra.

The spectral measurements from each of the MSV infection stages ($n = 4$) were resampled to HYMAP and Airborne Imaging Spectroradiometer for Applications (AISA) AISA Eagle spectra using ENVI 4.3 image processing software. The method used a Gaussian model with a full width at half maximum (FWMAP) equal to the band spacing provided (Mutanga, 2005). HYMAP is an airborne hyperspectral imaging spectrometer, comprising 126 wavelengths, operating over the spectral range 436.5–2485 nm, with average spectral resolutions of 15 nm (437–1313 nm), 13 nm (1409–1800 nm) and 17 nm (1953 nm–2485 nm) (Cho et al., 2007). AISA Eagle is an airborne hyperspectral imaging spectrometer, comprising 126 wavelengths. The spectral reflectance was resampled because the current operational airborne and space borne sensors such as HYMAP and AISA Eagle lack the fine spectral resolution of the ASD spectral reflectance (Mutanga et al., 2005). Furthermore, in line with the current availability of airborne sensors in South Africa, it is of interest if the specific spectral bands of these sensors can discriminate between infection stages of maize streak virus. In the event that the outcomes are certain, the mapping and monitoring of maize diseases could be operational on airborne hyperspectral platforms.

3.4 Optimizing using the random forest

The purpose of the optimization process was to define the best parameters for each classifier so as to get the best classification accuracies. The *ntree* and *mtry* for RF were optimized using the grid search and a ten-fold cross-validation method (Waske et al., 2009). That dataset was divided into ten subsets of equal size, RF model was then used to train nine subset samples, and tested on the omitted one. The process was repeated ten times until all subset samples were served as test samples. The pair of parameters for each classifier that reduces the classification error was then considered as the best values for final classification. Based on the recommendation of (Statnikov et al., 2008) *ntree* values up to 5000 were considered using intervals of 500 while the default *mtry* was used (for example 1/3, default *mtry*). The default value of *mtry* is based on the square root of the 272 AISA Eagle and 126 Hymap wavebands.

3.5 Variable Selection

The optimum bands are set of bands, with the least correlation among themselves, high information content and are able to discriminate the target. Optimum can be quantified using Random Forest's (RF) Forward variable selection and Guided Regularized Random Forest (GRRF). The variable importance as calculated by RF was used to rank the 272 AISA Eagle and 126 Hymap wavebands according to their ability to discriminate amongst healthy, early, moderate and severely infected maize leaves with maize streak virus. Forward variable selection method was executed to identify the least number of the spectral bands that produced the maximum classification accuracy, (Kohavi and John, 1997). Multiple RFs were iteratively fitted using the ranked wavebands in a sequential manner.

Initially, a new RF model was built using the highest two ranked bands and for the next iteration, the two highest bands were considered. This process was repeated until all the spectral variables used in this study ($n = 272$ for AISA and 126 for Hymap) were considered (Adam and Mutanga, 2009b). Finally, the subset of spectral bands that produced the lowest 10-fold cross-validated error was selected

as the optimum subset of spectral bands for classification. Random Forests was used to distinguish relevant from irrelevant variables in variable selection approaches e.g. (Díaz-Uriarte and de Andrés, 2006, Sandri and Zuccolotto, 2006, Yang et al., 2009).

The limitations of the random forest algorithm in measuring variable importance is that it does not select automatically the optimum number of variables that produce the best classification accuracy (Adam and Mutanga, 2009b). Therefore, GRRF was used to automatically select the optimal number of wavelengths based on the Random Forest measurement of variables importance (Adam and Mutanga, 2009b, Ismail and Mutanga, 2010). Here, a guided RRF (GRRF), was proposed in which the importance scores from the ordinary random forest were used to guide the feature selection process in RRF. Since the importance scores from an RF are calculated based on all trees in the RF and all the training data, GRRF is expected to perform better than RF.

When using the GRRF, the importance scores from an ordinary random forest (RF) are used to guide the feature selection process in RF (Daughtry and Walthall, 1998, Schmidt and Skidmore, 2003, Thenkabail et al., 2004, Thenkabail et al., 2002, Vaiphasa et al., 2005). Several studies show that GRRF, in general, is able to select compact feature sub sets and is better than RF, varSelRF and LASSO-logistic regression in terms of the accuracy of RF and a decision tree method (Thenkabail et al., 2002). Feature selection has been widely used in many applications as it can reduce the curse of dimensionality, improve interpretability and avoid unnecessary work of analyzing irrelevant and redundant features. A powerful classifier called random forest (RF) has been widely used for measuring feature importance as it naturally handles numerical and categorical variables, different scales, interactions, and nonlinearities, etc (Schmidt and Skidmore, 2003, Thenkabail et al., 2004). Though the random forest feature importance scores can be used to select K features with the highest importance scores (the K best features), there could be redundancy among the K variables, which is different from feature selection that selects a set of relevant but non-redundant features (the best K-feature sub set).

3.6 Accuracy assessment

The accuracy of each classifier was assessed using the 30% holdout sample. In order to evaluate the accuracy of the RF classifiers, the overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) were used (Mather and Tso, 2009). Overall accuracy is a ratio (%) between the number of correctly classified samples and the number of test samples, while User accuracy shows the likelihood that a sample belongs to the specific class and the classifier accurately assigns it such a class. Producer accuracy shows the probability of a certain class being correctly classified.

3.7 Results

3.7.1 Variable importance using the random forest algorithm

Using the default setting of traditional random forest, the importance of resampled AISA Eagle and Hymap bands was measured using the mean decrease in Gini index. What can be noticed from Figures 3.1a, and b is that there are key regions of the electromagnetic spectrum that are important for classifying the maize streak virus. The most important spectral wavebands for AISA are located in the red and the rededge (631.9 nm, 667.4 nm, 669.7 nm, and 683.0 nm), of the electromagnetic spectrum, and for Hymap are located in the red and red edge region (618 nm, 633.5 nm, 679.4 nm, 694.4 nm). The results shows that random forest algorithm has successfully explored and described the relative importance of each individual wavelength in discriminating among the four stages of MSV infection.

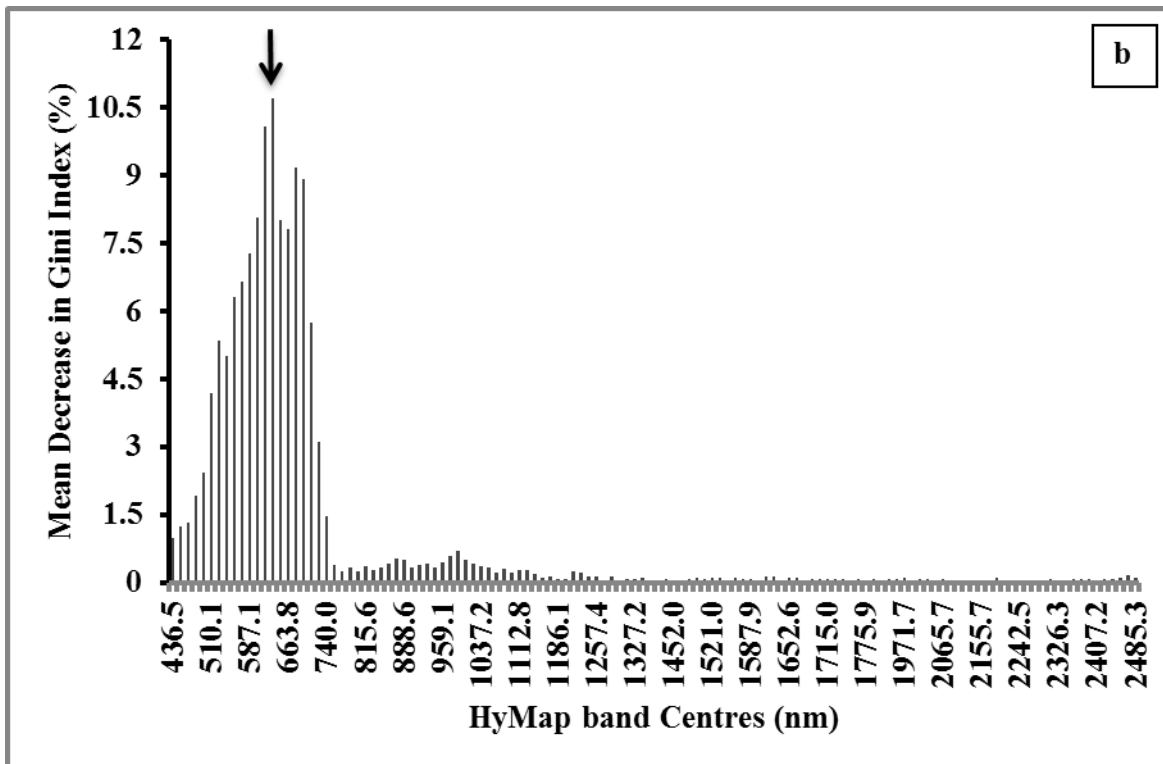
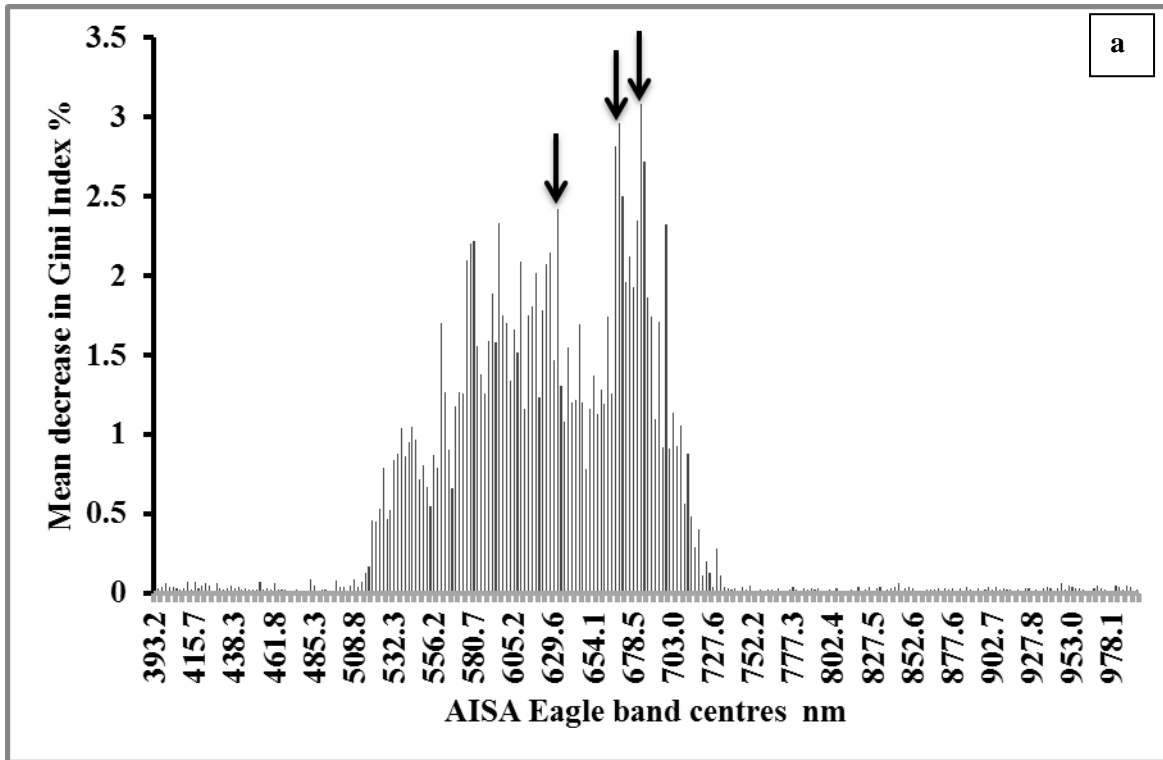


Figure 3.1 The variable importance as measured by traditional random forest for resampled AISA Eagle (a) and resampled Hymap (b). The errors were calculated using the mean decrease in the Gini index and the default settings. The black arrow indicates the most important wavelengths

3.7.2 Random Forest's Forward Variable selection

The study concentrated on the prediction performance of RF focusing on out-of-bag (OOB) error (Breiman, 2001b). The study used this prediction error estimate for three reasons: the main reason being that of comparing models instead of assessing models, the second is that it gives fair assessment compared to the usual alternative test set error even if it is considered as a little bit optimistic and the last one, is that it is a default output of the Random Forest procedure, so it is used by almost all users.

The forward variable selection method selected an optimal of 10 spectral bands from the resampled AISA Eagle and 8 spectral bands from the resampled Hymap using the ranking output of RF for discriminating amongst classes. The 10 spectral wavebands of AISA Eagle produced a minimal OOB error of 12% using the training dataset and a bootstrap error of 11% (Figure 3.2 a). The 8 spectral bands of Hymap produced a minimum OOB error of 10.58% and a bootstrap error of 11.23 % (Figure 3.2 b). These spectral wavebands were then used as the optimal input variables in the Random Forest classifier.

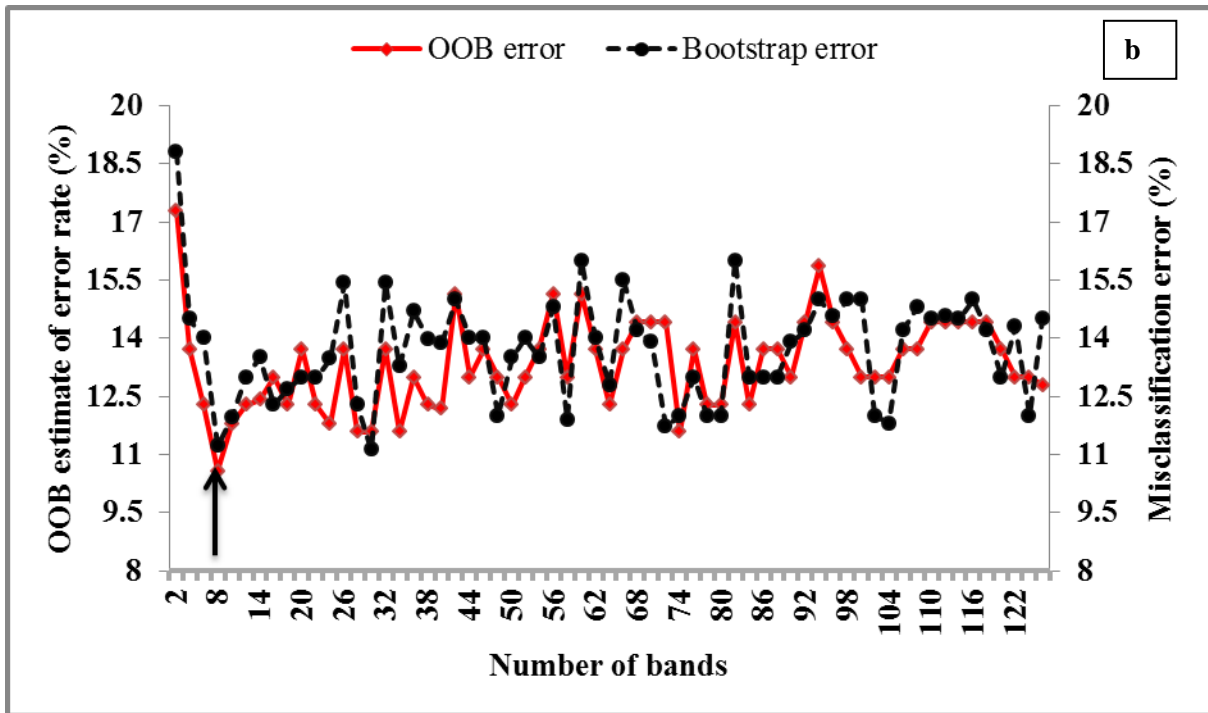
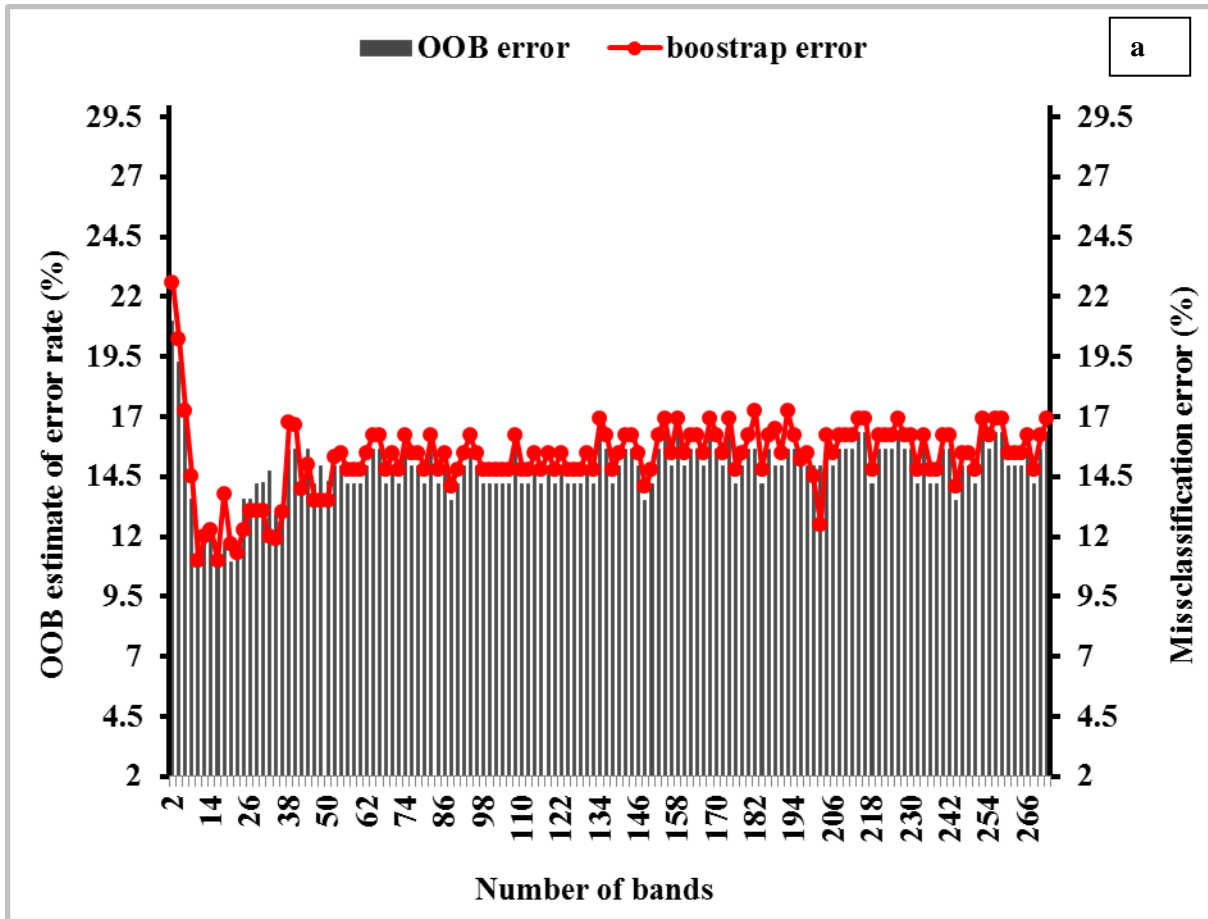


Figure 3.2: Forward variable selection method AISA dataset (a) and HyMap dataset (b)

3.7.3 Variable selection using GRRF

The variable selection has been extensively used in many applications as it can reduce the curse of dimensionality, improve interpretability and avoid unnecessary work analysing irrelevant and redundant features. Recently the regularized random forest (RRF) has been proposed for feature selection by building only one ensemble. However, in RRF the features are assessed by a part of the training data at each tree node, and thus the feature selection process may not be stable. In this study an enhanced RRF, referred to as guided RRF (GRRF), is proposed. When using the GRRF, the importance scores from an ordinary random forest (RF) are used to guide the feature selection process in RRF. Experimental studies show that GRRF, in general, is able to select compact feature sub sets and is better than RRF, varSelRF and LASSO-logistic regression in terms of the accuracy of RF and a decision tree method. Both RRF and GRRF were implemented in the “RRF” package available at CRAN (<http://cran.r-project.org/>), the official R package archive. The best-selected bands (502.4, 636.3, 669.7, 683, 729, 8 and 850.4) from AISA eagle dataset using GRRF yielded an OOB error of 8.4 %. Best bands (480.4, 571.9, 633.5 and 708,9) selected from HyMap dataset produced an OOB error of 7.25 %. The best subsets of bands were used as input variables in random forest classifier.

3.7.4 Accuracy Assessment

Accuracy assessment was done using the bands that were selected by GRRF and RF's FVS methods for both the two sensors in order to assess the prediction performance of the models using an independent test dataset.

Table 3.1: Shows bands selected by GRRF from AISA Eagle (n= 6) yielded a higher accuracy (89.17%) while those selected using FVS for AISA Eagle (n=10) yielded an overall accuracy of 85.85%

Sensor and Method	AISA FVS	AISA GRRF	Hymap FVS	Hymap GRRF
	683.0	502.4	633.5	480.4
	669.7	636.3	618.0	571.9
	667.4	669.7	679.0	633.5
	685.2	683.0	694.4	708.9
	671.9	729.8	602.5	
	631.9	850.4	648.7	
	680.8		663.8	
	596.3		587.1	
	698.6			
	580.7			
Total Number of bands	10	6	8	4
OA	85.85%	89.17%	87.60%	91.67%

Table 3.2: Confusion matrix for AISA Eagle and Hymap showing the classification accuracy for the different levels of maize streak virus severity (H- Healthy, E- Early stage, M-Moderate, and S-Severe)

	AISA (GRRF)					HYMAP (GRRF)					AISA (FVS)					HYMAP (FVS)				
	H	E	M	S	Total	H	E	M	S	Total	H	E	M	S	Total	H	E	M	S	Total
H	28	1	1	0	30	28	1	1	0	30	27	1	2	0	30	28	2	1	0	30
E	2	26	1	1	30	2	26	1	1	30	2	25	2	1	30	3	24	1	2	30
M	2	2	26	0	30	1	2	27	0	30	3	2	25	1	30	1	1	26	2	30
S	1	1	1	27	30	0	0	1	29	30	1	1	2	26	30	0	1	1	28	30
TOTAL	33	30	29	28	120	31	29	30	30	120	33	30	31	28	120	31	29	30	30	120

(30%): OOB error = 10.83%	(30%) OOB error = 8.33%	OOB error = 14.17%	OOB error = 12.40 %
Overall accuracy = 89.17%	= 91.67%	= 85.85%	= 87.60%
KHAT Value = 0.86	= 0.89	= 0.81	= 0.83
Producer Accuracy= 84.85%	= 90.32%	= 81.82%	= 87.5%
User Accuracy = 93.33%	= 93.33%	= 90%	= 90.32%

On a resampled Hymap, bands selected by GRRF (n=4) produced an accuracy of 91.67% while bands selected by RF's FVS (n=8) produced a lower accuracy (87.60%). On a resampled AISA, bands selected by GRRF (n=6) produced an accuracy of 89.17% bands selected by RF's FVS (n=10) produced a lower accuracy of 85.85%.

3.8 Discussion

The objective of this study was to test the ability of RF's FVS and GRRF in selecting optimum variables from resampled AISA Eagle and Hymap. The two methods yielded accuracies that can be compared. Bands selected by GRRF yielded better accuracy than bands selected by RF's FVS. Hyperspectral band selection allows for the interpretation of selected bands based on physiological and/or structural information of vegetation and for the selection of a set of optimal bands that produce a predictive model. Results from this analysis have shown that the GRRF is able to

select the best number of bands that produce an accurate model for the classification of maize streak virus infestation.

In comparison to related studies, the results produced by the guided regularized random forest are superior in classification accuracy and an optimal subset of bands that have best discriminatory power. This result was expected as recent hyperspectral studies have shown that reducing the number of input bands produces more accurate random forest classifier. Researchers attribute the improvement in classification accuracy due to, the trade-off between bias and variance, sensitivity to regularization parameters, sensitivity to *mtry* and *ntree* parameters. *Mtry* is always chosen based on the lowest error rate.

Bands selected by GRRF for both sensors yielded a higher accuracy than those selected by FVS. There is a significant difference in the number of bands selected by the two methods. GRRF selects fewer bands and produced relatively lower OOB error and higher overall accuracies for both the sensors than the RF's FVS. The researcher used MSV data sets for the experiments and the results show that GRRF can select compact feature sub sets, and the accuracy performance is better than that of RF's variable selection method. There is a relative difference in overall accuracy for the two methods and the GRRF selects fewer bands than the FVS. Several studies have concluded that choosing a set of appropriate variables could improve the accuracy of classification and reduce the training time and reduces the complexity of the problem. (Adam et al., 2012b, Bajcsy and Groves, 2004, Borges et al., 2007, Li et al., 2011, Mutanga et al., 2012, Pal, 2009, Abdel-Rahman et al., 2013). This is due to the fact that a model which has a large number of variables as input data does not guarantee a more accurate result because some of these variables might not be useful for modelling processes (Pal, 2009). Such variables are the same as noise to any statistical model even if it is unique and accurate (Bajcsy and Groves, 2004, Li et al., 2011, Pal, 2009).

3.9 Conclusion

This study utilises an analysis design in order to evaluate the validity and replicability of variable selection methods in terms of both prediction and description for comparative purposes. This study shows that the variable selection process is critical in order to ensure optimal and stable results in the estimation process so that methods can be compared. The results of this study present a successful application of the random forest's variable selection and guided regularized random forest in selecting the optimal variables. The results have shown that the GRRF algorithm has the potential to select compact feature sub sets, and the accuracy performance is better than that of RF's variable selection method. The GRRF was considered to be a vigorous model for dealing with redundancy in the complexity of hyperspectral data.

In the future, research should target ways that enhance the understanding of how *n_{tree}* and *m_{try}* parameters are set in the models. Overall, the feasibility and flexibility of using the GRRF algorithm for detecting MSV using hyperspectral data can assist in making decisions regarding site-specific applications of chemicals and fungicides. However, for the approach presented in this study to be operational, various available hyperspectral and newly-launched multispectral sensors, together with other environmental variables, such as rainfall and temperature, should be investigated and tested in different maize diseases and different climatic conditions.

3.10 References

- Abdel-Rahman, E.M., F.B. Ahmed, and R. Ismail. 2013. "Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data." *International Journal of Remote Sensing* 34 (2):712-28.
- Adam, E. M., O. Mutanga, D. Rugege, and R. Ismail. 2012. "Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP." *International Journal of Remote Sensing* 33 (2):552-69. doi: 10.1080/01431161.2010.543182.
- Adam, Elhadi, and Onesimo Mutanga. 2009. "Spectral discrimination of papyrus vegetation (*Cyperus papyrus* L.) in swamp wetlands using field spectrometry." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (6):612-20.
- Adam, Elhadi, Onesimo Mutanga, and Denis Rugege. 2010. "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review." *Wetlands Ecology and Management* 18 (3):281-96.
- Adjorlolo, Clement, Moses A. Cho, Onesimo Mutanga, and Riyad Ismail. 2012. "Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features." *Journal of Applied Remote Sensing* 6 (1):063560-1. doi: 10.1117/1.jrs.6.063560.
- Adjorlolo, Clement, Onesimo Mutanga, Moses A. Cho, and Riyad Ismail. In press. "Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification." *International Journal of Applied Earth Observation and Geoinformation* (0). doi: 10.1016/j.jag.2012.07.011.
- Archer, KJ, and RV Kimes. 2008. "Empirical characterization of random forest variable importance measures." *Computational statistics & data analysis* 52 (4):2249-60.

- Artigas, FJ, and JS Yang. 2005. "Hyperspectral remote sensing of marsh species and plant vigour gradient in the New Jersey Meadowlands." *International Journal of Remote Sensing* 26 (23):5209-20.
- Bajcsy, P, and P Groves. 2004. "Methodology for hyperspectral band selection." *Photogrammetric engineering and remote sensing* 70:793-802.
- Bassani, Cristiana, Rosa Maria Cavalli, Roberto Goffredo, Angelo Palombo, Simone Pascucci, and Stefano Pignatti. 2009. "Specific spectral bands for different land cover contexts to improve the efficiency of remote sensing archaeological prospection: The Arpi case study." *Journal of Cultural Heritage* 10:e41-e8.
- Bock, KR, EJ Guthrie, and RD Woods. 1974. "Purification of maize streak virus and its relationship to viruses associated with streak diseases of sugar cane and *Panicum maximum*." *Annals of Applied Biology* 77 (3):289-96.
- Borges, J.S, A.R.S Marcal, and J.M.B Dias. 2007. Evaluation of feature extraction and reduction methods for hyperspectral images. Paper presented at the New Developments and Challenges in Remote Sensing, Poland, 29 May- 2 June 2006.
- Bosque-Pérez, Nilsa A. 2000. "Eight decades of maize streak virus research." *Virus research* 71 (1):107-21.
- Breiman, L. 2001a. "Random forests." *Machine Learning* 45:5 - 32.
- Breiman, L. 2001b. "Random forests." *Machine learning* 45 (1):5-32.
- Cho, Moses A., Jan van Aardt, Russel Main, and Bongani Majeke. 2010. "Evaluating variations of physiology-based hyperspectral features along a soil - water gradient in *Eucalyptus grandis* plantation." *International Journal of Remote Sensing* 31 (12):3143–59.
- Cho, Moses Azong, Andrew Skidmore, Fabio Corsi, Sipke E Van Wieren, and Istiak Sobhan. 2007. "Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares

- regression." *International Journal of Applied Earth Observation and Geoinformation* 9 (4):414-24.
- Choe, Eunyong, Freek van der Meer, Frank van Ruitenbeek, Harald van der Werff, Boudewijn de Smeth, and Kyoung-Woong Kim. 2008. "Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain." *Remote Sensing of Environment* 112 (7):3222-33. doi: <http://dx.doi.org/10.1016/j.rse.2008.03.017>.
- Daughtry, CST, and CL Walthall. 1998. "Spectral discrimination of Cannabis sativa L. leaves and canopies." *Remote Sensing of Environment* 64 (2):192-201.
- Demarchi, Luca, Frank Canters, Claude Cariou, Giorgio Licciardi, and Jonathan Cheung-Wai Chan. 2014. "Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping." *ISPRS Journal of Photogrammetry and Remote Sensing* 87:166-79.
- Deng, H, and G Runger. 2012a. "Gene selection with regularized random forest " *Bioinformatics*:1-5.
- Deng, Houtao, and George Runger. 2012b. Feature selection via regularized trees. Paper presented at the Neural Networks (IJCNN), The 2012 International Joint Conference on.
- Díaz-Uriarte, R, and A de Andrés. 2006. "Gene selection and classification of microarray data using random forest." *BMC bioinformatics* 7 (1):3.
- ENVI. 2009. *ENVI 4.7: environment for visualizing images. Exelis Visual Information Solutions.*: ITT Industries, Colorado.
- Fava, F, R Colombo, S Bocchi, M Meroni, M Sitzia, N Fois, and C Zucca. 2009. "Identification of hyperspectral vegetation indices for Mediterranean pasture characterization." *International Journal of Applied Earth Observation and Geoinformation* 11 (4):233-43.

- Granitto, PM, C Furlanello, F Biasioli, and F Gasperi. 2006. "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products." *Chemometrics and Intelligent Laboratory Systems* 83 (2):83-90.
- Guyon, I, and A Elisseeff. 2003. "An introduction to variable and feature selection." *The Journal of Machine Learning Research* 3:1157-82.
- Guyon, Isabelle, Amir Saffari, Gideon Dror, and Gavin Cawley. 2010. "Model selection: Beyond the bayesian/frequentist divide." *The Journal of Machine Learning Research* 11:61-87.
- Hamza, M, and D Larocque. 2005. "An empirical comparison of ensemble methods based on classification trees." *Journal of Statistical Computation and Simulation* 75 (8):629-43.
- Han, P , X. Zhang, R.S Norton, and Z.P Feng. 2007. Reducing overfitting in predicting intrinsically unstructured proteins. Paper presented at the the 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'2007), Nanjing, China.
- Hsu, PH. 2007. "Feature extraction of hyperspectral images using wavelet and matching pursuit." *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (2):78-92.
- Ismail, R, and O Mutanga. 2011a. "Discriminating the early stages of Sirex noctilio infestation using random forest and shortwave infrared (SWIR) wavelengths." *International Journal of Remote Sensing*.
- Ismail, R, O Mutanga, and F Ahmed. 2007. "Discriminating Sirex noctilio attack in pine forest plantations in South Africa using high spectral resolution data." In *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*, edited by M Kalacska and A Sanchez-Azofeifa, 350 Rutledge, USA: Taylor and Francis: CRC Press.
- Ismail, R. 2009. "Remote sensing of forest health: The detection and mapping of Pinus patula trees infested by Sirex noctilio." dissertation, University of KwaZulu-Natal,.

- Ismail, R., and O. Mutanga. 2011b. "Discriminating the early stages of *Sirex noctilio* infestation using classification tree ensembles and shortwave infrared bands." *International Journal of Remote Sensing* 32 (15):4249-66.
- Kavzoglu, T, and PM Mather. 2002. "The role of feature selection in artificial neural network applications." *International Journal of Remote Sensing* 23 (15):2919-37.
- Kohavi, R, and GH John. 1997. "Wrappers for feature subset selection." *Artificial Intelligence* 97 (1-2):273-324.
- Lawrence, R. L, S.D Wood, and R.L Sheley. 2006. "Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForests)." *Remote Sensing of Environment* 100:356-62.
- Li, Shijin, Hao Wu, Dingsheng Wan, and Jiali Zhu. 2011. "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine." *Knowledge-Based Systems* 24 (1):40-8.
- Mansour, Khalid, Onesimo Mutanga, Terry Everson, and Elhadi Adam. 2012. "Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution." *ISPRS Journal of Photogrammetry and Remote Sensing* 70:56-65.
- Mather, Paul, and Brandt Tso. 2009. *Classification methods for remotely sensed data*: CRC press.
- Mutanga, O. 2005. "Discriminating tropical grass canopies grown under different nitrogen treatments using spectra resampled to HYMAP." *International Journal of Geoinformatics* 1 (2):21-32.
- Mutanga, O, AK Skidmore, L Kumar, and J Ferwerda. 2005. "Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain." *International Journal of Remote Sensing* 26 (6):1093-108.

- Mutanga, Onisimo, Elhadi Adam, and Moses Azong Cho. 2012. "High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm." *International Journal of Applied Earth Observation and Geoinformation* 18:399-406.
- Nicodemus, Kristin, James Malley, Carolin Strobl, and Andreas Ziegler. 2010. "The behaviour of random forest permutation-based variable importance measures under predictor correlation." *BMC bioinformatics* 11 (1):110.
- Pal, M. 2005. "Random forest classifier for remote sensing classification." *International Journal of Remote Sensing* 26 (1):217-22.
- Pal, Mahesh. 2009. "Margin-based feature selection for hyperspectral data." *International Journal of Applied Earth Observation and Geoinformation* 11 (3):212-20. doi: <http://dx.doi.org/10.1016/j.jag.2009.02.001>.
- R Development Core Team. 2007. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan. 2012. "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture." *Remote Sensing of Environment* 121 (0):93-107. doi: 10.1016/j.rse.2011.12.003.
- Sandri, Marco, and Paola Zuccolotto. 2006. "Variable selection using random forests." In *Data analysis, classification and the forward search*, 263-70. Springer.
- Schmidt, KS, and AK Skidmore. 2003. "Spectral discrimination of vegetation types in a coastal wetland." *Remote Sensing of Environment* 85 (1):92-108.
- Shaw, G, and D Manolakis. 2002. "Signal processing for hyperspectral image exploitation." *IEEE Signal Processing Magazine* 19 (1):12-6.
- Shepherd, Dionne N, Darren P Martin, Eric van der Walt, Kyle Dent, Arvind Varsani, and Edward P Rybicki. 2010. "Maize streak virus: an old and complex 'emerging' pathogen." *Molecular plant pathology* 11 (1):1-12.

- Statnikov, Alexander, Lily Wang, and Constantin F Aliferis. 2008. "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification." *BMC bioinformatics* 9 (1):319.
- Storey, Harold Haydon, and Südafrikanische Union. 1925. *Streak disease of sugarcane*.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. "Conditional variable importance for random forests." *BMC bioinformatics* 9 (1):307.
- Svetnik, V, A Liaw, C Tong, JC Culberson, RP Sheridan, and BP Feuston. 2003. "Random forest: a classification and regression tool for compound classification and QSAR modeling." *Journal of Chemical Information and Computer Scienci* 43 (6):1947-58.
- Thenkabail, Prasad S, Ronald B Smith, and Eddy De Pauw. 2000. "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics." *Remote Sensing of Environment* 71 (2):158-82.
- Thenkabail, PS, EA Enclona, MS Ashton, and B Van Der Meer. 2004. "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications." *Remote Sensing of Environment* 91 (3-4):354-76.
- Thenkabail, PS, RB Smith, and E De Pauw. 2002. "Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization." *Photogrammetric engineering and remote sensing* 68 (6):607-22.
- Thottappilly, G, NA BOSQUE-PÉREZ, and HW Rossel. 1993. "Viruses and virus diseases of maize in tropical Africa." *Plant Pathology* 42 (4):494-509.
- Vaiphasa, C, S Ongsomwang, T Vaiphasa, and AK Skidmore. 2005. "Tropical mangrove species discrimination using hyperspectral data: A laboratory study." *Estuarine, Coastal and Shelf Science* 65 (1-2):371-9.
- Vaiphasa, Chaichoke, Andrew K. Skidmore, Willem F. de Boer, and Tanasak Vaiphasa. 2007. "A hyperspectral band selector for plant species

discrimination." *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (3):225-35. doi: DOI: 10.1016/j.isprsjprs.2007.05.006.

Verikas, A., A. Gelzinis, and M. Bacauskiene. 2011. "Mining data with random forests: A survey and results of new tests." *Pattern Recognition* 44 (2):330-49. doi: DOI: 10.1016/j.patcog.2010.08.011.

Waske, Björn, Jon Atli Benediktsson, Kolbeinn Árnason, and Johannes R Sveinsson. 2009. "Mapping of hyperspectral AVIRIS data using machine-learning algorithms." *Canadian Journal of Remote Sensing* 35 (sup1):S106-S16.

Yang, Lian, Husheng Ding, Zhennan Gu, Jianxin Zhao, Haiqin Chen, Fengwei Tian, Yong Q Chen, Hao Zhang, and Wei Chen. 2009. "Selection of single chain fragment variables with direct coating of aflatoxin B1 to enzyme-linked immunosorbent assay plates." *Journal of agricultural and food chemistry* 57 (19):8927-32.

Chapter 4

TESTING THE UTILITY OF IN-SITU HYPERSPECTRAL DATA IN DETECTING THE SEVERITY OF MAIZE STREAK VIRUS

This chapter is based on:

Inos Dhau, Elhadi Adam, Onesimo Mutanga & Kingsley K. Ayisi (2017) Detecting the severity of maize streak virus infestations in maize crop using *in situ* hyperspectral data, Transactions of the Royal Society of South Africa, 73:1, 8-15,
DOI: [10.1080/0035919X.2017.1370034](https://doi.org/10.1080/0035919X.2017.1370034), Pages 8-15

Abstract

Maize streak geminivirus (MSV) causes maize streak disease, a major disease limiting maize production over wide-spread areas of Africa. There has always been an urgency about the need for developing quick and efficient methods of detecting such disease for control purposes as well as increased food production and security. The utility of remote sensing techniques in detecting the geminivirus infected maize was evaluated in this study based on experiments in Ofcolaco, Tzaneen in South Africa. Specifically, the potential of hyperspectral data in detecting different levels of MSV in maize was tested based on Guided Regularized Random Forest (GRRF). Specifically, the optimal bands for detecting different levels of maize streak disease in maize were 552 nm, 603 nm, 683 nm, 881 nm, and 2338 nm based on the GRRF algorithm. The findings from this study illustrate the strength of hyperspectral data in detecting different levels of MSV infections. This study highlights the potential of remotely sensed data in the accurate detection of food crop diseases such as MSV and their severity which is critical in crop monitoring to foster food security, especially in the resource-limited sub-Saharan Africa.

Key Words: Field spectroscopy, Random forest, Remote sensing, Maize Streak Virus (MSV)

4.0 Introduction

Maize (*Zea mays L.*) is a staple food for over 100 million people and it is the most important cereal crop in sub-Saharan Africa where approximately 15 million ha are planted annually (Olaniyan, 2015). Maize streak virus (MSV) is one of the most severe and widespread diseases that adversely reduces maize production thereby posing a threat to food security (Thottappilly et al., 1993). This disease was reported in the year 1901 in South Africa and its symptoms were first described as 'mealie variegation' (Fuller, 1901) and later renamed 'maize streak'. Maize streak starts as tiny, pale round spots on young leaves after infection (Shepherd et al., 2010). With time, newer leaves develop containing streaks that measure to several millimeters in length along the leaf veins. The streaks mainly coagulate along secondary and tertiary veins. Usually, the streaks are fused laterally, appearing as broken thin chlorotic stripes stretching along the leaves. The colour of these stripes varies from white to yellow. Occasionally, high MSV loads cause red pigmentation on the leaves as well as stunted plants (Shepherd et al., 2010).

Since Fuller's report, little was done until the 1920s when research on the disease and its causal agent began in Kenya. McClean, (1935) demonstrated that it is transmitted by *Balclutha (=Cicadulina) mbila Naude* (Homoptera:Cicadellidae) leaf hoppers. The damage to the maize crop by MSV varies each year with temperature and precipitation, but with a widespread occurrence, it can destroy crops to a yield loss of hundred percent (Alegbejo et al., 2002). Specifically, field trials depending on natural infection in East Africa reported yield losses of between 33 and 56% (Guthrie, 1977), while losses of 100% were reported in many countries in West Africa (Fajemisin et al., 1976a, Bosque-Perez et al., 1998). Research conducted between 1983 and 1985, presented by Fajemisin et al. (1976b), reported a yield reduction of 71 to 93% in maize due to MSV. Under conditions of natural infection, yield losses ranged from 24 to 76% (Fajemisin et al., 1976b). Across the African tropics, maize is grown predominantly as a subsistence crop and MSV is the most significant viral disease of Africa's most important food crop costing between US\$120m and US\$480m per year (Bosque-Pérez, 2000). Consequently, there is still an urgent need for novel techniques of forecasting epidemics such as MSV in Africa.

Monitoring the health of agricultural crops is a critical step in controlling stress induced by insects and diseases, which often results in high yield losses, poor quality produce and uncertainties when fostering food security interventions. It is still a common practice for farmers to indiscriminately apply agro-chemicals throughout the entire field in controlling diseases such as the MSV which brings about exorbitant economic losses. The quantification of the spatial extent and real-time distribution of such damages has been largely hinged on visual surveys. This procedure is time-consuming, tedious and, highly subjective at landscape scales. Therefore, to minimise economic costs and environmental pollution while ensuring food security, affordable, quick and consistent methods for agricultural crop monitoring and diseases forecasting are urgently required, especially in southern Africa where diseases such as MSV are frequent and severe. This can provide useful information for decision making on the necessity and appropriate timing of the application of insecticides.

Earth observation technologies have emerged as a quick and reliable procedure in detecting plant disease with a potential for being continuously utilised in remotely monitoring the physiology of agricultural crops (Chemura et al., 2017). Earth observation facilities offer spatially explicit information about the damage, real-time spatial distribution of disease infestation over large geographic areas (Deery et al., 2014, Raji et al., 2016). Remote sensing techniques have been widely utilised in plant pathology and crop protection (Yuan et al., 2017, Devadas et al., 2015, Yuan et al., 2014, Bauriegel and Herppich, 2014). Yuan et al., (2017) specifically, used remotely sensed data and analysis of variance to discriminate wheat insects and diseases. Williams et al., (2012) noted that hyperspectral narrow band channels 1900 nm and 2136 nm were the most optimal wavebands for detecting *Fusarium verticillioides* bacterial infections in maize based on partial least squares discriminant analysis. Water-stressed corn absorbed a lesser amount of light in the visible and more light in the NIR regions of the spectrum than the less water stressed and unstressed plants based on field hyperspectral data. Meanwhile, Devadas et al. (2015) successfully used hyperspectral data derived vegetation indices to discriminate the effect of stripe rust infections and nitrogen deficiency in wheat based on a Pearson correlation and simple regression algorithms. Meanwhile, notable challenges of analysing hyperspectral data such as the hyper-dimensionality cannot

be ignored (Liu et al., 2010; Pal, 2009; Hu et al., 2009). Despite the high dimensionality set, the above-cited works attest to the utility of field spectroscopy in detecting different agricultural crop diseases.

Considering the utility and robustness of field spectroscopy, it is hypothesised that integrating it with machine learning algorithms could facilitate an easy, timely and accurate method of discriminating agricultural crop diseases such as MSV (Adam et al., 2012). Machine learning algorithms such as random forest (RF), an ensemble learning procedure by Breiman, (2001) could discriminate various infection levels on crops such as maize. RF is based on the grouping of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001), designed to increase the discrimination process (Dube and Mutanga, 2015, De'Ath, 2007). When compared to other algorithms, random forest (i) is more accurate; (ii) has a unique procedure of selecting important variables; (iii) has the ability to process complex data affected by high dimensionality such as hyperspectral data; (iv) does not require data to be normally distributed and (v) it can deal with data with missing values (Cutler et al., 2007). Simplifying the high dimensionality of hyperspectral data has been indicated as a major success of random forest in remote sensing applications (Adam et al., 2012, Vincenzi et al., 2011, Abdel-Rahman et al., 2012, Adelabu and Dube, 2015). However, the challenge with random forest algorithm is that it does not automatically select the optimal number of variables that produce the best classification accuracy (Vincenzi et al., 2011). Several studies have shown that Random Forest has a preference to highly correlated predictor variable in identifying variables in high-dimensional spectral space (Strobl et al., 2008, Adjorlolo et al., 2013). Therefore, the guided regularized random forest (GRRF) proposed by Deng and Runger (2013) uses the importance scores from a RF built on the complete training data to complement the information gain in a local node and performs better than simple Random Forest. Consequently, GRRF has a better potential of selecting compact variables for discriminating different levels of MSV infection on maize.

Given that the MSV disease is a viral disease (Ward et al., 1996) and plant leaves have a well-known spectral signature, the integration of remote sensing technologies such as hyperspectral data and machine learning algorithms could be used to

identify spectral channels that could be used for early detection of the disease and also providing useful spatial information on the severity of infections. This information can also be handy in making decisions on the necessity and appropriate timing of agrochemical applications (Nutter Jr and Schultz, 1995). In that regard, this study tested whether field spectrometry measurements could discriminate between the various stages of MSV infection on maize and healthy maize leaves. More specifically the potential of hyperspectral data and GRRF machine learning algorithm was tested in detecting different levels of MSV infestation. Finally, the study sought to determine the optimum spectral bands that could be important for detecting MSV.

4.1 Methodology

4.1.1 Study Area

Field spectral measurements were taken at Ofcolaco experimental farm, located 60 km south of Tzaneen- in the Limpopo province, of South Africa (Figure 4.1). Maize was planted on the 25th of February 2015. Agronomists then visually assessed the stages of MSV disease across different growth stages of the maize prior to taking measurements (Sibiya et al., 2011).

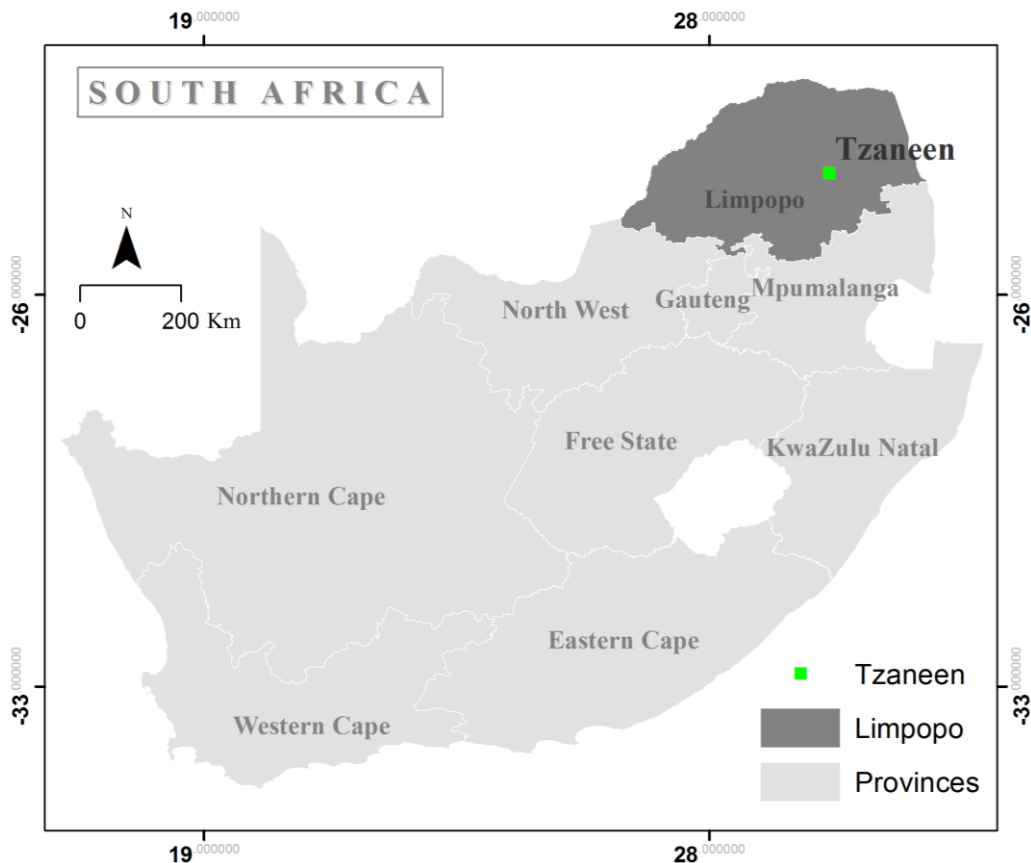


Figure 4.1 Location of the study area.

4.2 Maize leaves sampling

Leaves with less than 18% (Figure 4.2b) leaf area showing the first appearance of MSV symptoms were considered to be at a mild stage, while leaves with 19 to 42% (c) of leaf area with MSV symptoms were considered to be in the moderate stage of infection. Leaves with 43% to 100 % (d) of leaf area showing MSV symptoms were considered to be severely infected. These leaves were removed from maize plants and carried in a preservative container to a laboratory where their reflectance was measured within a few minutes of acquisition.

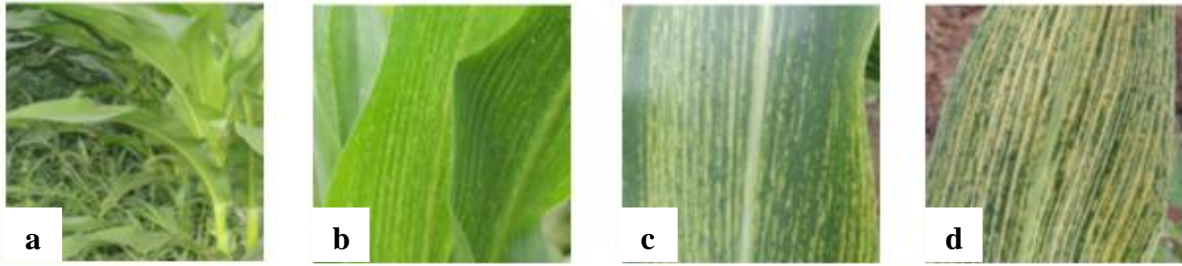


Figure 4.2 (a) Health, (b) Early, (c) Moderate and (d) Severe stages of maize streak virus infections on maize plants (courtesy Dhau, 2015).

4.3 Remote sensing of different levels of MSV infections

Spectral measurements of healthy, mild, moderate and severely infected maize leaves were conducted using the Analytical Spectral Devices (ASD) FieldSpec® 4 optical sensor (Analytical Spectral Devices, Inc., Boulder, CO, USA). In measuring the reflectance of maize at different MSV infections, a plant probe foreoptic with a leaf clip holder and an integrated light source was used in the laboratory. The contact probe foreoptic has a 10 mm field of view and an integrated 100W halogen reflector lamp. The instrument was warmed up for 90 minutes prior to measurement to increase the quality and homogeneity of spectral data. Instrument optimization and reflectance calibration were performed prior to sample acquisition. Furthermore, an average of 25 dark-current and 25 barium sulfate white reference (Spectralon, Labsphere, North Sutton, NH, USA) measurements were conducted at different intervals before and during the acquisition of maize reflectance.

The ASD FieldSpec® 4 spectrometer has a 350 - 2,500 nm spectral range, with 1.4 nm and 2 nm sampling interval for the UV/VNIR (350–1,000 nm) and SWIR (1,000–2,500 nm) regions, respectively. In that regard, at least 5 measurements were made from each leaf sample. Pre-processing to smooth the spectrum and reduce signal noise was not necessary, because reflectance spectra were assessed under constant light and temperature conditions with the plant probe foreoptic. One hundred (100) samples were taken for each disease status. A total of four hundred spectra were used in the spectral analysis (Figure 4.3). One of the most notable difficulties in hyperspectral data processing is the hyper-dimensionality of the data,

which requires sufficient training samples to simplify the complexity of classification and prediction processes (Hu 2009).

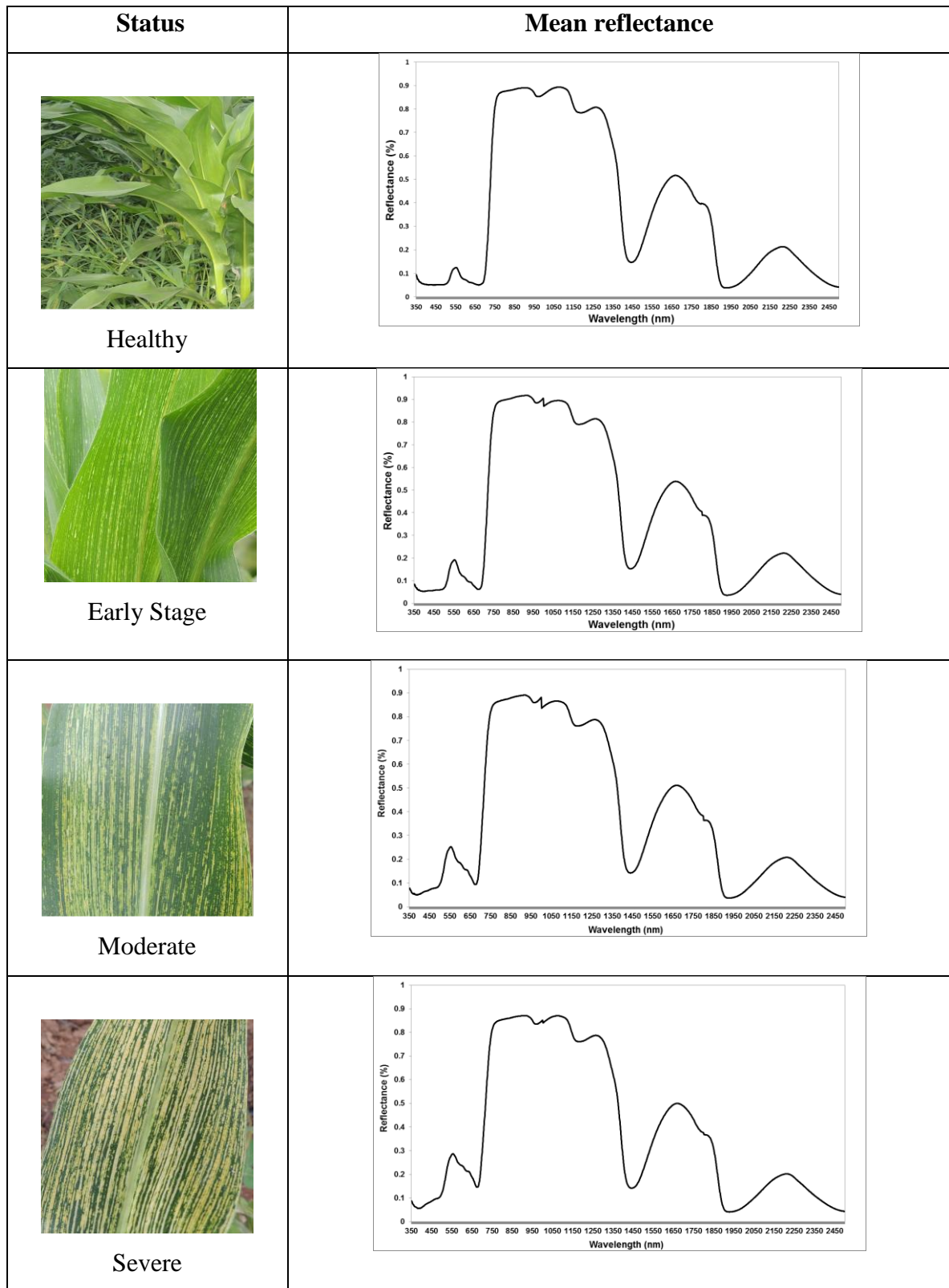


Figure 4.3 Field spectra for healthy, early, medium and severely infected maize leaves by the MSV

4.5 Discriminating against different levels of MSV infection using remotely sensed data.

Noise spectral channels between 904.5 and 994.5 nm, 1807.2 and 2027.7 nm and between 2182.4 and 2503.4 nm were discarded from the analysis. Thus, only 1825 wavelengths were used for the spectral analysis. To moderate high dimensionality on field measured spectra, discriminating healthy (HS) maize leaves from those that are at the early (ES) moderate (MS) and severe (SS) stages of MSV infection, Deng and Runger's (2013) guided regularized random forest (GRRF) algorithm was adopted for variable importance measurements and selection optimal variables.

4.6 Accuracy assessment

Prior to discrimination based on GRRF, data was split into training (70%) and testing (30%). The 30% testing sample was used in the 10-fold cross-validation procedure. A confusion matrix was also derived from the classification process. The overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) were derived using the confusion matrix. Furthermore, the kappa coefficient was also computed and used to evaluate the effectiveness of the GRRF algorithm in discriminating different levels of MSV infections on maize. If the kappa coefficients are equal to one or close to one, then there is perfect agreement (Safaralizade et al., 2014).

4.7 Results

4.7.1 Spectral separability of MSV disease infestation levels on maize

Figure four illustrates the spectral curves of healthy maize leaves in relation to those on the early, medium and severe stages of MSV infection. It can be observed that the four different levels of MSV infection can be discriminated effectively using the visible, red edge, near-infrared (NIR), shortwave NIR and slightly using the longwave NIR sections of the electromagnetic spectrum (Figure 4).

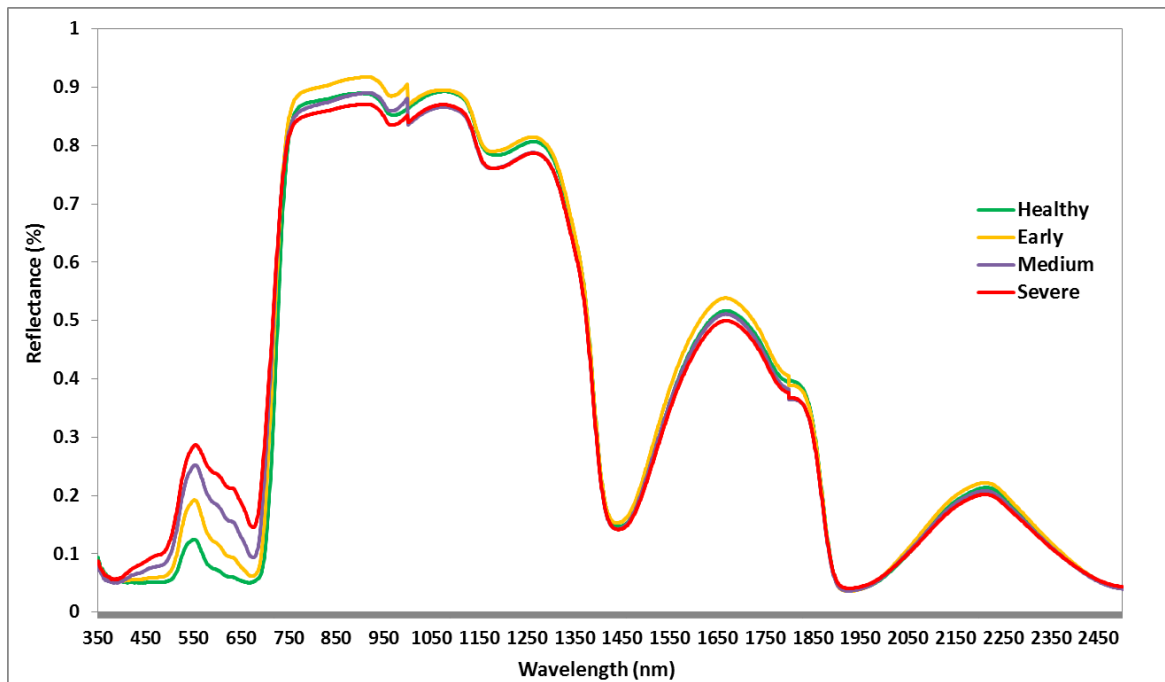


Figure 4.4 Spectral profiles of healthy, early, medium and severely infested maize leaves.

4.7.2 Selection of optimal variables for discriminating different levels of MSV infections.

GRRF selected wavebands from the visible through the red edge to the NIR regions of the electromagnetic spectrum for best discriminating healthy maize leaves from those in the mild stage, moderate and severely MSV infection stages (Figure 5). Specifically, GRRF identified five of 1825 narrow wavebands as optimal bands for detecting and discriminating different levels of MSV disease infections. Bands, 552 nm, 603 nm, 683 nm, 881 nm, and 2338 nm were selected by GRRF as the most important variables for distinguishing different levels of MSV infection although band 881 nm and 2338 nm yielded the least mean decrease in accuracy. The selected variables, 881 nm is close to 910 nm which associated with proteins (Curran, 1989), while 2338 nm was close to 2340 nm which associated with cellulose (Curran, 1989, Fourty et al., 1996, Kumar et al., 2001). The five selected wavebands were then used as input variables in random forest classifier for discriminating healthy maize leaves from those in the early stage, moderate and severely infected by MSV stages.

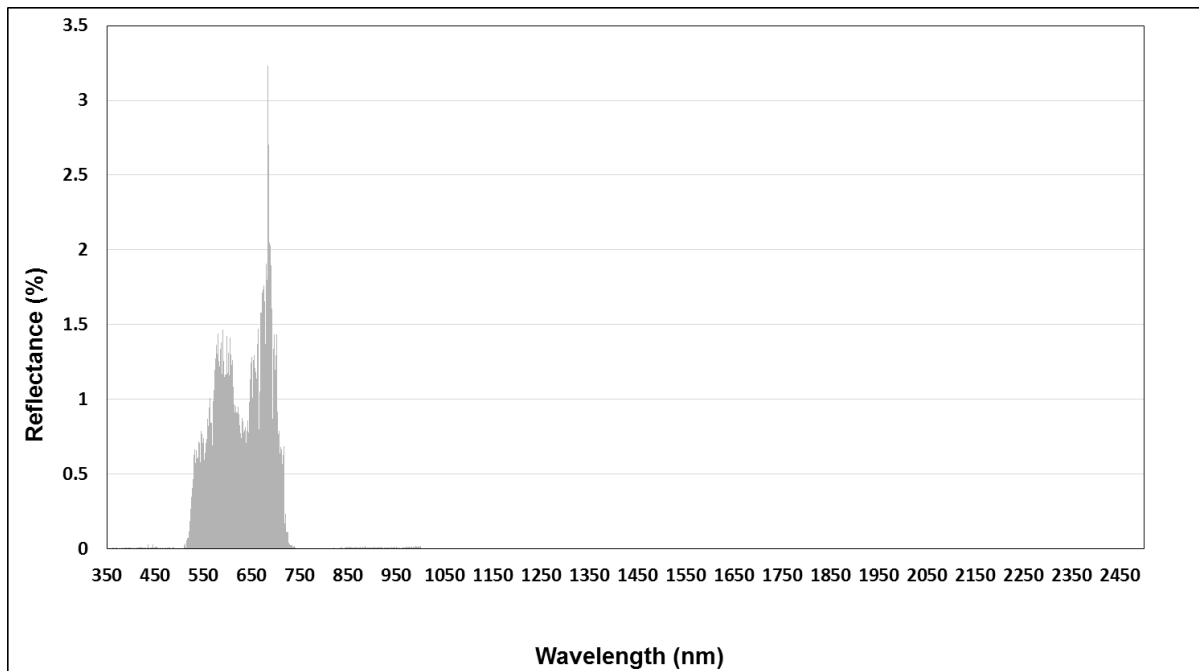


Figure 4.5 Important bands for detecting MSV on maize leaves.

4.7.3 Discrimination accuracy assessment

The optimal bands identified by the GRRF algorithm ($n = 5$) were used as input variables. The least error of (0.03) (3%) was attained based on the best combination of the 138 *mtry* and 9500 *n tree*. The discrimination model based on the 5 selected variables produced an Overall Accuracy of 95.83 % and a Kappa of 0.94. Table 4.1 illustrates the confusion matrix derived using the optimal variables selected by the GRRF algorithm in discriminating different levels of maize infections by MSV. Figure 6 illustrates high producer and user accuracies derived in discriminating the healthy, early, moderate, and severe stages of MSV infection in maize based on the optimal ($n = 5$) hyperspectral wavebands. It can be observed that healthy and moderate stage have the same user accuracy of (98%) whereas the early stage and the severe stage ranged between 90 and 94% respectively. On the other hand, the producer accuracy shows some slight differences for all the stages of infection.

Table 4.1: Confusion matrix derived based on the optimal variables selected by the GRRF algorithm in discriminating different levels of maize infections by MSV.

Class	HS	ES	MS	SS	TOTAL
H	29	0	0	0	29
E	1	30	2	0	33
M	0	0	26	0	26
S	0	0	2	30	32
TOTAL	30	30	30	30	120

HS is healthy, ES is the Early, MS is the moderate, and SS is the severe stage of MSV infection in maize.

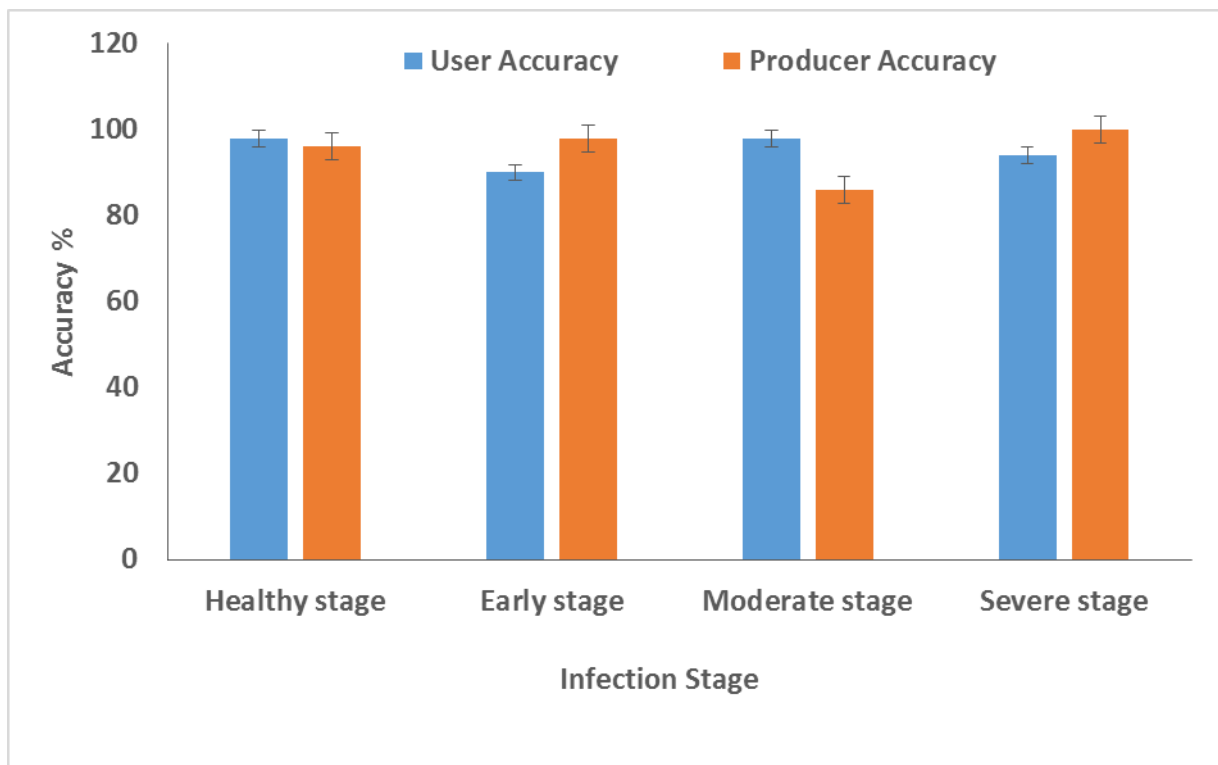


Figure 4.6 Producer and User accuracies derive based on the optimal variables in discriminating different levels of MSV infections.

4.8 Discussion

The essence of this study was to evaluate the utility of hyperspectral data in detecting different levels of MSV disease infection on maize. Furthermore, this study sought to identify optimal wavebands for discriminating the HS, ES, MS, and SS stages of MSV infection in maize. The utility of remotely sensed data in discriminating different levels of disease infestations such as MSV will help in detecting infections at an early stage. This is a critical step required in initiating mitigation and remedial intervention measures for preventing yield losses.

4.8.1 Remote sensing different levels of MSV infections on maize

Results of this work indicated that hyperspectral data could be used in distinguishing spectral differences between healthy maize plants and those that are; ES, MS, and SS infected by MSV. The GRRF algorithm selected 5 wavebands from the visible and the near infrared sections of the electromagnetic spectrum as optimal bands for discriminating the healthy, early, moderate, and severe stages of MSV infection in maize. The discrimination based on the wave bands from the visible section of the electromagnetic spectrum (552 nm, 603 nm, 683 nm) could be explained by the effect of MSV in maize plants, particularly the leaves. Specifically, the literature indicates that the MSV disease starts as tiny and spherical chlorotic spots on the leave of younger maize plants which are yellow in colour (Magenya et al., 2008, Bosque-Perez et al., 1998). These spots later amalgamate into continuous chlorotic streaks along the leaves and on the veins of the leaves. Meanwhile, on severely affected plants, the chlorotic lines combine into a pale green or yellow and sometimes white in colour on leaf surfaces. In this regard, the yellow colour of spots and stripes as well the pale green through to yellow and at times white chlorotic portions in the early and moderate infected maize leaves could explain the discrimination of these infection stages on the visible section of the electromagnetic spectrum. Considering that the selected optimal wavebands 560 nm, 603 nm, and 683 nm are close to Chlorophyll b absorption region (460 nm and 660 nm) (Curran, 1989, Kumar et al., 2001), the high chlorophyll content in healthy maize plants could also explain the optimal discrimination. Results of this study are consistent with Mahlein et al. (2013) who noted that the hyperspectral wave bands within the visible

section of the electromagnetic spectrum (including 552 nm, 603 nm, 683 nm) provided critical information for discriminating different levels of agricultural crop leaf diseases such as *Cercospora* leaf spot, sugar beet rust and powdery mildew in developing spectral indices for detecting and identifying plant diseases.

The other critical waveband that was selected by the GRRF algorithm in distinguishing spectral differences between maize plants which were HS, ES, MS and SS infection stages of MSV was 881 nm which falls within the NIR region of the electromagnetic spectrum. NIR is one of the regions that is renowned for discriminating and detecting various plant properties including the health of crops (Mariotto et al., 2013, Thenkabail et al., 2013). Healthy plants usually have a high NIR reflectance and varying reflectance in the short-wave infrared region (Liu et al., 2010). Similarly, the discrimination of different levels of MSV infection based on the NIR waveband could be explained by the reduced chlorophyll content resulting in lower reflectance of infected yellow, white and pale green maize plants when compared with the healthy maize plants reflecting high in the NIR region of the electromagnetic spectrum. Results of this study correspond to those of Liu et al. (2010) who noted that the NIR bands were influential in discriminating different levels of fungal infection in rice based on neural networks and principal components analysis. The final waveband that was selected by the GRRF algorithm as an optimal waveband for discriminating different levels of MSV was from the far near infrared section (FNIR) of the electromagnetic spectrum. Literature indicates that this section is associated with the cellulose absorption regions (Mariotto et al., 2013, Thenkabail et al., 2000). The optimal influence of the FNIR band 2338 nm which is close to 2340 nm associated with cellulose (Curran, 1989, Fourty et al., 1996, Kumar et al., 2001) could be have attributed to the variability in the cellulose concentration between healthy leaves and those that are infected by MSV.

4.9 Conclusions

This study sought to test whether field spectrometry measurements could discriminate between the various stages of MSV infection in maize. Specifically we tested the potential of hyperspectral data and GRRF machine learning algorithm in detecting healthy, early, moderate, and severely MSV infected maize plants. Finally, this study intended to determine the optimal spectral bands for detecting MSV. Based on the findings of this study it was concluded that:

- Remotely sensed data, particularly wavebands 552 nm, 603 nm, 683 nm, 881 nm, and 2338 nm were selected by GRRF as the most important variables for distinguishing different levels of MSV infection although band 881 and 2338 yielded the least mean decrease in accuracy.
- GRRF algorithm can be used to effectively detect and select optimal wavelengths urgently required to discriminate between different levels of MSV crop infections

The study underscores that remotely sensed data could offer a fast, accurate and effective approach urgently required for monitoring and forecasting agricultural crop disease epidemics such as MSV which has severely affected food production in Africa. The findings of this study are a stage towards a method for comprehensively monitoring and forecasting agricultural crop disease epidemics at a landscape scale required in building initiatives of improving food production in areas such as sub-Saharan Africa affected by frequent droughts, poverty, and limited resources. However, the findings of this study still need to be up scaled to sensor resolutions.

4.10 References

- ABDEL-RAHMAN, E., AHMED, F., ISMAIL, R., GONCALVES, J. & CORREIA, K. 2012. Random forest regression for sugarcane yield prediction in umfolozi, South Africa based on landsat TM and ETM+ derived spectral vegetation indices. *Sugarcane: Production, Cultivation and Uses*, 257-284, Volume 34.
- ADAM, E., MUTANGA, O., RUGEGE, D. & ISMAIL, R. 2012. Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing*, 33, 552-569.
- ADELABU, S. & DUBE, T. 2015. Employing ground and satellite-based QuickBird data and random forest to discriminate five tree species in a Southern African Woodland. *Geocarto International*, 30, 457-471.
- ADJORLOLO, C., MUTANGA, O., CHO, M. A. & ISMAIL, R. 2013. Spectral resampling based on user-defined inter-band correlation filter: C 3 and C 4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21, 535-544.
- ALEGBEJO, M., OLOJEDE, S., KASHINA, B. & ABO, M. 2002. Maize streak mastrevirus in Africa: distribution, transmission, epidemiology, economic significance and management strategies. *Journal of Sustainable Agriculture*, 19, 35-45.
- BAURIEGEL, E. & HERPPICH, W. B. 2014. Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to *Fusarium* spec. infections on wheat. *Agriculture*, 4, 32-57.
- BOSQUE-PEREZ, N., OLOJEDE, S. & BUDDENHAGEN, I. 1998. Effect of maize streak virus disease on the growth and yield of maize as influenced by varietal resistance levels and plant stage at time of challenge. *Euphytica*, 101, 307-317.
- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- CHEMURA, A., MUTANGA, O. & DUBE, T. 2017. Integrating age in the detection and mapping of incongruous patches in coffee (*Coffea arabica*) plantations using multi-temporal Landsat 8 NDVI anomalies. *International Journal of Applied Earth Observation and Geoinformation*, 57, 1-13.

- CURRAN, P. J. 1989. Remote sensing of foliar chemistry. *Remote sensing of Environment*, 30, 271-278.
- CUTLER, D. R., EDWARDS, T. C., BEARD, K. H., CUTLER, A., HESS, K. T., GIBSON, J. & LAWLER, J. J. 2007. Random forests for classification in ecology. *Ecology*, 88, 2783-2792.
- DE'ATH, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology*, 88, 243-251.
- DEERY, D., JIMENEZ-BERNI, J., JONES, H., SIRAUULT, X. & FURBANK, R. 2014. Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy*, 4, 349-379.
- DENG, H. & RUNGER, G. 2013. Gene selection with guided regularized random forest. *Pattern Recognition*, 46, 3483-3489.
- DEVADAS, R., LAMB, D., BACKHOUSE, D. & SIMPFENDORFER, S. 2015. Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat. *Precision agriculture*, 16, 477-491.
- DUBE, T. & MUTANGA, O. 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 36-46.
- FAJEMISIN, J., COOK, G., OKUSANYA, F. & SHOYINKA, S. 1976a. Maize streak epiphytotic in Nigeria [Virus diseases, Cicadulina mbila, insect vectors]. *Plant Disease Reporter*.
- FAJEMISIN, J., SHOYINKA, S., WILLIAMS, L., GORDON, D. & NAULT, L. Maize streak and other maize virus diseases in West Africa. Proceedings, International maize virus disease colloquium and workshop., 1976b. Ohio Agricultural Research and Development Center., 52-60.
- FOURTY, T., BARET, F., JACQUEMOUD, S., SCHMUCK, G. & VERDEBOUT, J. 1996. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote sensing of Environment*, 56, 104-117.
- FULLER, C. 1901. Mealic variegation. *First Report of the Government Entomologist Natal 1899-1900*, 17-19.

- GUTHRIE, E. Virus diseases of maize in East Africa. International Maize Virus Disease Colloquium and Workshop. Wooster, Ohio (USA). 16-19 Aug 1976., 1977.
- HU, C.-C., HSU, Y.-H. & LIN, N.-S. 2009. Satellite RNAs and satellite viruses of plants. *Viruses*, 1, 1325-1350.
- KUMAR, L., SCHMIDT, K., DURY, S. & SKIDMORE, A. 2001. Imaging spectrometry and vegetation science. *Imaging spectrometry*. Springer.
- LIU, Z.-Y., WU, H.-F. & HUANG, J.-F. 2010. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. *Computers and Electronics in Agriculture*, 72, 99-106.
- MAGENYA, O., MUEKE, J. & OMWEGA, C. 2008. Significance and transmission of maize streak virus disease in Africa and options for management: a review. *African Journal of Biotechnology*, 7, 4897-4910.
- MAHLEIN, A.-K., RUMPF, T., WELKE, P., DEHNE, H.-W., PLÜMER, L., STEINER, U. & OERKE, E.-C. 2013. Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment*, 128, 21-30.
- MARIOTTO, I., THENKABAIL, P. S., HUETE, A., SLONECKER, E. T. & PLATONOV, A. 2013. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HypSIRI mission. *Remote Sensing of Environment*, 139, 291-305.
- MCCLEAN, A. P. D. "STREAK DISEASE OF SUGAR CANE." Proc. Int. Soc. Sugar Cane Technol. Vol. 5. 1935.
- NUTTER JR, F. W. & SCHULTZ, P. M. 1995. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Canadian Journal of Plant Pathology*, 17, 174-184.
- OLANIYAN, A. B. 2015. Maize: Panacea for hunger in Nigeria. *African Journal of Plant Science*, 9, 155-174.
- PAL, M. 2009. Extreme-learning-machine-based land cover classification. *International Journal of Remote Sensing*, 30, 3835-3841.
- RAJI, S. N., SUBHASH, N., RAVI, V., SARAVANAN, R., MOHANAN, C. N., MAKESHKUMAR, T. & NITA, S. 2016. Detection and Classification of Mosaic Virus Disease in Cassava Plants by Proximal Sensing of Photochemical

- Reflectance Index. *Journal of the Indian Society of Remote Sensing*, 44, 875-883.
- SAFARALIZADE, E., HUSSEINZADE, R., PASHAZADE, G. and KHOSRAVI, B., 2014. Assessing the accuracy of the pixel-based algorithms in classifying the urban land use, using the multi spectral image of the IKONOS satellite (Case study, Uromia city). *International Letters of Natural Sciences*, 6.
- SHEPHERD, D. N., MARTIN, D. P., VAN DER WALT, E., DENT, K., VARSANI, A. & RYBICKI, E. P. 2010. Maize streak virus: an old and complex 'emerging' pathogen. *Molecular plant pathology*, 11, 1-12.
- SIBIYA, J., TONGOONA, P., DERERA, J., VAN RIJ, N. & MAKANDA, I. 2011. Combining ability analysis for *Phaeosphaeria* leaf spot resistance and grain yield in tropical advanced maize inbred lines. *Field crops research*, 120, 86-93.
- STROBL, C., BOULESTEIX, A.-L., KNEIB, T., AUGUSTIN, T. & ZEILEIS, A. 2008. Conditional variable importance for random forests. *BMC bioinformatics*, 9, 307.
- THENKABAIL, P. S., MARIOTTO, I., GUMMA, M. K., MIDDLETON, E. M., LANDIS, D. R. & HUENNRICH, K. F. 2013. Selection of hyperspectral narrowbands (HNBS) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6, 427-439.
- THENKABAIL, P. S., SMITH, R. B. & DE PAUW, E. 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote sensing of Environment*, 71, 158-182.
- THOTTAPPILLY, G., BOSQUE-PÉREZ, N. & ROSSEL, H. 1993. Viruses and virus diseases of maize in tropical Africa. *Plant Pathology*, 42, 494-509.
- VINCENZI, S., ZUCCHETTA, M., FRANZOI, P., PELLIZZATO, M., PRANOVI, F., DE LEO, G. A. & TORRICELLI, P. 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecological Modelling*, 222, 1471-1478.

- WARD, J., HOHLS, T., LAING, M. & RIJKENBERG, F. 1996. Fungicide responses of maize hybrids to grey leaf spot. *European journal of plant pathology*, 102, 765-771.
- WILLIAMS, P. J., GELADI, P., BRITZ, T. J. & MANLEY, M. 2012. Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, 55, 272-278.
- YUAN, L., HUANG, Y., LORAAMM, R. W., NIE, C., WANG, J. & ZHANG, J. 2014. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Research*, 156, 199-207.
- YUAN, L., ZHANG, H., ZHANG, Y., XING, C. & BAO, Z. 2017. Feasibility assessment of multi-spectral satellite sensors in monitoring and discriminating wheat diseases and insects. *Optik-International Journal for Light and Electron Optics*, 131, 598-608.

Chapter 5

UNDERSTANDING THE SUBTLE SPATIAL DISTRIBUTION OF MSV DISEASE USING HIGH MULTISPECTRAL RESOLUTION DATA

This chapter is based on:

Inos Dhau, Elhadi Adam, Kingsley K. Ayisi & Onesimo Mutanga (2018) Detection and mapping of maize streak virus using RapidEye satellite imagery, Geocarto International, DOI: [10.1080/10106049.2018.1450448](https://doi.org/10.1080/10106049.2018.1450448)

Abstract

Maize streak virus is one of the most prevalent plant diseases which affects maize productivity in Africa. In South Africa, the disease is common in Limpopo, Northern Cape, KwaZulu-Natal, and Mpumalanga Provinces. The aim of this study was to detect and map maize streak virus using RapidEye multispectral sensor in Ofcolaco farm. To achieve this objective, the acquired RapidEye sensor covering the study area was classified using the robust Random Forest algorithm. Furthermore, the variable importance technique was used to determine the influence of each spectral band and indices on the mapping accuracy. For better performance of image data, commonly used vegetation indices were tested if they can significantly improve the classification accuracy. The results revealed that the use of RapidEye spectral bands in detection and mapping of maize streak virus yielded good classification results with an overall accuracy of 82.75%. The inclusion of vegetation indices computed from RapidEye sensor improved the classification accuracies by 3.4%. The most important RapidEye spectral bands in classifying maize streak virus were near infrared, blue and red-edge. On the other hand, the most important vegetation indices were the Soil adjusted vegetation index (SAVI), Enhanced Vegetation Index (EVI), Red index (RI) and Normalized Vegetation Index (NDVI). The current study recommends future studies to assess the importance of multi-temporal remote sensing applications in detecting and monitoring the spread of maize streak virus.

Keywords: Remote sensing, detection, mapping, RapidEye, vegetation indices.

5.0 Introduction

Maize streak virus (MSV) causes a plant disease, known as maize streak disease. It is an insect-transmitted maize pathogen that is endemic in sub-Saharan Africa (Martin et al., 2009). Although the origins of MSV is difficult to trace, the disease is suspected to have originated from pre grasses and eventually spread to maize by infected grass and leafhoppers (Redinbaugh et al., 2002, Nault and Ammar, 1989). Research findings reported by the International Service for the Acquisition of Agri-biotech Application (ISAAA) revealed maize streak virus has been a problem that requires immediate attention to many subsistence farmers in Africa. Farmers have been struggling to control maize streak virus and consequently, they have incurred high yield losses especially during severe infection periods. (Martin and Shepherd, 2009). Since leafhoppers prefer shade, farmers have been cultivating open lands without dense vegetation in an attempt to reduce the spread of maize streak virus infection Redinbaugh et al., (2002). The effective control and management of MSV in crop production is determined by the knowledge of the farmers to accurately detect the spatial distribution of the disease at an earlier stage. There is a need for farmers to develop adequate technology to identify and quantify all the sections of the farms infected by MSV. However, remote sensing technology has the potential to detect and monitor the spatial and temporal distribution of MSV, therefore assisting relevant stakeholders in the agricultural sector to implement appropriate management strategies.

The applications of remote sensing technology in providing reliable environmental data has been tested and reviewed in various disciplines such as water resources management (Shoko et al., 2015, Bastiaanssen et al., 2005), invasive species monitoring (Matongera et al., 2016, Singh et al., 2014) and biomass studies (Dube and Mutanga, 2015); (Drake et al., 2003). Similarly, remotely sensed data may be used as an indispensable tool for assessing and monitoring crop diseases and crop conditions. Research studies have revealed that medium spatial resolution multispectral sensors can be utilized in detection and monitoring of crop diseases severity (Ashourloo et al., 2014, Kranz and Rotem, 2012). The aforementioned studies used broadband multispectral data such as Landsat to detect plant diseases. However, the application of these data sets have been limited by their coarse spatial

and low radiometric resolution. On the other hand, the recently launched fine spatial resolution satellite imagery like RapidEye presents a powerful combination of characteristics within one sensor, which comprise of enormous swath width, improved temporal and radiometric resolution, and supplementary spectral bands. The aforementioned sensor offers a great potential for detection and mapping of crop infections, owing to its fine spatial resolution (6.5 m pixel size), a swath width of 77 km, and an acquisition capacity of 1500km per orbit (RapidEye, 2011b). Furthermore, the sensor has five strategically positioned spectral bands that are situated within the visible to near-infrared parts of the electromagnetic spectrum. The sensor also comprises of valuable additional spectral data from the red-edge band that is located in the wavelength range of 690–730 nm. However, the successful application of remote sensing satellite imagery in detecting plant diseases also depends on the classification algorithm chosen. Recently, robust classifiers such as Random Forest algorithm have emerged.

The robust Random Forest algorithm has a powerful group of tree classifiers in which each classifiers is created by means of a simple unsystematic vector which is then sampled autonomously using the original input vector (Pal, 2005, Breiman, 1999). Therefore, every single tree contributes one vote towards determining the final class of an object. Additionally, the algorithm has the ability to select the optimum variables that were used during a classification. The variable importance technique is essential in determining the influence of specific RapidEye bands and derived vegetation indices on the overall classification accuracy of maize streak virus. Research studies have shown that the performance of remotely sensed data can be improved by the inclusion of vegetation indices (Matongera et al., 2017, Ramoelo et al., 2015, Huete et al., 2002). Satellite derived vegetation indices are popular for being used as proxies of vegetation greenness (Rodríguez-Pérez et al., 2007). Vegetation indices used in this study are designed considering the reflectance property of infected crops, because the reflectance obtained due to interaction of the electromagnetic radiation is affected by the chemical and physical characteristics of targeted objects. The vegetation indices are built on the observation that various pigments and chemical composition of the crop properties have strong response on satellite data (Joshi, 2011).

Although several studies have successfully detected crop diseases and stress using remotely sensed data, to the best of my knowledge, the utility of RapidEye multispectral sensor and derived vegetation indices combined with the robust Random Forest algorithm in detection and mapping of maize streak virus in South Africa has not yet been tested. In this regard, this study hypothesized that RapidEye and derived vegetation indices have the potential to an invaluable data base for remote sensing of maize streak virus mapping.

5.1 Materials and Methods

5.1.1 Description of the study area

The research was done at the 20ha Ofcolaco Farm located 60km south of Tzaneen- in the Limpopo province, South Africa (Figure 5.1). Geographically, the Ofcolaco Farm is located between S 24° 4' 59.99" and E 30° 22' 0.01". In terms of the climatic characteristic, the study area is mainly subtropical, with warm summers and mild winters. The Ofcolaco farm is predominantly in the semi-arid region. In summer, daily temperatures range from about 25°C and 26°C. Research has shown that in Limpopo, rainfall is seasonal with 95% of the precipitation received from October to March (M'marete, 2003). Furthermore, the farm is situated within the Limpopo metamorphic belt which joins the Kaapvaal craton to the south with the Zimbabwe craton to the north. The belt consists of high-grade metamorphic rocks that have undergone a long cycle of high temperature and pressure deformation (Lal, 2001). The farm is underlain by gneiss, quartzite and granulite of the Malala Drift Group of the Beit Bridge Complex which forms part of the Limpopo belt (Barton et al., 2006). These rocks have been intruded locally by younger diabase dykes. Limpopo Province produces a diversity of agricultural products which range from fruits, vegetables and cereals (Aphane, 2011). The two distinctive agricultural productions systems in Limpopo are commercial and subsistence farming which is being practiced by the majority.

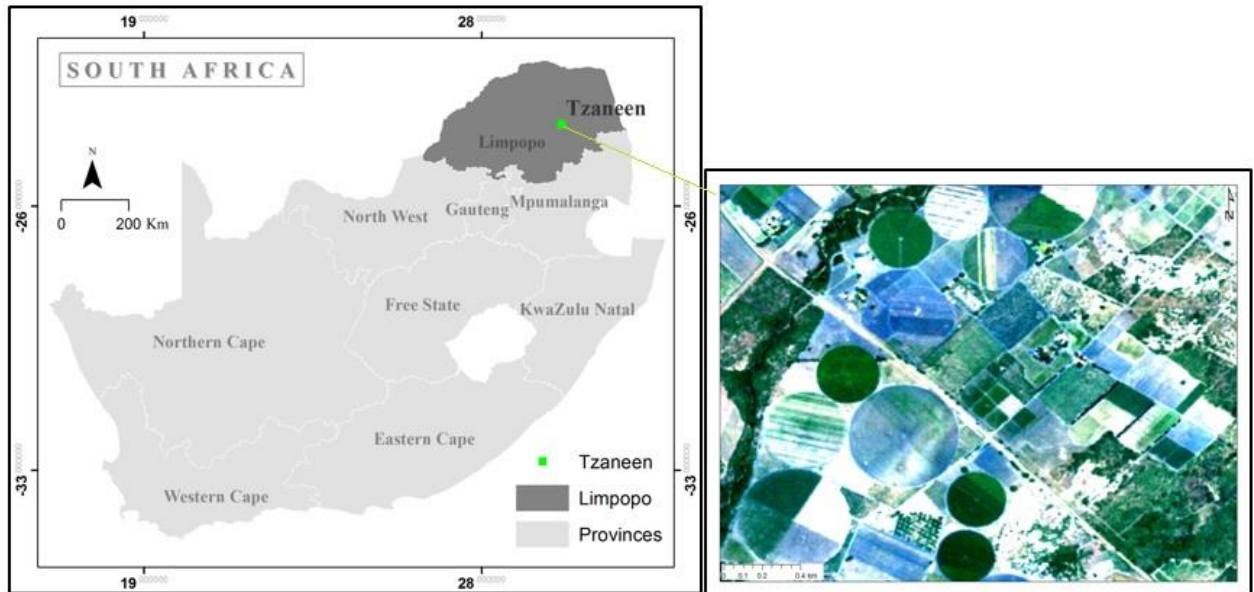


Figure 5.1 Location of the study area

5.2 Field data collection

The ancillary data was collected to compliment remote sensing data. Data collection was conducted to record healthy and disease stricken maize, using a hand held Leica GS20 Global positioning system (GPS) with a sub meter accuracy. The field data collection was conducted in January 2016. The period at which the field data was collected coincided with the image acquisition dates. Purposive sampling strategy was adopted in choosing sampling sites within the farm. Two sampling sites were chosen, one site with healthy maize field (NIM) and the other with MSV infected maize (IM). ArcGIS 10.3 was used to generate random sampling points within the two sampling sites. The other land cover classes which were identified and their locations recorded include riverine vegetation (VEG), grassland (GR), fruit trees (FT) and bare soil (BS). An average of 65 GPS locations were recorded per landcover class. Therefore, a total of 390 sample points were used in this study. The GPS-measured locations were recorded in a table format in Microsoft excel and then converted into point map in ArcGIS 10.3 software. The ancillary data was divided into 70% training and 30% for validation data sets.

5.3 Satellite data acquisition and pre-processing

The RapidEye image which covers Ofcolaco Farm was obtained in January 2016. By description, RapidEye is an assemblage of five unique satellite sensors that have the potential to acquire high spatial resolution and large swath width remotely sensed data daily. An orthorectified product (Level 3A) was provided for the study with corrected radiometry and geometry. For the orthorectification process, an accuracy of 1 or less pixel was attained (RapidEye, 2011a). At the time of delivery, the RapidEye image was already resampled to a 5 m x 5 m spatial resolution. Due to flat terrain of the study area, no terrain related corrections were performed. For surface reflectance retrieval, atmospheric correction was performed through the topographic and atmospheric software (ATCOR2).

5.4 Spectral analysis and characterization

The evaluation of the separability of infected and non-infected maize (figure 5.2) in this study was done with the transformed divergence separability index (TDSI). Reflectance values extracted from the land cover types based on RapidEye satellite data were used in conducting the TDSI. Specifically, the TDSI index was used in this study based on its previous performance in other studies (Matongera et al., 2017, Chemura and Mutanga, 2016, Huang et al., 2016). The TDSI was calculated using the following formula:

$$TDSI = (\sqrt{2(1 - e^{-\alpha})})^2 \quad \text{Equation (1)}$$

where α is the Bhattacharya distance (Equation (2))

$$\alpha = \frac{1}{8} (\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{i^{(C_i + C_j)}}{\sqrt{|C_i| |C_j|}} \right) \quad \text{Equation (2)}$$

Where i and j are the two separate signature classes being compared, C_i represents the covariance matrix of the signature i , μ_i is the mean vector of vector i and $|C_i|$ is the determinant of C_i (Asmala, 2012). According to the TDSI principle, values which are more than 1.9 demonstrates that those spectral classes can easily be separated

while values which falls below 1.0 indicate that classes are not statistically different in order to achieve an accurate land cover classification.

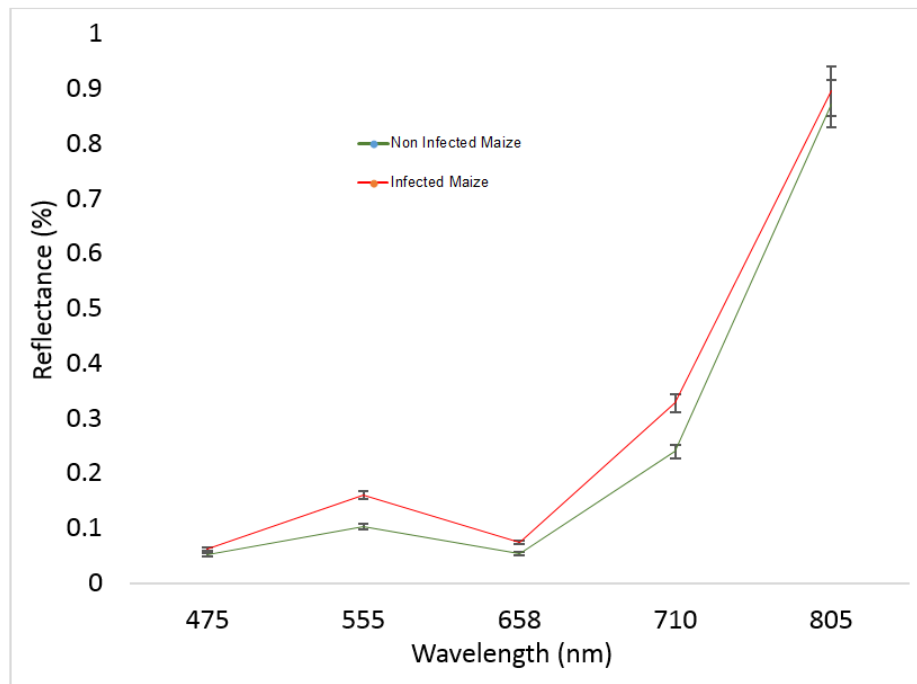


Figure 5.2 Spectral characteristics of infected and non-infected maize

5.5 Image classification and accuracy assessment

To discriminate the infected maize crops from non-infected maize crops and other landcover classes, the robust and advanced Random Forest (RF) algorithm was used to classify RapidEye imagery. Several research studies have extensively used the RF algorithm for classifying remotely sensed data (Odindi et al., 2014, Gislason et al., 2004). Basically, an individual tree represents a single vote when deciding the final class of an object under classification. During the classification process, RF constructs various classification trees using bootstrap models that use substitutes that are drawn from the provided data set. In this regard, the RF library that was developed in R statistical software was used to run the RF classifier and the classification maps prepared in ENVI 4.8. RF also provides variable importance measurement, which is a measure of the contribution of each variable (band) in the final predictive model. In this study, variable importance measurement was used to assess the importance of each band and Vegetation Indices in detecting MSV infection. In order to compare the original class with the class that have been

classified by the RF classifier a confusion matrix was generated giving overall, user and producer accuracies.

5.6 Improving classification using vegetation indices

Vegetation indices were calculated from the RapidEye image to improve the overall classification accuracy. Various vegetation indices have been designed and have their specific utility for vegetation mapping (Joshi, 2011). Vegetation indices used in this study did not only help to discriminate maize streak virus but are also essential in identifying and mapping of other land cover classes. An overlay was conducted between the point map of GPS locations of land cover classes, the remotely sensed data and their derivatives, in ENVI 4.8. The extracted spectral data was then exported from ENVI in a tabular format in notepad. The extracted spectral data was exported to excel for Vegetation Indices calculation (Table 5.1).

Table 5.1: Summary of RapidEye derived indices used in this study

Vegetation Index	Equation	Reference
Normalised difference vegetation index <i>a</i>	$\text{NIR} - \text{Red} / \text{NIR} + \text{Red}$	(Rouse Jr et al., 1974)
Normalised difference vegetation index <i>b</i>	$\text{NIR} - \text{Red edge} / \text{NIR} + \text{Red edge}$	(Rouse Jr et al., 1974)
Simple ratio	NIR / Red	(Gitelson and Merzlyak, 1994)
Soil adjusted vegetation index	$\{(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + \text{L})\} (1 + \text{L})$	(Huete, 1988)
Enhanced vegetation index	$2.5 \{(\text{NIR} - \text{Red}) / (\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{blue} + 1)\}$	(Huete et al., 1999)
Green index	$(\text{NIR} / \text{Green}) - 1 / (\text{NIR} / \text{Red}) - 1$	(Gitelson et al., 2002)
Red index		(Gitelson et al., 2002)

5.7 Results

5.7.1 Spectral separability of land cover classes

Spectral separability analysis results revealed that the land cover classes under study were spectrally separable (TDS > 1.0) meaning that they could be successfully discriminated using RapidEye satellite imagery. The TDS results for all the classes under study are presented in Table 5.2. It can be observed that infected maize and non-infected maize can be spectrally separated using RapidEye data with the TDS above one.

Table 5.2: RapidEye transformed divergence indices showing interclass separability of land cover classes

	IM	NIM	VEG	GR	FT	BS
IM	-----	1.13	1.44	1.35	1.98	1.87
NIM	1.13	-----	1.33	1.28	1.96	1.88
VEG	1.44	1.33	-----	1.19	1.54	1.66
GR	1.35	1.28	1.19	-----	1.89	1.98
FT	1.98	1.96	1.54	1.89	-----	2.10
BS	1.87	1.88	1.66	1.98	2.10	-----

5.7.2 Image classification

5.7.2.1 Classification results using RapidEye spectral bands

The classification results attained using RapidEye spectral bands data set are shown in Table 5.3. As illustrated in the table, the results revealed that the use of RapidEye spectral bands in detection and mapping of maize streak virus yielded good classification results with an overall accuracy of 82.75%. Similarly, good user and producer accuracies were also reported. On average, user accuracies were slightly higher than producer accuracies by 3.4%. Of all the six classes under study, grassland produced the least producer (68.10%) and user (72.58%) accuracies. On the other hand, the fruit trees class had the highest producer (89.81%) and user (94.78%) accuracies. In this regard, RapidEye derived spectral bands demonstrated the high potential of detecting and mapping of maize streak virus.

5.7.2.2 Classification results using computed vegetation indices

The classification results for calculated vegetation indices are presented Table 5.3. The results revealed that the addition of vegetation indices in detection and mapping of maize streak virus produced improved classification results with an overall accuracy of 86.10%. When compared to the use of raw spectral bands, the inclusion of vegetation indices improved overall accuracy by 3.6%, whereas for producer and user accuracies the improvement was slightly marginal ranging between 0 and 2.5%. Correspondingly, the inclusion of vegetation indices also reported improved user and producer accuracies. Similar to the results obtained from the use of vegetation indices, user accuracies were slightly higher than producer accuracies by 3.3%. Although they were above average, the grassland class produced the least producer (69.31%) and user (74.78%) accuracies. The fruit trees class had the highest producer (94.45%) and user (96.84%) accuracies. Therefore, computed vegetation indices improved classification accuracy results in detection and mapping of maize streak virus.

Table 5.3: Classification accuracies (%) for RapidEye derived spectral bands and computed vegetation indices

	Spectral bands		Vegetation indices	
	PA	UA	PA	UA
Infected maize	74.21	76.11	76.89	81.41
Non-infected maize	75.10	78.74	74.73	79.21
Riverine vegetation	87.41	89.01	87.10	87.78
Grassland	68.10	72.58	69.31	74.78
Fruit trees	89.81	94.78	94.45	96.84
Bare soil	86.25	90.17	91.27	94.13
OA	82.75		86.10	

(PA= Producer Accuracy, UA= User Accuracy and OA =Overall Accuracy)

5.7.3 Variable importance ranking

To generate the optimum variables for the classification, the average decline in overall accuracy was executed to determine the significance of all RapidEye spectral bands (n=5) in detection and mapping of MSV. The robust Random Forest classifier determined the most significant spectral bands for the classification (Figure 5.3a). The most important RapidEye spectral bands were near infrared (760-850 nm), blue (440-510 nm) and red-edge (690-730 nm). Furthermore, the variable importance measurement was also used to rank the importance of vegetation indices used in this study. The Random Forest classifier successfully measured the importance of the vegetation indices in detecting and mapping. The significance of each vegetation index computed from RapidEye spectral bands in detecting and mapping maize streak virus are presented in Figure 5.3b. The most important vegetation indices were; Soil adjusted vegetation index (SAVI), Enhanced vegetation index (EVI), Red index (RI) and Normalized Vegetation Index (NDVI).

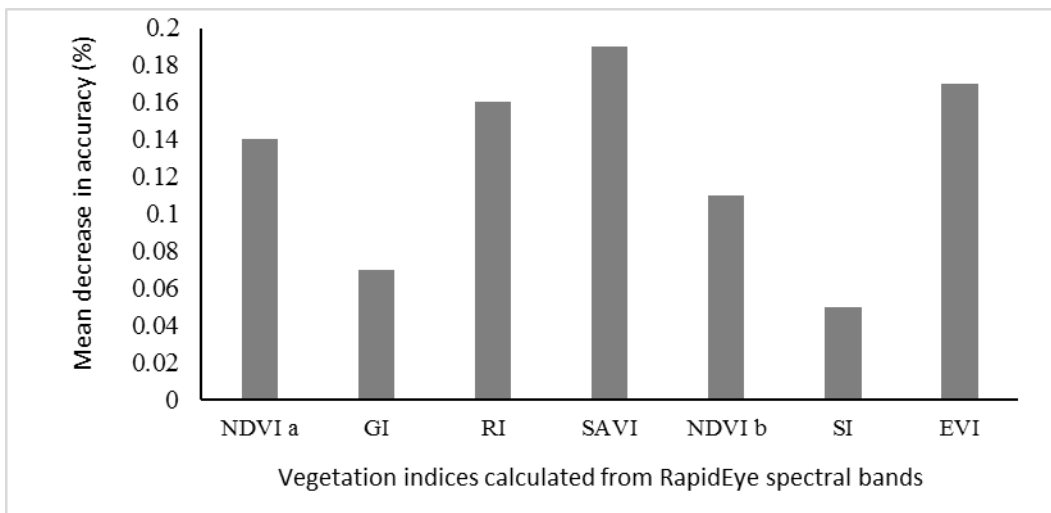
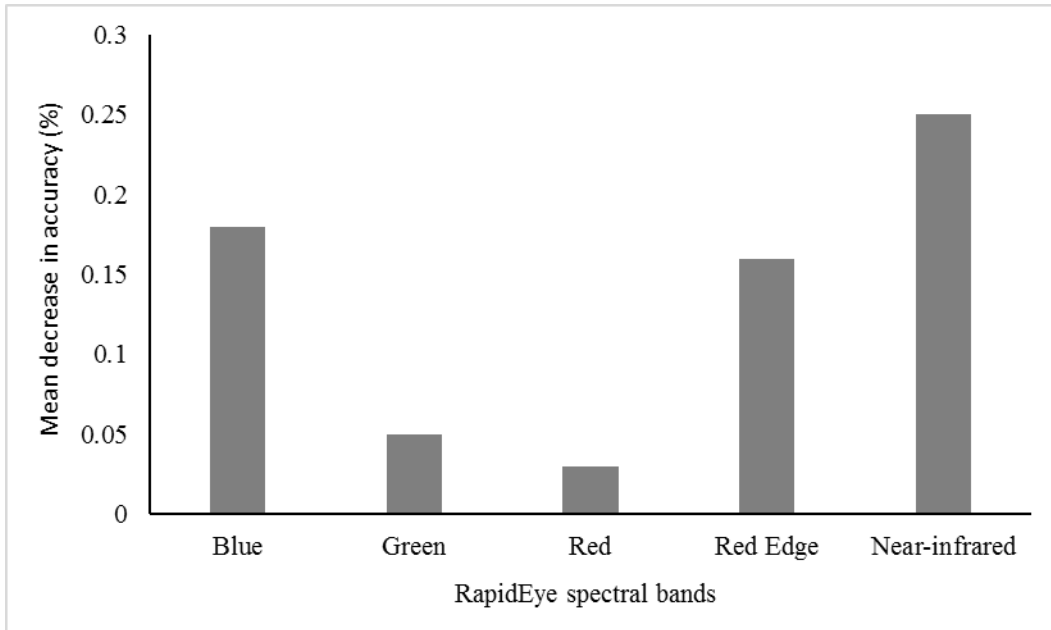


Figure 5.3 a and b: The importance of RapidEye spectral bands and derived vegetation indices in detection and mapping MSV

5.7.4 Spatial distribution of MSV infected maize and other land cover classes

The spatial distribution of MSV infected maize and other land cover classes at Ofcolaco farm is shown in Figure 5.4. The land cover map shows that Maize Streak Virus was widely distributed in the fields, especially in the North-eastern sections of the map. According to the spatial distribution map, MSV hardly co-exist with healthy maize. The MSV disease spreads and infects nearby healthy maize and becomes concentrated in infected fields. Other major land cover classes significant in Ofcolaco farm include riverine vegetation and fruit trees.

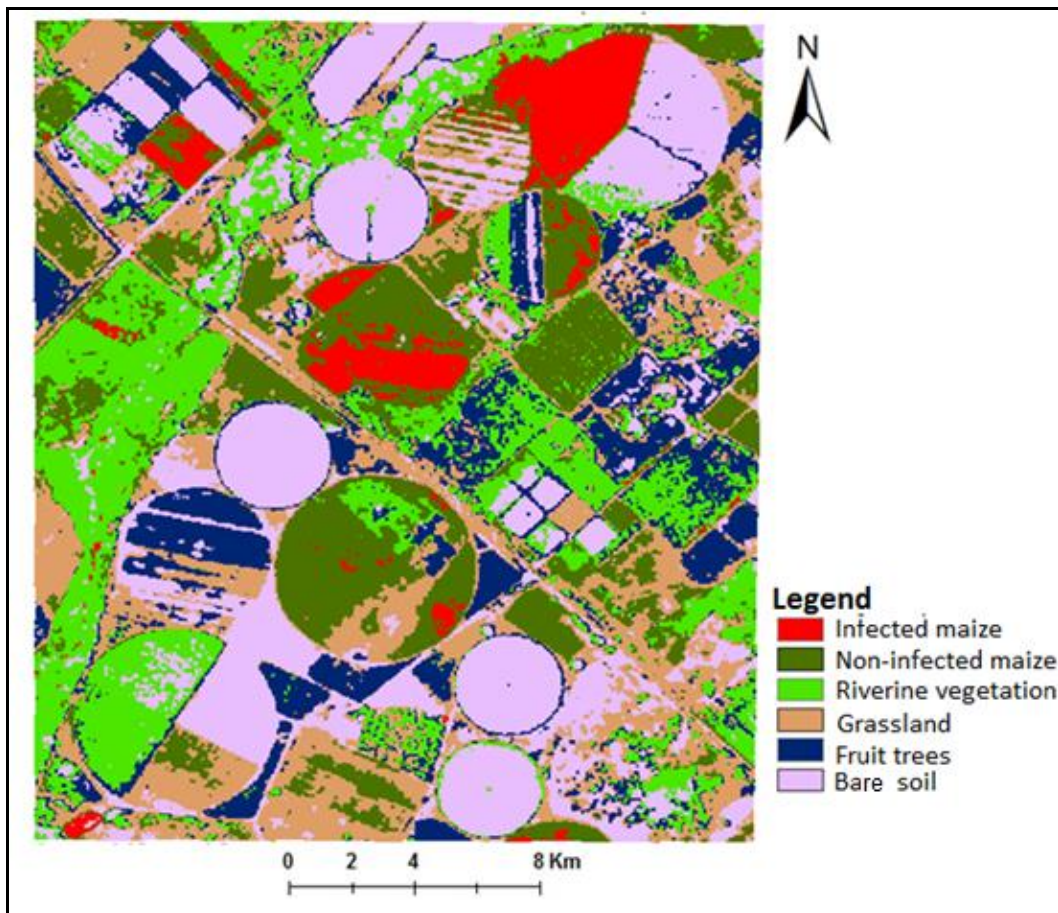


Figure 5.4 Land cover classification map of infected maize and other classes

5.8 Discussion

The maize streak virus (MSV) was successfully detected and mapped using RapidEye multispectral imagery in the current research study. The inclusion of satellite based vegetation indices improved the classification accuracy. The classification process using RapidEye spectral bands yielded 82.75%, whilst the computed vegetation indices improved the classification accuracy by 3.4%. This observation is also confirmed by the TDSI which indicated that the infected maize can be spectrally separated with a TDS greater than one. For instance, literature shows that any separability test with a TDS value greater than one can be spectrally separated from other classes (Matongera et al 2017). The high performance of RapidEye sensor in detecting and mapping maize streak virus in Ofcolaco farm could be possibly be credited to RapidEye's improved spatial and radiometric resolution and its distinctive design. For example, RapidEye sensor has high spatial resolution (5m) which is vital in detecting MSV even in its early stages. Additionally, the strategically located RapidEye red edge band is important in vegetation discrimination. The strategically located red edge band entrenched in the sensor was vital in the detection of maize streak virus in this study. Furthermore, the significance of the red edge band gives the recently launched modern satellite sensors an advantage of quantifying biochemical properties of plants absorptions at a large spatial scale (Ramoelo et al., 2015, Eitel et al., 2011).

The utility of RapidEye sensor in disease detection, mapping and monitoring was also successfully tested in several other studies (Ortiz et al., 2013, Marx, 2010, Marx et al., 2010, Eitel et al., 2011). The results reported in this study concur with research findings reported by the aforementioned studies. For example, (Marx et al., 2010) detected beetle infestations in spruce stands using multi-temporal RapidEye sensor. Their findings revealed that the use of multi-temporal RapidEye imagery, ground truth data of bark beetle infestation and the application of data mining techniques, enabled the recognition and differentiation of the infestation stages. The analysis reveals a weak trend for identifying newly infested groups of trees, which are still widely green. Similarly, Eitel et al. (2011) established that the red edge spectral band significantly improves early plant stress detection in a New Mexico conifer forest. The similarity of these previous findings with those established in this study mainly

emphasises the importance of the strategically located RapidEye red edge spectral band.

The inclusion of vegetation indices computed from RapidEye sensor improved the classification accuracies. The most important vegetation indices in detecting and mapping maize streak virus were the Soil Adjusted Vegetation Index (SAVI), Enhanced Vegetation Index (EVI), Red Index (RI) and Normalized Vegetation Index (NDVI). The SAVI was essential in this study because of its ability to reduce soil brightness from the spectral vegetation indices that include red and NIR bands (Huete, 1988). On the other hand, the commonly used NDVI was also influential in detecting maize streak virus in this study. The NDVI was sensitive to maize streak virus probably because at a later stage of the virus the crop will be identified by low chlorophyll content and subsequently changes in leaf area. The responsiveness of NDVI in crop stress and disease detection has also been previously reported by (Eitel et al., 2011). More recently, Random Forest algorithm has been ranked as a powerful classification algorithm in remote sensing studies (Belgiu and Drăguț, 2016, Maxwell et al., 2016, Peerbhay et al., 2016). Findings from this study endorse Random Forest's robustness and accuracy. Results in this study are consistent with previous studies that reported Random Forest's exceptional technique which identifies key variables which enhances classification results.

The spatial distribution of MSV infected maize and other land cover classes in Ofcolaco farm shows maize infected with maize streak virus was widely distributed in the fields, especially in the North-eastern sections of the map. According to the spatial distribution map, MSV hardly co-exist with healthy maize. In most cases, the co-existence will eventually lead to infection of non-infected crops. Generally, based on the results from this study, RapidEye sensor which provides red-edge data presents a great potential for the development of early warning systems for detecting, mapping and monitoring crop infections and stress.

5.9 Conclusion

The current study focused on the detection and mapping of maize streak virus using RapidEye multispectral imagery. In this regard, findings from this study concluded that RapidEye multispectral imagery can successfully be used in detection and mapping of MSV in Ofcolaco farm. The high performance of RapidEye sensor can be credited to the sensors improved spatial resolution and its additional strategically located red edge spectral band. Several vegetation indices were found to be important in detecting the virus. The most important vegetation indices were the Soil Adjusted Vegetation Index (SAVI), Enhanced Vegetation Index (EVI), Red Index (RI) and Normalized Vegetation Index (NDVI). Rapid Eye's ability to frequently monitor large areas with MSV during the growing season, at an affordable cost, presents a great advantage of remote sensing in disease detection and quantification. Since the study utilized single date remotely sensed data, the current study recommends future studies to assess the importance of multi-temporal remote sensing in detecting and monitoring the spread of maize streak virus. However due to the cost involved in acquiring commercial multi-temporal images, like RapidEye, a freely available Landsat-8 OLI data can offer a new data source that is useful for maize diseases estimation, especially in the developing countries where the use of high commercial remote sensing data for precision agriculture remains challenging due to the limited resources.

5.10 References

- Aphane, M. M. 2011. Small-scale mango farmers, transaction costs and changing agro-food markets: evidence from Vhembe and Mopani districts, Limpopo Province.
- Ashourloo, D., Mobasheri, M. R. & Huete, A. 2014. Developing two spectral disease indices for detection of wheat leaf rust (*Puccinia triticina*). *Remote Sensing*, 6(6), 4723-4740.
- Asmala, A. 2012. Analysis of maximum likelihood classification on multispectral data. *Applied Mathematical Sciences*, 6(129-132), 6425-6436.
- Barton, J., Klemd, R. & Zeh, A. 2006. The Limpopo belt: A result of Archean to Proterozoic, Turkic-type orogenesis? *Geological Society of America Special Papers*, 405, 315-332.
- Bastiaanssen, W., Noordman, E., Pelgrum, H., Davids, G., Thoreson, B. & Allen, R. 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of irrigation and drainage engineering*, 131(1), 85-93.
- Belgiu, M. & Drăguț, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- Beukes, D., Bennie, A., Hensley, M., van Duivenbooden, N., Pala, M. & Studer, C. 1999. Optimization of soil water use in the dry crop production areas of South Africa. *Proc. of the 1998 (Niger)(April 26-30 April) and*, 9-13.
- Breiman, L. 1999. Random forests. *UC Berkeley TR567*.
- 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.
- Chan, J. C.-W. & Paelinckx, D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6), 2999-3011.
- Chemura, A. & Mutanga, O. 2016. Developing detailed age-specific thematic maps for coffee (*Coffea arabica* L.) in heterogeneous agricultural landscapes using

- random forests applied on Landsat 8 multispectral sensor. *Geocarto International*, 1-18.
- Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B. & Hofton, M. 2003. Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. *Global ecology and biogeography*, 12(2), 147-159.
- Dube, T. & Mutanga, O. 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 36-46.
- Eitel, J. U., Vierling, L. A., Litvak, M. E., Long, D. S., Schulthess, U., Ager, A. A., Krofcheck, D. J. & Stoscheck, L. 2011. Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, 115(12), 3640-3646.
- Gislason, P. O., Benediktsson, J. A. & Sveinsson, J. R. Year: Published. Random forest classification of multisource remote sensing and geographic data. *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International. IEEE*, 1049-1052.
- Gitelson, A. & Merzlyak, M. N. 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143(3), 286-292.
- Gitelson, A. A., Zur, Y., Chivkunova, O. B. & Merzlyak, M. N. 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and photobiology*, 75(3), 272-281.
- Huang, H., Roy, D. P., Boschetti, L., Zhang, H. K., Yan, L., Kumar, S. S., Gomez-Dans, J. & Li, J. 2016. Separability analysis of Sentinel-2A multi-spectral instrument (MSI) data for burned area discrimination. *Remote Sensing*, 8(10), 873.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X. & Ferreira, L. G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, 83(1), 195-213.
- Huete, A., Justice, C. & Van Leeuwen, W. 1999. MODIS vegetation index (MOD13). *Algorithm theoretical basis document*, 3, 213.

- Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309.
- Joshi, P. C. 2011. Performance evaluation of vegetation indices using remotely sensed data. *International Journal of Geomatics and Geosciences*, 2(1), 231.
- Kranz, J. & Rotem, J. 2012. *Experimental techniques in plant disease epidemiology*, Springer Science & Business Media.
- Lal, R. 2001. Soil degradation by erosion. *Land degradation & development*, 12(6), 519-539.
- M'marete, C. 2003. Climate and water resources in the Limpopo Province. *Agriculture as the cornerstone of the economy in the Limpopo province. A study commissioned by the Economic Cluster of the Limpopo Provincial Government under the leadership of the Department of Agriculture*, 1-49.
- Martin, D., Shepherd, D. & Rybicki, E. 2009. Maize streak virus. *Desk Encyclopedia of Plant and Fungal Virology*, 209.
- Martin, D. P. & Shepherd, D. N. 2009. The epidemiology, economic impact and control of maize streak disease. *Food Security*, 1(3), 305-315.
- Marx, A. 2010. Detection and classification of bark beetle infestation in pure norway spruce stands with multi-temporal RapidEye imagery and data mining techniques. *Photogrammetrie-Fernerkundung-Geoinformation*, 2010(4), 243-252.
- Marx, A., Sagischewski, H., Sossna, I. & Chmara, S. 2010. Detecting bark beetle infestation in spruce stands using multi-temporal RapidEye Satellite Data. *Forst und Holz*, 65(6), 36-40.
- Matongera, T. N., Mutanga, O., Dube, T. & Lottering, R. T. 2016. Detection and mapping of bracken fern weeds using multispectral remotely sensed data: A review of progress and challenges. *Geocarto International*, (just-accepted), 1-31.
- Matongera, T. N., Mutanga, O., Dube, T. & Sibanda, M. 2017. Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor. *International Journal of Applied Earth Observation and Geoinformation*, 57, 93-103.
- Maxwell, A. E., Warner, T. A. & Strager, M. P. 2016. Predicting Palustrine Wetland Probability Using Random Forest Machine Learning and Digital Elevation

- Data-Derived Terrain Variables. *Photogrammetric Engineering & Remote Sensing*, 82(6), 437-447.
- Mzezewa, J., Misi, T. & Van Rensburg, L. D. 2010. Characterisation of rainfall at a semi-arid ecotope in the Limpopo Province (South Africa) and its implications for sustainable crop production. *Water SA*, 36(1), 19-26.
- Nault, L. R. & Ammar, E.-D. 1989. Leafhopper and planthopper transmission of plant viruses. *Annual review of entomology*, 34(1), 503-529.
- Odindi, J., Adam, E., Ngubane, Z., Mutanga, O. & Slotow, R. 2014. Comparison between WorldView-2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm. *Journal of Applied Remote Sensing*, 8(1), 083527-083527.
- Ortiz, S. M., Breidenbach, J. & Kändler, G. 2013. Early detection of bark beetle green attack using TerraSAR-X and RapidEye data. *Remote Sensing*, 5(4), 1912-1931.
- Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-222.
- Peerbhay, K., Mutanga, O., Lottering, R. & Ismail, R. 2016. Mapping *Solanum mauritianum* plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote Sensing of Environment*, 182, 39-48.
- Ramoelo, A., Dzikiti, S., Van Deventer, H., Maherry, A., Cho, M. A. & Gush, M. 2015. Potential to monitor plant stress using remote sensing tools. *Journal of Arid Environments*, 113, 134-144.
- RapidEye, A. 2011a. RapidEye Standard Image Product Specifications. *RapidEye AG: Brandenburg an der Havel, Germany*, 1-54.
- RapidEye, A. 2011b. Satellite imagery product specifications. *Satellite imagery product specifications: Version*.
- Redinbaugh, M., Seifers, D., Meulia, T., Abt, J., Anderson, R., Styer, W., Ackerman, J., Salomon, R., Houghton, W. & Creamer, R. 2002. Maize fine streak virus, a new leafhopper-transmitted rhabdovirus. *Phytopathology*, 92(11), 1167-1174.
- Richter, R. & Schläpfer, D. 2011. Atmospheric/topographic correction for airborne imagery. *ATCOR-4 user guide*, 565-02.
- Rodríguez-Pérez, J. R., Riaño, D., Carlisle, E., Ustin, S. & Smart, D. R. 2007. Evaluation of hyperspectral reflectance indexes to detect grapevine water

status in vineyards. *American Journal of Enology and Viticulture*, 58(3), 302-317.

Rouse Jr, J. W., Haas, R., Schell, J. & Deering, D. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351, 309.

Shoko, C., Clark, D., Mengistu, M., Bulcock, H. & Dube, T. 2015. Estimating spatial variations of total evaporation using multispectral sensors within the uMngeni catchment, South Africa. *Geocarto International*, 1-22.

Singh, K., Forbes, A. & Akombelwa, M. 2014. The Evaluation of High Resolution Aerial Imagery for Monitoring of Bracken Fern. *South African Journal of Geomatics*, 2(4), 296-208.

Chapter 6

MAPPING MAIZE STREAK VIRUS DISEASE USING LANDSAT 8 DATA

This chapter is based on:

Inos Dhau, Elhadi Adam, Onesimo Mutanga & Kingsley K. Ayisi (Under Review)
Investigating the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations. *African Journal of Ecology*,

Abstract

Crop diseases monitoring is critical in understanding the effects of diseases on crop production and associated implications on food security. The aim of this study was to assess the utility of the medium-resolution multispectral Landsat-8 OLI data set, in detecting and mapping maize streak virus disease at Ofcolaco farms in Tzaneen, South Africa. More specifically, the study sought to spectrally discriminate and map maize infected with maize streak virus from other land-cover classes using Landsat 8 data. To achieve this objective two analysis approaches were used: spectral analysis (Test I: spectral bands; Test II: spectral bands + spectral vegetation indices) using random forest. The use of Landsat 8 spectral bands in detection and mapping of maize streak virus disease yielded an overall accuracy of 50.32% whereas the use of vegetation indices improved the accuracies by 1.29%. Blue, short wavelength infrared 1 and short wavelength infrared 2, NDVI, SAVI, MSAVI, CI and EVI were selected as the most important bands for classifying maize streak virus disease. The results of the study show that, the medium-resolution multispectral Landsat 8-OLI data set can be used to detect and map maize infected by maize streak virus.

Keywords: maize disease, mapping, food security, satellite applications, costs

6.0 Introduction

Food insecurity remains a major global concern with approximately one billion people currently experiencing starvation, malnutrition and severe food shortages. These impacts are largely felt in sub-Saharan Africa, where estimates indicate that about 239 million people suffer from hunger, with an anticipated increase in the near future (Rademacher, 2012). National statistics associated this problem with a number of factors, which include poor crop production means, poor governance, climate change and weather variations and more importantly the spread of diseases. Pests and diseases that are affecting crops are becoming a serious issue in the context of climate change, posing a serious challenge to our food production.

Maize which is a staple food crop in Southern Africa and other parts of the world, particularly in Central America and one of the top three cereal crops grown in the world, along with rice and wheat remains one of the most threatened cereal crop (Redinbaugh and Zambrano, 2014). A large number of people all over the world are dependent on maize as their source of labour, income and food (Redinbaugh and Zambrano, 2014). Maize is produced throughout South Africa under diverse environments of varying climatic conditions. According to the South African Department of Agriculture, efficient and sustainable maize production depends mostly on the precise application of production inputs, such as fertilisation, disease control, marketing and financial resources that enhances sustainable agriculture and environment. However, although not well documented maize yield in the region can be affected by various biotic and abiotic factors that most certainly lead to a decline in the overall maize production (Lobell et al., 2009).

Crop production depends on both of these factors for instance, weeds are established in an environment to its way to grow and reproduce, thus leading to frequent competition for resources needed for the crops to grow and produce the goods needed in the society. Climate change is perceived to affect the survival, rate of reproduction, direction, and dispersal of various vectors and pathogens that are responsible for the transmission of different crop diseases (Weintraub and Beanland, 2006). For example, normal drought conditions or irregular rains are often associated with outbreaks of the disease (Rossel and Thottappilly, 1985).

Consequently, yield losses vary with the time of infection and varietal resistance (Bosque-Perez et al., 1998).

Maize streak virus disease (MSV) known for posing a great threat to maize crops, is endemic in sub-Saharan Africa, West Africa, East Africa; and Indian Ocean islands (Bosque-Perez, 2000; Opong et al. 2015; Redinbaugh and Zambrano, 2014). MSV was first described as 'mealie variegation' (Fuller 1901) and after sometime the virus was renamed 'maize streak' (Storey 1925). Maize streak virus disease serves as one of the challenges facing farmlands of sub-Saharan African countries (Opong et al. 2015; Sibley et al 2014). The maize streak virus can lead to the formation and establishment of maize streak disease in maize crops, which can cause major reductions and severe chlorosis on newly emerged leaves (Bosque-Perez 2000).

The maize streak virus can spread and be transmitted by the virus-infected leafhoppers, such as *Cicadulina mbila*, *Cicadulina arachidis*, *Cicadulina ghaurii* and *Cicadulina latens Fennah*, (Bosque-Perez 2000; Van antwerpen et al. 2011; Pioneer 2010). This virus originated from the grasses and it is the most damaging pathogen of maize that leads to loss of income and an increase in food insecurity. Beside the fact that maize crops of sub-Saharan African farmers are being affected, there is an increase in remote sensing technologies and methods for elevating the farmer's abilities of detecting crop diseases as early as possible so as to hinder the possible loss of yields and further damages of maize crops and help in developing informed pesticides spraying protocols and frameworks.

Mapping the spatial distribution of diseases over a large area can therefore assist in understanding the disease infection status, which is critical for loss assessment. Literature has shown that the use of traditional, manual assessment and field survey are still a major way to collect disease distribution information, which is not only labour intensive and time consuming, but also impossible to monitor disease occurrence and severity in a spatially continuous manner over a large area. Maize streak virus disease is one of the most serious viral disease of maize in Africa (Shepherd et al., 2010) MSV is naturally erratic, varying from insignificant in some years to epidemic proportions in others (Kim et al., 1989).

The focus of this study is to detect and map maize crops that have been infected by maize streak virus at Ofcolaco Farms using Landsat 8 data. Landsat 8 is made up of a land imager and the thermal infrared sensor. These two instruments of Landsat 8 can offer signal-to-noise (SNR) radiometric performance over 12-bit dynamic range. The improved signal to noise performance of Landsat 8 images gives a way for a better differentiation of land cover features and the resulting conditions for that particular ground features which are captured after every 16 days (USGS, 2017).

6.1 Materials and Methods

6.1.1 Description of the study area

The research was conducted at farms around Ofcolaco in Tzaneen, in the Limpopo Province, of South Africa, (Figure 6.1). The study area is located between S 24° 4' 59.99" and E 30° 22' 0.01". High humidity and heavy rainfall characterises the climate of the study area, and MSV disease frequently occurs in most of the years. The Limpopo Province produces a variety of agricultural produce ranging from tropical fruits such as bananas, mangoes to cereals such as maize, wheat and vegetables such as tomatoes, onion and potatoes (Aphane, 2011). There are two distinct types of agricultural production systems in Limpopo, these include the large scale commercial farming systems and the small holder subsistence farming system.

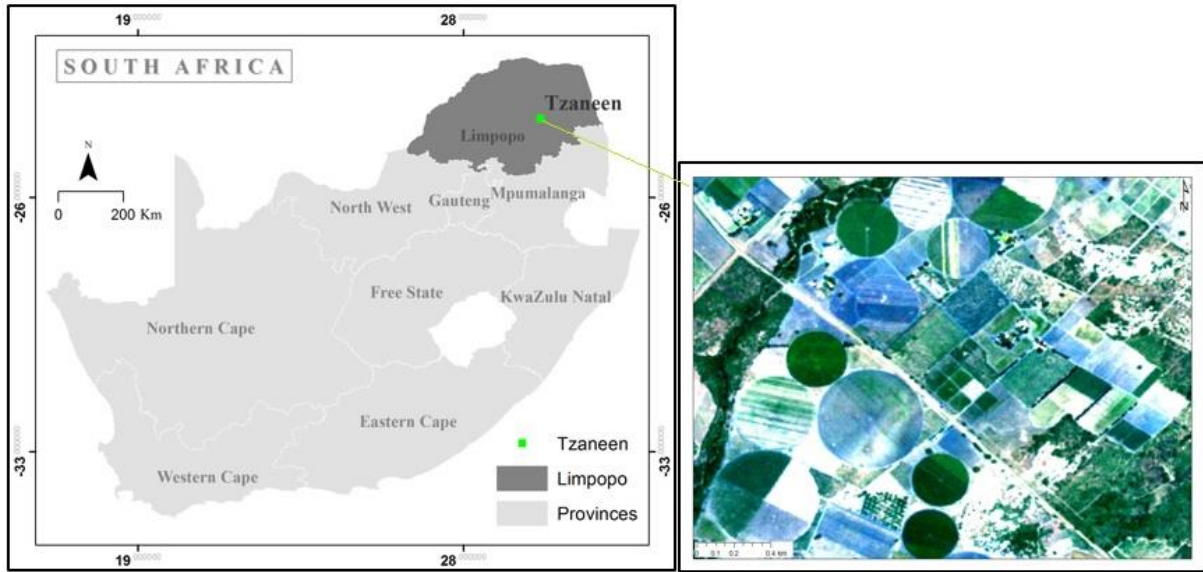


Figure 6.1 Location of the study area

6.1.2 Field data collection and pre-processing

6.1.2.1 Field Measurements and image pre-processing

GPS data of healthy and MSV infected maize were collected using a hand held Leica GS20 Global positioning system (GPS), with a sub meter accuracy. Field data collection was conducted in January 2016 during the time the image was acquired. Purposive sampling strategy was adopted in choosing sampling sites within the farm. Sampling sites were chosen, one site with healthy maize field and the other with MVS infected maize. The other land cover classes which were identified and their locations recorded include bushveld, grassland, fruit trees and base soil. Field data was randomly split into 70% training and 30% validation data sets.

Landsat 8 image covering the area of interest was acquired during the time that coincided with the field measurement date. The Landsat image was acquired during sunny and clear sky day conditions, with cloud cover less than 10%, a sun azimuth angle of 39.78 and a sun elevation angle of 33.14. The satellite image was obtained from the USGS Earth Resources Observation and Science (EROS) Center archive (<http://earthexplorer.usgs.gov/>) with eight spectral bands (Table 1). The Landsat-8 OLI satellite image was obtained in digital number (DN), and then it was converted to

reflectance values. Conversion from DN to reflectance of the Landsat-8 OLI, was applied in ENVI environment, following the approach described on the USGS website (<http://landsat.usgs.gov>). The image was geometrically and atmospherically corrected.

Table 6.1: Landsat 8 OLI spectral bands (<http://landsat.usgs.gov>)

Spectral Band	Wavelength	Resolution
Band 1 - Coastal / Aerosol	0.433 – 0.453 μm	30 m
Band 2 - Blue	0.450 – 0.515 μm	30 m
Band 3 - Green	0.525 – 0.600 μm	30 m
Band 4 - Red	0.630 – 0.680 μm	30 m
Band 5 - NIR	0.845 – 0.885 μm	30 m
Band 6 - SWIR	1.560 – 1.660 μm	30 m
Band 7 - SWIR	2.100 – 2.300 μm	30 m
Band 8 - Panchromatic	0.500 – 0.680 μm	15 m

6.2 Image classification and accuracy assessment

Landsat 8 imagery was classified using Random forest (RF) classification ensemble into different land cover types. RF was used in many studies for classifying remotely sensed data (Odindi et al., 2014, Chan and Paelinckx, 2008, Gislason et al., 2004). Random forest is an ensemble learning technique developed for the classification and regression trees by combining large set of decision trees (Breiman 2001). Each tree contributes a single vote in deciding the final class of an object (Breiman 2001). The algorithm constructs numerous binary classification trees (*n*tree) using bootstrap samples with replacements drawn from the original data set. Random forest library that was developed in R statistical software was used to execute the RF algorithm and the classification maps prepared in ENVI 4.8. In order to compare the true class with the class assigned by the classifier a confusion matrix was developed. The overall, user's and producer's accuracies were also calculated. In order to test whether vegetation indices can improve the classification accuracy, at least nine indices (Table 6.2) were chosen as variables in the classification process. Joshi

(2011) assessed the utility of vegetation indices for vegetation mapping. An overlay was conducted between the point map of GPS locations of land cover classes, and the remotely sensed data, in ENVI 4.8 to extract spectral data. The extracted spectral data was then exported from ENVI in a tabular format in notepad.

Table 6.2: Vegetation indices used in the classification process

Vegetation index	Equation	Reference
NDVI	$NDVI = \frac{NIR - Red}{NIR + Red}$	Qinghan, Herman and Zhongxin 2008
EVI	$EVI = G * ((NIR - R) / (NIR + C1 * R - C2 * B + L))$	Lawley et al, 2016 and jiang et al 2008
SAVI	$SAVI = ((Band\ 5 - Band\ 4) / (Band\ 5 + Band\ 4 + 0.5)) * (1.5)$	Jiang et al 2008
MSAVI	$MSAVI = (2 * Band\ 5 + 1 - \sqrt{(2 * Band\ 5 + 1)^2 - 8 * (Band\ 5 - Band\ 4)}) / 2.$	Hunt Jr et al. 2013.
NDMI	$NDMI = (Band\ 5 - Band\ 6) / (Band\ 5 + Band\ 6).$	Xu 2006
CI	Coloration Index, $C (R-G)/(R+G)$	
GLI	$GLI = \frac{2 * GREEN - RED - BLUE}{2 * GREEN + RED + BLUE}$	Nguy-Robertson et al. 2012
NGRDI	$NGRDI = \frac{RED - GREEN}{RED + GREEN}$	Yuan et al. 2017 In addition, Hunt Jr et al. 2013.
NBR2	$NBR2 = (Band\ 6 - Band\ 7) / (Band\ 6 + Band\ 7)$	

6.3 Results

6.3.1 Spectral response of maize infected with MSV and other land cover types

The results show that infected maize can be spectrally discriminated from other landcover classes using the Blue, SWIR1 and SWIR2 regions of the electromagnetic spectrum as shown on Figure 6.2.

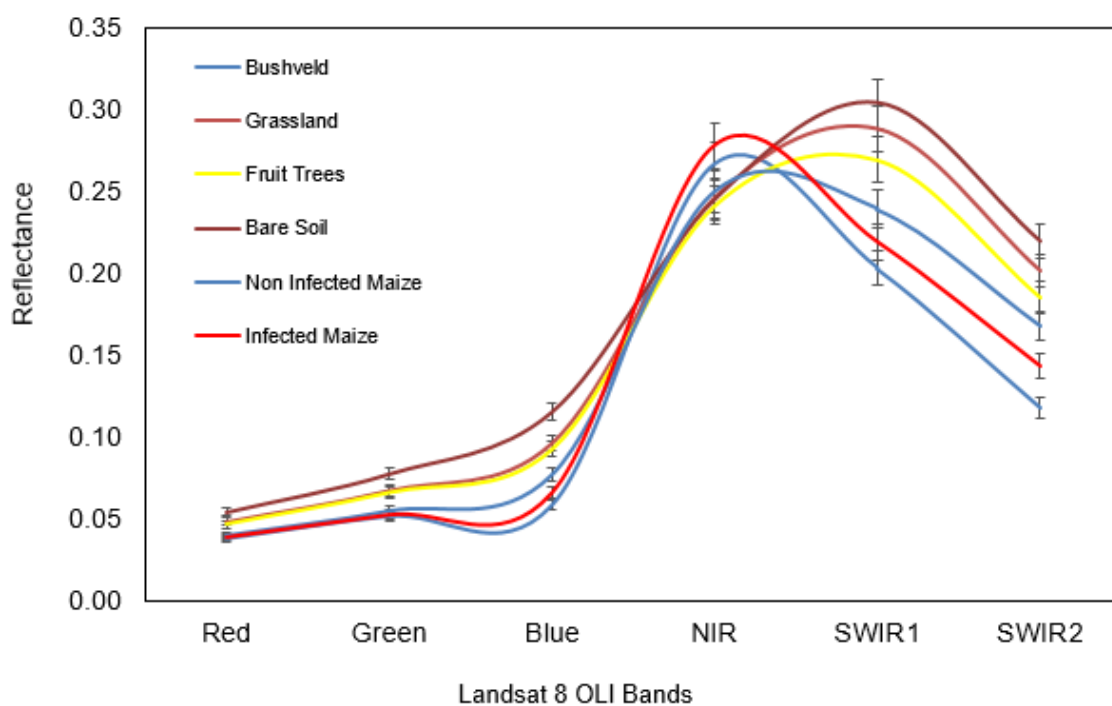


Figure 6.2 Spectral reflectance of maize infected with MSV and other landcover types

6.4 Image classification

6.4.1 Classification results using Landsat 8 spectral bands

The classification results attained using Landsat 8 spectral bands data set are shown in figure 6.3. As illustrated in figure 6.3, the results revealed that the use of Landsat 8 spectral bands in detection and mapping of maize streak virus yielded classification

results with an overall accuracy of 50.32%. In this regard, Landsat 8 derived spectral bands demonstrated potential of detecting and mapping of maize streak virus.

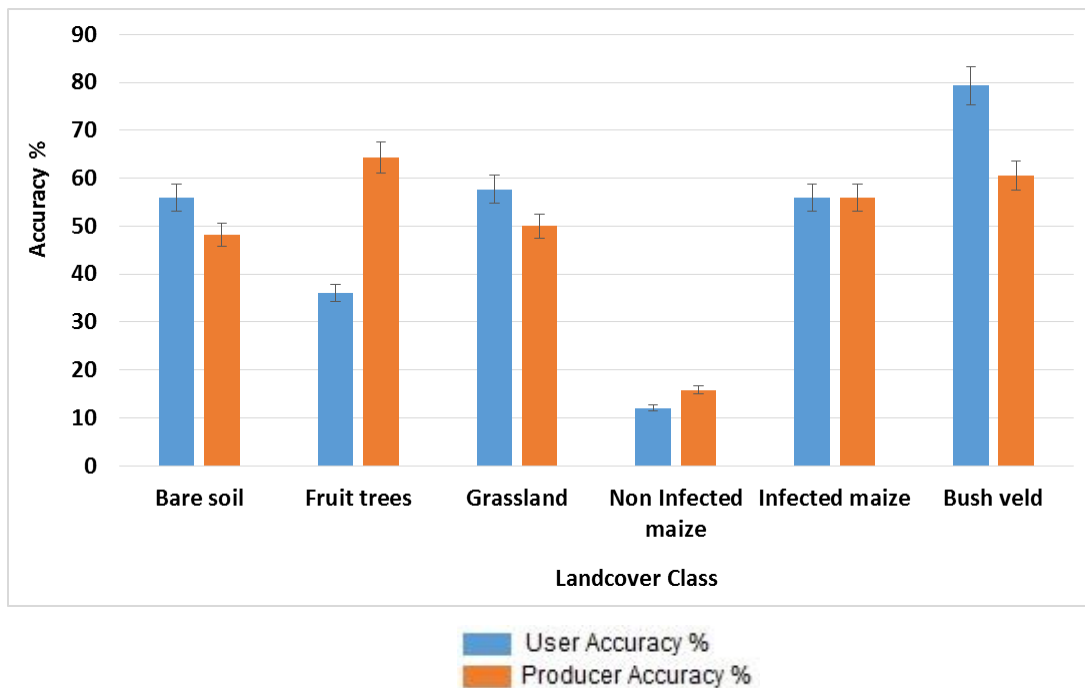


Figure 6.3 Classification accuracy for the infected and health maize and other classes using spectral bands values

6.4.2 Classification results using computed vegetation indices

Computed vegetation indices classification results are presented in Figure 6.4. The results showed that the inclusion of vegetation indices in detection and mapping of maize streak virus produced slightly improved classification results with an overall accuracy of 51.61%. Compared to the use of raw spectral bands, the inclusion of vegetation indices improved overall accuracy by 1.29%. Therefore, computed vegetation indices improved classification accuracy results in detection and mapping of maize streak virus.

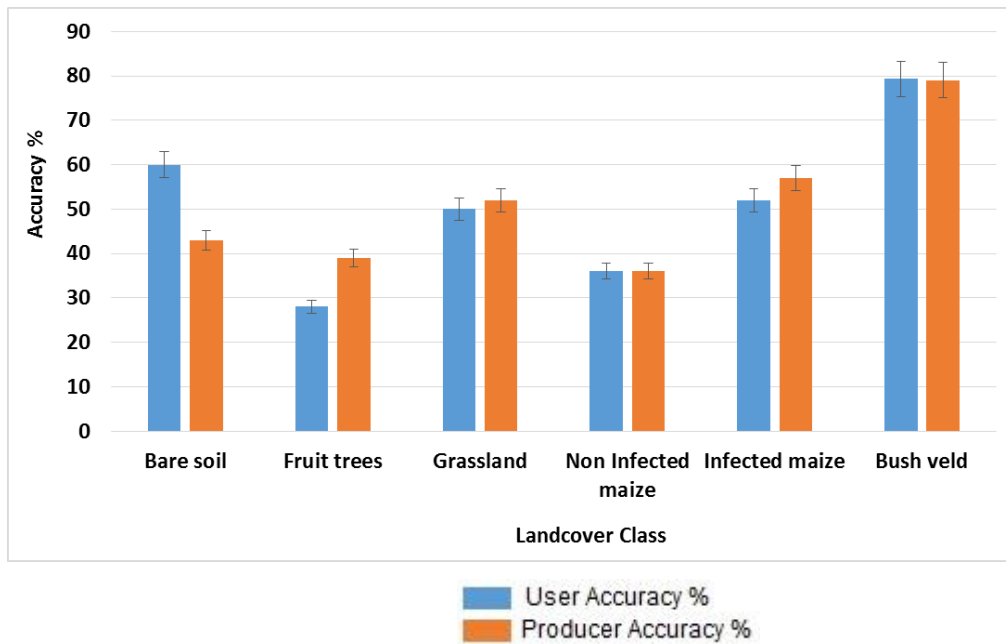


Figure 6.4 Classification accuracy for the infected and health maize and other classes using spectral bands values and vegetation indices

6.5 Spatial distribution of MSV infected maize and other land cover classes

The spatial distribution of maize infected by MSV and other land cover classes at Ofcolaco Farms is shown in Figure 6.5 below. The land cover map shows that the infected maize is mainly distributed in the central and southeast part of Ofcolaco farms. The map also shows that the western part of Ofcolaco area is not much affected as there are few maize farms. According to the spatial distribution map, MSV hardly co-exist with healthy maize. The MSV disease spreads and infects nearby healthy maize and becomes concentrated in infected fields.

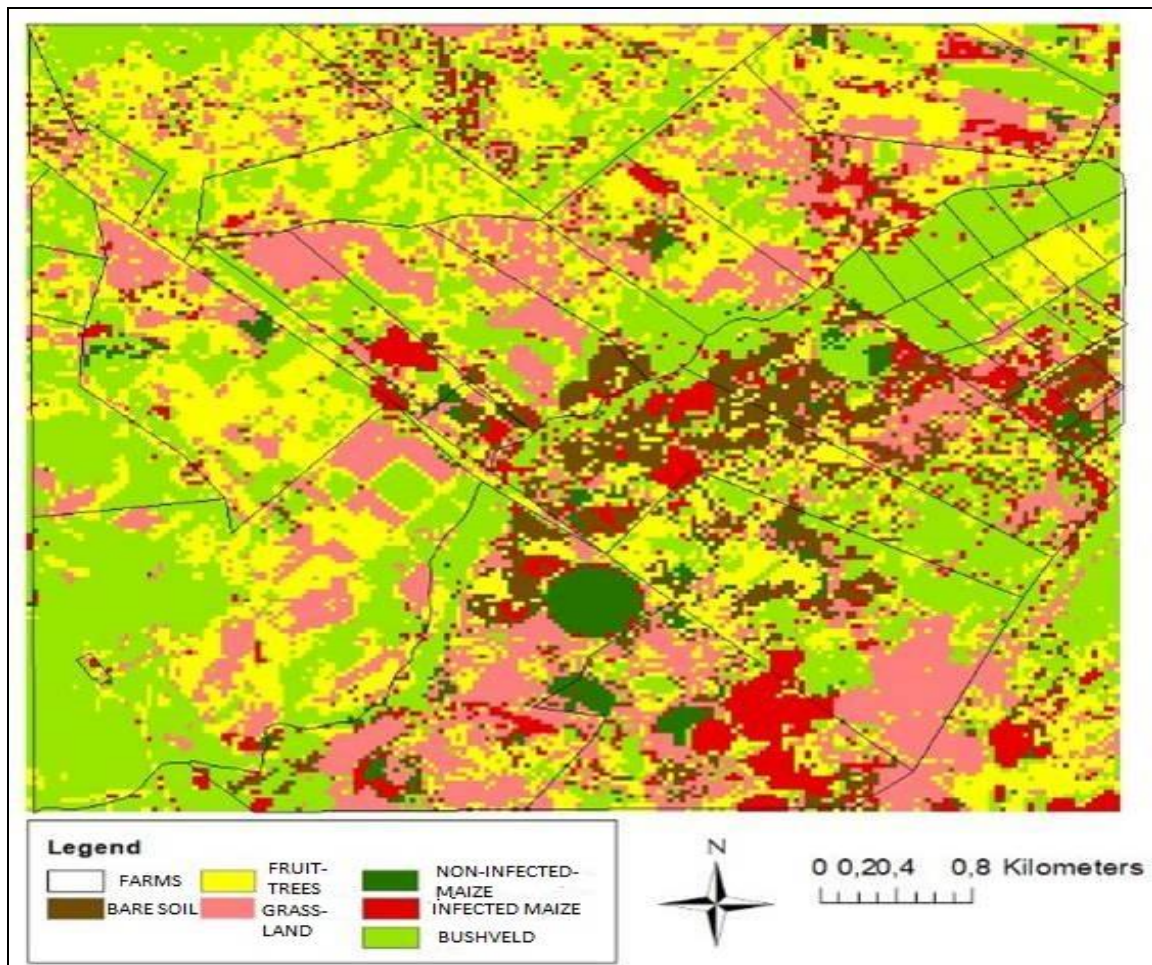


Figure 6.5 Land cover classification map of infected maize and other classes

The results show that infected maize crops cover 7.52 percent of land, while healthy maize covers 2.77 percent. The bare soil covers 7.14 percent, fruit trees covers 28.80 percent, grassland 19.84 percent and bushveld covers 33.93 percent. Figure 6.6 shows that the area covered by infected maize is greater than the area covered by healthy maize which shows that the greater part of the maize crops is affected by MSV.

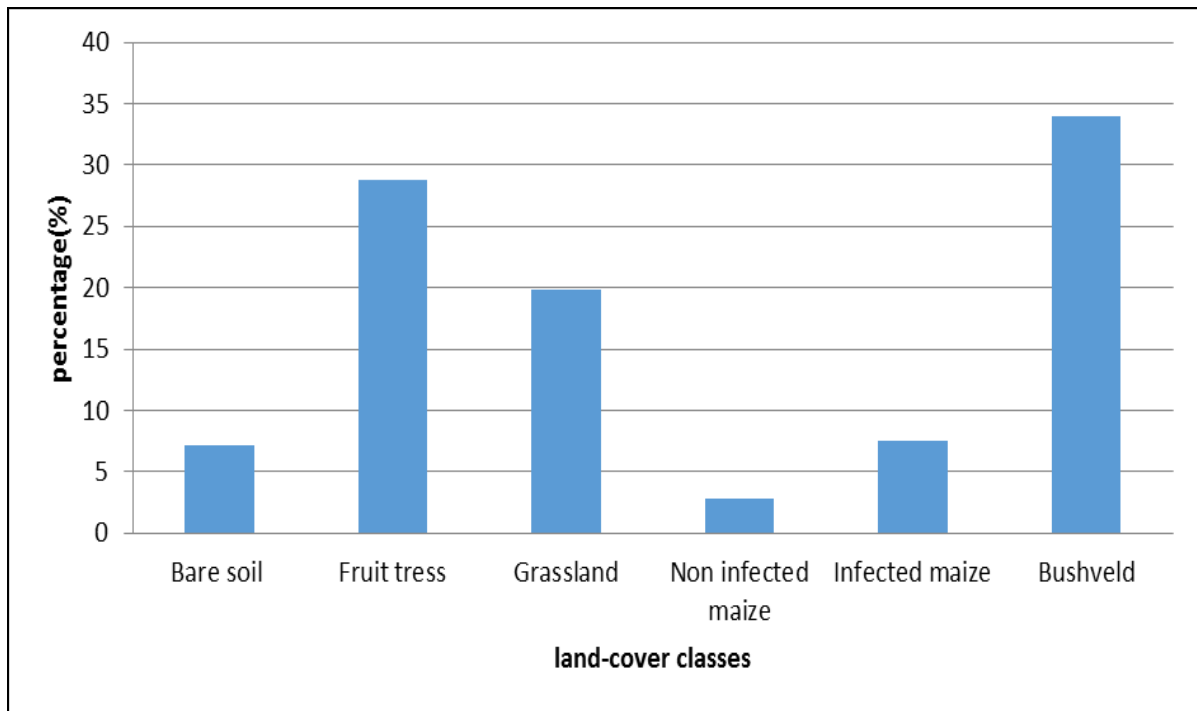


Figure 6.6 Estimated percentage of each land-cover class in Ofcolaco farm, 2016.

6.6 Discussion

This study aimed at examining the potential of a freely available Landsat-8 OLI sensor in mapping and detecting MSV disease in Ofcolaco farms in Tzaneen, South Africa. The estimation of MSV disease provides an important input data and information required for accurate modelling and monitoring of MSV infestation in maize farms. This study therefore focused on investigating whether the Landsat-8 OLI has the capability of detecting and mapping MSV disease at regional level. Landsat 8 multispectral data managed to detect and map maize streak virus infestation with an acceptable level of accuracy which was then improved by combining Landsat data with derived vegetation indices. The performance of Landsat 8 sensor in detecting and mapping maize streak virus infestation in Ofcolaco farm can be attributed to the sensors improved spectral and radiometric resolution.

Several studies have reported similar findings using remotely sensed data when detecting and mapping crop diseases (Beck et al., 2000, Blakeman, 1990, Lamb and Brown, 2001, Liu et al., 2006a, Mirik et al., 2011, Qin and Zhang, 2005, Vogelmann and Rock, 1989, Vogelmann et al., 2012, Zhang et al., 2012, Liu et al., 2006b). For

instance Liu et al., (2006b) used three Landsat Thematic Mapper (TM) satellite images and three Envisat Advanced Synthetic Aperture Radar (ASAR) satellite images acquired from reviving stage to milking stage of winter wheat. This data was successfully used to monitor crop condition and forecast grain yield and protein content. Results from this study indicated that both multi-temporal Envisat ASAR and Landsat TM imagery could provide accurate information about crop conditions.

The classification process using Landsat 8 spectral bands yielded an overall accuracy of 50.32%, whilst the computed vegetation indices improved the classification accuracy by 1.29%. The inclusion of vegetation indices will improve the accuracy of classification. Vegetation indices are well known proxies of vegetation greenness (Broge and Leblanc, 2001). The spatial distribution of MSV infected maize and other land cover classes shows that maize streak virus was distributed in the central part of the study area, and the South-eastern parts of the map. Maize streak virus infection can cause the infection of non-infected maize crops. The results from this study, encourages the application of other satellites with high spatial and spectral resolution.

6.7 Implications of the study on food security in Sub Saharan Africa

Food security, which involves the balance between food production and demand in the face of climate change, diseases and pests, has of late become an imperative issue, especially in the developing world. Food security is currently under threat from the spread of diseases and pests which have largely compromised yield in terms of quality and quantity. Therefore, routine, timely, and spatially explicit information is critical for food security, particularly in sub-Saharan Africa where hunger and poverty have reached unprecedented levels. Earth observation technologies have of late been found to provide an effective way to acquire timely crop information over relatively large geographical areas. However, its adoption in most developing countries for decision making purposes has been hampered by the cost associated with high resolution data. The advent of free archival data by NASA, therefore provides new opportunities for crop and disease monitoring on large geographic scales. Information derived using this data will help farmers from poor countries to

prioritize the spraying and other-related disease control mechanisms and minimize costs by avoiding the spraying of the entire farm. Therefore, the findings of this study provide the most necessary baseline information on the importance of earth observation in alleviating food security against the scourge of climate change and its effects on crop production.

6.8 Conclusion

The objective of this study was to detect and map maize streak virus disease using Landsat 8 multispectral imagery. The findings from this study established that Landsat 8 multispectral imagery can be used in detection and mapping of maize streak virus disease at Ofcolaco Farms with an acceptable level of accuracy. The performance of Landsat 8 sensor can be attributed to the sensor's improved radiometric resolution. The ability to monitor large areas with MSV disease at large scale during the maize growing season, using freely available Landsat data, is however an added advantage of remote sensing in disease detection and mapping. However, for the approaches presented in these studies to be operational, various available hyperspectral, newly-launched multispectral sensors, and the influence of derived vegetation indices together with environmental variables, such as elevation, slope, rainfall and temperature, should be investigated and tested in different maize diseases under different climatic conditions.

Acknowledgements

The author would like to thank the Ofcolaco communities and farmers for their support during field data collection as well as for allowing us to use their farms as experimental sites. The author would also like to thank the South African weather services for the provision of climatic data. NASA is also appraised for providing Landsat and DEM data. A word of appreciation goes to the VLIR-IUC Programme for providing logistical support.

6.9 References

- APHANE, M. M. 2011. Small-scale mango farmers, transaction costs and changing agro-food markets: evidence from Vhembe and Mopani districts, Limpopo Province.
- BECK, L. R., LOBITZ, B. M. & WOOD, B. L. 2000. Remote sensing and human health: new sensors and new opportunities. *Emerging infectious diseases*, 6, 217.
- BLAKEMAN, R. 1990. The identification of crop disease and stress by aerial photography. *Applications of remote sensing in agriculture*, 229-254.
- BOSQUE-PEREZ, N., OLOJEDE, S. & BUDDENHAGEN, I. 1998. Effect of maize streak virus disease on the growth and yield of maize as influenced by varietal resistance levels and plant stage at time of challenge. *Euphytica*, 101, 307-317.
- 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, 156-172.
- CHAN, J. C.-W. & PAELINCKX, D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999-3011.
- GISLASON, P. O., BENEDIKTSSON, J. A. & SVEINSSON, J. R. Random forest classification of multisource remote sensing and geographic data. Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International, 2004. IEEE, 1049-1052.
- KIM, S.-K., EFRON, Y., FAJEMISIN, J. & BUDDENHAGEN, I. 1989. Mode of gene action for resistance in maize to maize streak virus. *Crop science*, 29, 890-894.
- LAMB, D. & BROWN, R. B. 2001. Pa—precision agriculture: Remote-sensing and mapping of weeds in crops. *Journal of Agricultural Engineering Research*, 78, 117-125.

- LIU, D., KELLY, M. & GONG, P. 2006a. A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote sensing of environment*, 101, 167-180.
- LIU, L., WANG, J., BAO, Y., HUANG, W., MA, Z. & ZHAO, C. 2006b. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images. *International Journal of Remote Sensing*, 27, 737-753.
- LOBELL, D. B., CASSMAN, K. G. & FIELD, C. B. 2009. Crop yield gaps: their importance, magnitudes, and causes. *Annual review of environment and resources*, 34.
- MIRIK, M., JONES, D., PRICE, J., WORKNEH, F., ANSLEY, R. & RUSH, C. 2011. Satellite remote sensing of wheat infected by wheat streak mosaic virus. *Plant Disease*, 95, 4-12.
- ODINDI, J., ADAM, E., NGUBANE, Z., MUTANGA, O. & SLOTOW, R. 2014. Comparison between WorldView-2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm. *Journal of Applied Remote Sensing*, 8, 083527-083527.
- QIN, Z. & ZHANG, M. 2005. Detection of rice sheath blight for in-season disease management using multispectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 7, 115-128.
- RADEMACHER, A. 2012. Food Insecurity in Sub-Saharan Africa. *Published: October, 26, 2012*.
- REDINBAUGH, M. G. & ZAMBRANO, J. L. 2014. Control of virus diseases in maize. *Advances in virus research*. Elsevier.
- ROSSEL, H. & THOTTAPPILLY, G. 1985. Virus diseases of important food crops in Tropical Africa. *Virus diseases of important food crops in tropical Africa*.
- SHEPHERD, D. N., MARTIN, D. P., VAN DER WALT, E., DENT, K., VARSANI, A. & RYBICKI, E. P. 2010. Maize streak virus: an old and complex 'emerging' pathogen. *Molecular plant pathology*, 11, 1-12.
- VOGELMANN, J. E. & ROCK, B. N. 1989. Use of Thematic Mapper data for the detection of forest damage caused by the pear thrips. *Remote Sensing of Environment*, 30, 217-225.
- VOGELMANN, J. E., XIAN, G., HOMER, C. & TOLK, B. 2012. Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in

selected forest and rangeland ecosystems. *Remote Sensing of Environment*, 122, 92-105.

WEINTRAUB, P. G. & BEANLAND, L. 2006. Insect vectors of phytoplasmas. *Annu. Rev. Entomol.*, 51, 91-111.

ZHANG, J.-C., PU, R.-L., WANG, J.-H., HUANG, W.-J., YUAN, L. & LUO, J.-H. 2012. Detecting powdery mildew of winter wheat using leaf level hyperspectral measurements. *Computers and Electronics in Agriculture*, 85, 13-23.

Chapter 7

SPATIAL MODELLING OF MAIZE STREAK VIRUS DISEASE USING ENVIRONMENTAL VARIABLES IN SEMI-ARID ENVIRONMENTS

This Chapter is based on:

Inos Dhau, Elhadi Adam, Onesimo Mutanga & Kingsley K. Ayisi (Under Review)
Spatial modelling of maize streak virus disease using environmental variables in semi-arid environments. *African Journal of Ecology*.

Abstract

In this study, the aimed was to examine the influence of climatic, environmental and remotely sensed variables on the spread of MSV disease on maize farms at the Ofcolaco farms in Tzaneen, South Africa. The study integrated environmental and climatic variables together with Landsat 8 derived vegetation indices to predict the probability of MSV occurrence at Ofcolaco maize farms in Limpopo, South Africa. Correlation analysis was used to relate vegetation indices, environmental and climatic variables to incidences of maize streak virus disease. The variables used to predict the distribution of MSV were elevation, rainfall, slope, temperature, and vegetation indices. It was found that MSV disease infestation is more likely to occur on low-lying altitudes and areas with high NDVI located at an altitude ranging between 350 and 450 m.a.s.l. The suitable areas are characterized by temperatures ranging from 24°C to 25°C. The results indicate the potential of integrating Landsat 8 derived vegetation indices, environmental and climatic variables to improve the prediction of areas that are likely to be affected by MSV disease outbreaks in maize fields in semi-arid environments.

Keywords: climatic variables, elevation, food security, remotely sensed data, sub-Saharan Africa

7.0 Introduction

Maize (*Zea Mays L.*) plays a major role as a food security crop in both rural and urban communities. However, food production is facing challenges from crop diseases, such as maize streak virus, amongst others (Mahuku et al., 2015). Maize streak virus infestations can be quantified by visually assessing the disease severity and intensity (Nutter Jr et al., 2010, Fajemisin et al., 1984). However, the accuracy and precision of visual disease assessments performed by different people continue to be a cause for concern (Fajemisin et al., 1984). The integration of remote sensing and geographic information systems offers new opportunities to obtain, process, and analyze geospatially referenced data in a timely and spatial manner (Nelson et al., 1999). Thus, pathogen data and host populations, biotic and abiotic risk factors, can be mapped, overlaid, and displayed at multiple spatial scales to explain associations and cause and effect relationships among data layers (Hess et al., 2002, Iqbal et al., 2010).

There is a need to develop metrics and spatial tools for evaluating and monitoring integrated disease management strategies (Hamerschlag and Kaplan, 2007). Remote sensing, when coupled with environmental variables, has the potential to assess crop health over time and space, with greater accuracy and precision (Donatelli et al., 2017). Therefore, the integration of remote sensing and environmental variables represents a new pattern in that disease management strategies could, in the future, be evaluated and deployed based upon the capability of a disease management program to produce and maintain healthy maize crop (Nutter Jr et al., 2010). Determining the factors that influence the spatial distribution of maize diseases is one of the most complex problems in plant pathology. This is because the spatial distribution is not influenced by a single factor, but by a complex array of interacting factors, both biotic and abiotic (Seagle and McNaughton, 1992).

Maize crops are susceptible to MSV disease at all phenological stages. The rate of MSV disease infection and damage is high, affecting yields and causing crop loss, therefore have significant impacts on food security in sub-Saharan Africa (Monjane et al., 2011). Researchers have shown that MSV disease occurrence tends to be influenced by rainfall, temperature, soil moisture and slope (Dabrowski et

al., 1991). The amount of damage caused by crop disease pathogens is strongly correlated with high precipitation levels (Hasan et al., 2009). In South Africa, maize streak virus is prevalent in parts of Limpopo, KwaZulu-Natal, the Northern Cape, and Mpumalanga, particularly in areas experiencing high rainfall and temperatures. At the landscape scale, topographic factors, such as slope, aspect, and altitude (Rose, 1978), together with remote sensing data are critical in predicting areas that are likely to be affected by maize streak virus. An investigation of the interaction between environmental and climatic factors in influencing the spatial distribution of MSV disease is critical as it aids in climate change modelling. The number of MSV disease vectors available during the rainy season is determined by the size of the cultivated area, with precipitation favouring the development of its hosts. With hot temperatures, there is a higher risk of epidemics of MSV disease, with a reversal of the situation at the end of the dry season (Rose, 1978).

Remote sensing techniques together with the spatial analysis is a good and readily available spatial tool for modelling the spatial distribution of MSV as compared to conventional ground surveys. With the advancement of geostatistical models, such as spatial regression, MSV disease can be linked to its causal factors and spatial patterns established (Rose, 1978). Moreover, such studies help in epidemiological surveillance which is vital in developing any MSV disease control strategies. With the use of spatial analysis and remote sensing, such studies could be geared towards site-specific MSV disease control measures. A MSV disease probability map can be utilized by multiple stakeholders to guide on location-based interventions, with emphasis on high severity areas. Farmers could be advised to plant maize at favourable MSV disease -free model-derived environmental conditions. It is at these locations characterised by high-to-medium MSV disease severities that farmers are also advised to use MSV disease control measures, such as insecticides to divide off MSV disease insect vectors, and early planting at low-to-zero model derived environmental conditions. This study relied on MSV disease presence and absence data as the response variable. Further research is required based on MSV disease presence and absence or any other relevant response variable using spatial and ecological models that account for spatial dependence and autocorrelation in order to generate comparable MSV disease risk maps for South Africa. The aim of this study was thus to identify climatic, environmental and remotely sensed variables that

correlate with MSV disease at a landscape scale. The relationship between MSV disease and landscape variables (slope, altitude, aspect, selected vegetation indices) were examined in this study.

7.1 Materials and Methods

7.1 1 Study Area

The research was conducted at a 20 (ha) Ofcolaco Farm located 60km south of Tzaneen- Limpopo province, South Africa in the Limpopo province, South Africa, (Figure 7.1). Geographically, the study area is located between S 24° 4' 59.99" and E 30° 22' 0.01". The climate of the study area is mainly subtropical, with warm summers and mild winters. The area is characterised by intensive maize production.

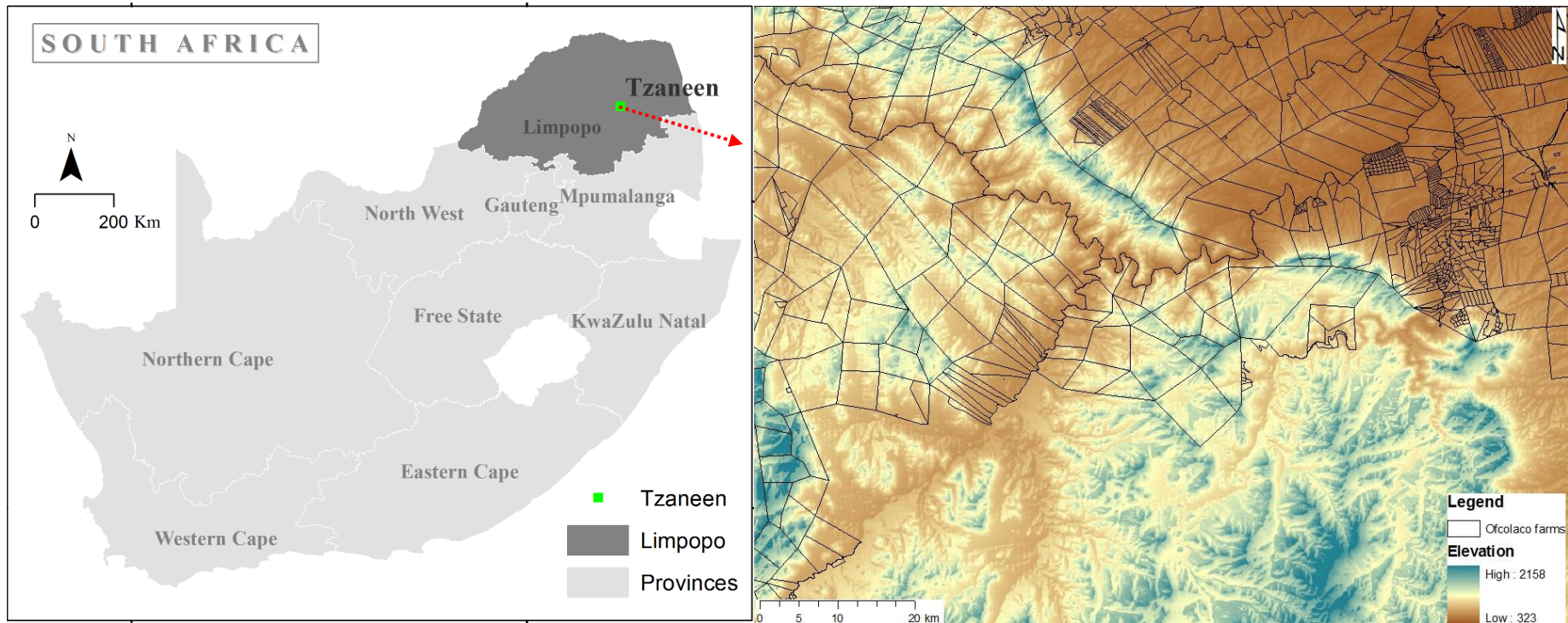


Figure 7.1 Location of the study area

7.1.2 Field data collection

Field data were collected in January 2016. Data collection was conducted to record the presence and absence of MSV disease, using a hand held Leica GS20 Global positioning system (GPS) with a sub-meter accuracy. A purposive sampling strategy was adopted in choosing sampling sites within the farm. Sampling sites were chosen, one site with healthy maize field and the other with MSV infected maize. The other land cover classes that were recorded include bushveld, grassland, fruit trees, and base soil.

The environmental variables that were used in this study are provided in Table 7.1 and Figure 7.2. The Digital Elevation Model (DEM) was used at a spatial resolution of 30 m and its derivatives were derived in ArcGIS 10.4, using spatial analyst extension. For climatic factors, rainfall and temperature were used. Rainfall and temperature data were taken as point values of monthly rainfall and temperature recorded at two stations in the study area. Rainfall and temperature point data was also interpolated to obtain its spatial variability across the study area; this was performed using ordinary Kriging interpolation method in ArcGIS 10.4. Centroids of areas affected by MSV were used to extract values to points of the variables under study.

Table 7.1: Climatic, environmental and remotely sensed variables

Variable	Definition	Source
P/A	Presence and absence of MSV disease.	Field-based measurements
VIs	Vegetation Indices	NASA
Elevation	Height above sea level, in meters	DEM
Slope	Elevation steepness, in degrees from 0 (flat) to 90 (steep)	DEM
Rainfall	Monthly total, in millimeters (mm).	SAWS
Temperature	Average, in degrees Celsius (°C).	SAWS

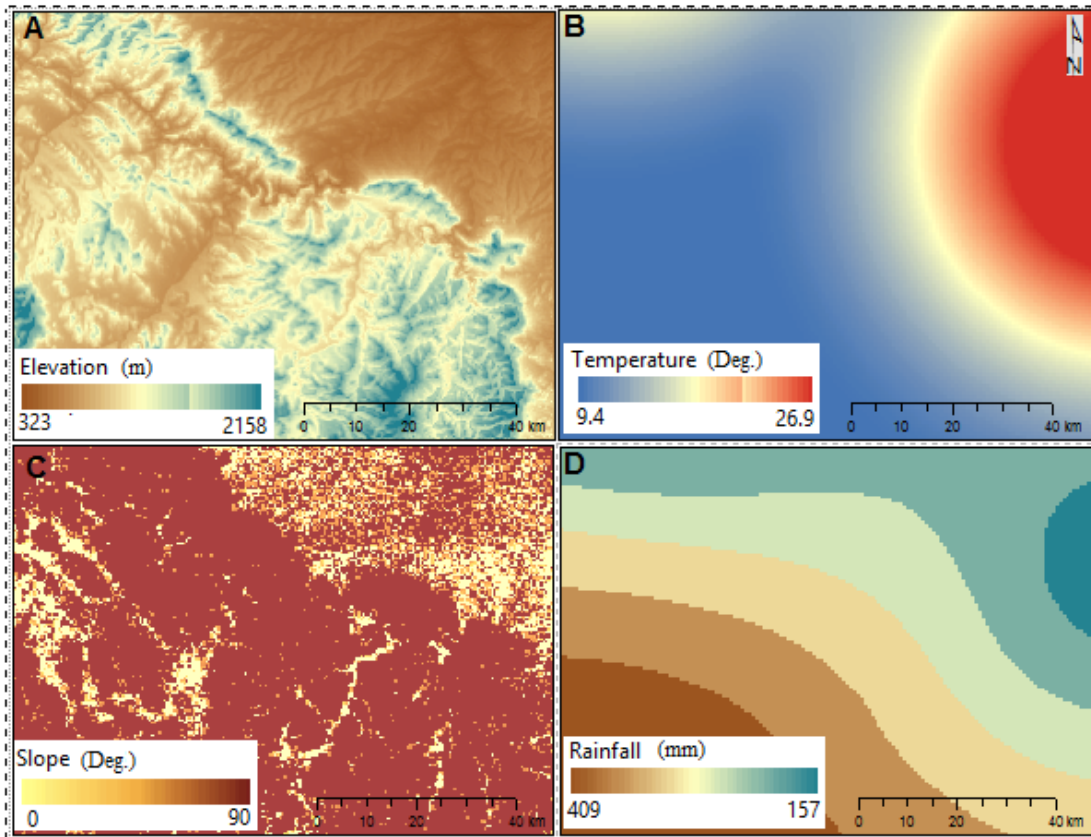


Figure 7.2 Environmental variables that were used in this study

7.2 Modelling

Correlation analysis was done to establish the relationship between environmental variables and the area affected by MSV disease. The collected presence and absence data was used together with elevation, slope, temperature, rainfall, and vegetation indices to predict areas that are likely to be affected by maize streak virus disease. Stepwise logistic regression was used to select significant factors that can explain the presence of MSV disease. The continuous variables were used in ArcGIS 10.4 to predict the probability of occurrence of disease.

7.3 Results

Table 7.2 shows descriptive statistics for the fourteen climatic factors, remotely sensed and topographic derivatives considered in this study. For the remotely sensed derivatives, NDVI exhibited the highest maximum value (0.89), followed by NBR with 0.77 whereas the GDVI and NGRDI had the least values. In terms of the standard deviation, NDR, NDVI, and NDMI had the highest. For instance, for the three variables the following results were observed, where $0.19 > \text{std.dev} > 0.25$. For environmental factors, elevation ranged between 591 and 625 meters above sea level (m.a.s.l), on the other hand, slope ranged between 0 and 90° . Rainfall was very low with a maximum of 257.60mm and a minimum of 94.65mm, whereas temperature was in the magnitude of 24.18 to 25.27°C.

Table 7.2: Descriptive statistics

Variable	Minimum	Maximum	Mean	Std. deviation
NDVI	0.12	0.89	0.47	0.19
EVI	0.03	0.70	0.29	0.15
SAVI	0.04	0.62	0.29	0.13
MSAVI	0.03	0.67	0.27	0.14
NDMI	-0.30	0.53	-0.02	0.20
NBR	-0.21	0.77	0.17	0.25
NBR2	0.05	0.42	0.21	0.08
CI	-0.08	0.62	0.44	0.14
GDVI	-0.11	0.15	-0.01	0.06
NGRDI	-0.28	0.13	-0.14	0.10
Elevation	591	625	607	8.12
Slope	0.00	89.99	88.92	9.68
Rainfall	94.65	257.60	214.02	50.80
Temperature	24.18	25.27	24.76	0.25

7.3.1 Pearson's correlation analysis

Results of Pearson's correlation analysis in Table 7.3 generally show that the independent variables are both positively and negatively correlated with the area affected by maize streak virus. A weak correlation with all environmental variables is observed. Relatively weak correlations were obtained between environmental variables and area affected by MSV. It can be observed that most vegetation indices exhibited a positive correlation with the area affected by maize streak virus disease except for CI with a negative correlation (-0.294). However, it is imperative to note that the positive correlation exhibited between area affected and vegetation indices were generally weak, with the value ± 0.20 . Topographic derivatives (elevation and slope) exhibited a negative correlation with the value ± 0.01 . While climatic factors (rainfall and temperature) had a positive correlation with temperature showing a significant correlation ($\alpha = 0.05$). Correlation analysis was also implemented between the variables and a strong positive correlation was observed among all the vegetation indices and climatic factors.

Table 7.3: Correlation between variables and area affected with MSV disease

Variables	P(area)	NDVI	EVI	SAVI	MSAVI	NDMI	NBR	NBR2	CI	GDVI	NGRDI	Elevation	Slope	Rainfall	Temperature
P(area)	1	0.231*	0.212*	0.200*	0.208*	0.222*	0.213*	0.228*	-0.294*	0.196*	0.244*	-0.016	-0.023	0.105	0.287*
NDVI	0.231*	1	0.977*	0.978*	0.965*	0.951*	0.964*	0.941*	-0.889*	0.932*	0.943*	-0.088	-0.027	0.222*	0.217*
EVI	0.212*	0.977*	1	0.998*	0.998*	0.972*	0.974*	0.948*	-0.881*	0.916*	0.929*	-0.045	-0.024	0.187*	0.207*
SAVI	0.200*	0.978*	0.998*	1	0.996*	0.967*	0.973*	0.943*	-0.855*	0.913*	0.915*	-0.036	-0.026	0.189*	0.206*
MSAVI	0.208*	0.965*	0.998*	0.996*	1	0.971*	0.970*	0.942*	-0.868*	0.902*	0.913*	-0.026	-0.018	0.171*	0.203*
NDMI	0.222*	0.951*	0.972*	0.967*	0.971*	1	0.991*	0.938*	-0.888*	0.909*	0.927*	-0.017	-0.007	0.190*	0.184*
NBR	0.213*	0.964*	0.974*	0.973*	0.970*	0.991*	1	0.966*	-0.879*	0.923*	0.932*	-0.004	-0.021	0.221*	0.196*
NBR2	0.228*	0.941*	0.948*	0.943*	0.942*	0.938*	0.966*	1	-0.888*	0.898*	0.921*	0.030	-0.031	0.225*	0.216*
CI	-0.294*	-0.889*	-0.881*	-0.855*	-0.868*	-0.888*	-0.879*	-0.888*	1	-0.877*	-0.957*	0.081	0.012	-0.201*	-0.213*
GDVI	0.196*	0.932*	0.916*	0.913*	0.902*	0.909*	0.923*	0.898*	-0.877*	1	0.979*	-0.141*	-0.026	0.276	0.196*
NGRDI	0.244*	0.943*	0.929*	0.915*	0.913*	0.927*	0.932*	0.921*	-0.957*	0.979*	1	-0.124*	-0.021	0.255	0.210*
Elevation	-0.016	-0.088	-0.045	-0.036	-0.026	-0.017	-0.004	0.030	0.081	-0.141*	-0.124*	1	0.054	-0.105	0.158*
Slope	-0.023	-0.027	-0.024	-0.026	-0.018	-0.007	-0.021	-0.031	0.012	-0.026	-0.021	0.054	1	0.039	0.094
Rainfall	0.105	0.222*	0.187*	0.189*	0.171*	0.190*	0.221*	0.225*	-0.201*	0.276*	0.255*	-0.105	0.039	1	0.313*
Temperature	0.287*	0.217*	0.207*	0.206*	0.203*	0.184*	0.196*	0.216*	-0.213*	0.196*	0.210*	0.158*	0.094	0.313*	1

*significant

7.3.2 Selection of optimally MSV disease predictor variables

Of the total variables considered in this study, nine were selected as the most significant variables for predicting MSV disease occurrence. These included eight remotely sensed and one climatic variable. For climatic variables only temperature was selected as significant with a p- value of 0.000. Rainfall, NDMI, elevation, slope and SAVI exhibited no effect on MSV disease occurrence with p-value greater than 0.05.

Table 7.4: Optimally selected remotely sensed and climatic variables in predicting MSV disease prevalence

Variables	P<0.05
NDVI	0.002
EVI	0.022
Temperature	0.000
MSAVI	0.035
NBR	0.024
NBR2	0.037
CI	0.008
GDVI	0.015
NGRDI	0.013

7.3.3 Probability of MSV disease occurrence as a function of the selected variables

The results in figure 3 indicate that the eight remotely sensed derivatives had a positive relationship with MSV disease occurrence. For instance, when NDVI increase, the probability of MSV disease occurrence also increases exponentially. Maize crops with low NDVI values around ± 2 had a low probability of disease occurrence of about 0.37. A similar trend is also observed for MSAVI, NBR2, NGRDI, GDVI, EVI, and NBR. However, CI exhibited an inverse relationship when compared to the other variables. For example, Low CI (0.00) has a higher probability

of disease occurrence of ± 0.95 , whereas when the CI increases to 0.6, MSV disease probability of occurrence decreases to ± 0.25 . As for climatic variables, only temperature exhibited an influence on MSV disease occurrence. For instance, a one degree Celsius increase in temperature triggered an increase in MSV disease by a magnitude of $\pm 60\%$ that is from 0.2 to 0.6 probability of occurrence. For this study rainfall had a negligible influence on MSV disease occurrence.

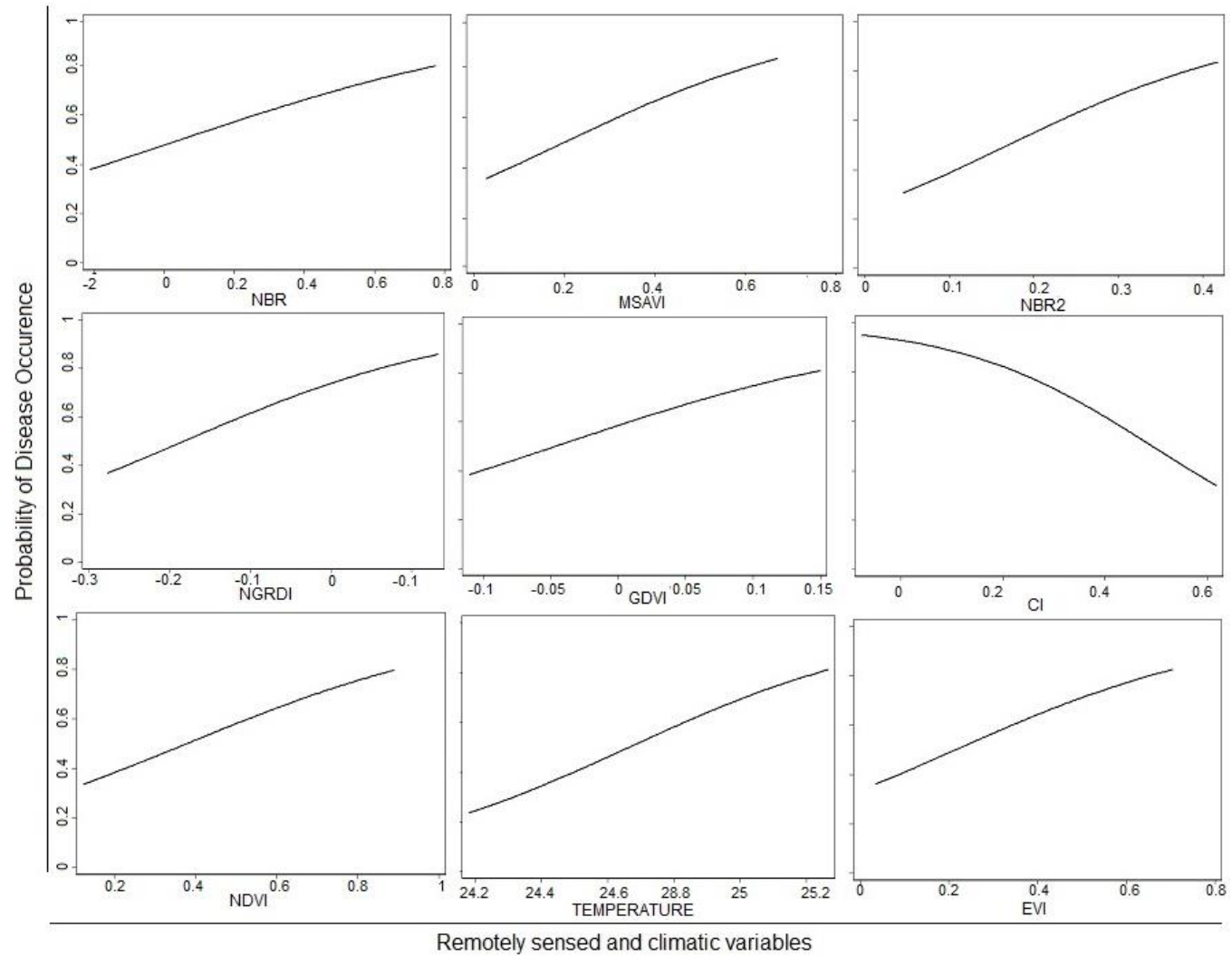


Figure 7.3: Relationship between disease occurrence and remotely sensed and climatic variables

7.4 Probability mapping

Figure 7.4 shows the areas that are likely to be affected by maize streak virus disease at Ofcolaco Farms. Most of the north eastern part of the map shows areas that are likely to be affected by maize streak virus. Areas that are likely to be affected by MSV disease are in the low lying areas with an altitude of ± 300 (m.a.s.l), while areas with high altitude are less likely to be affected by maize streak virus.

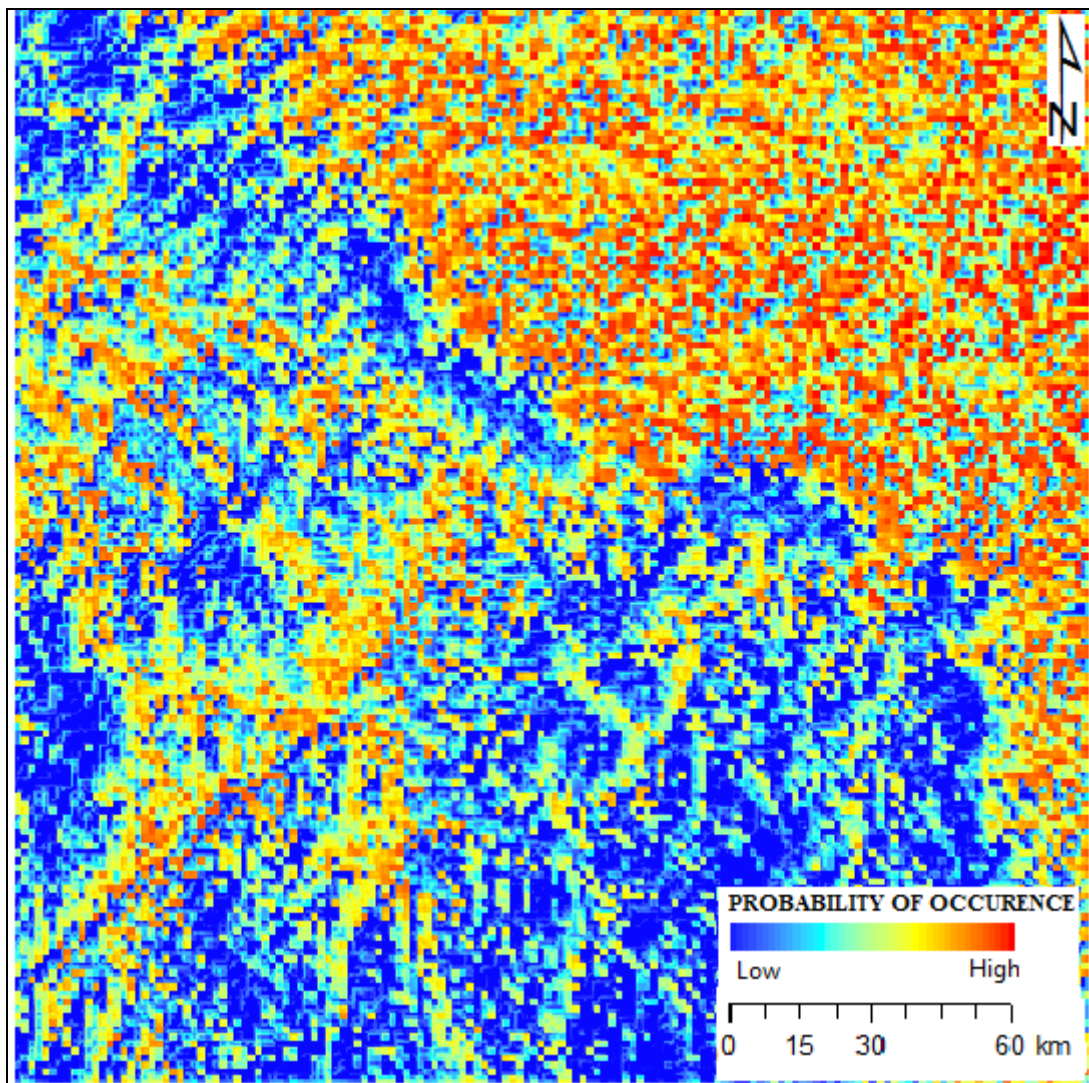


Figure 7.4: Maize streak virus probability of occurrence map

7.5 Discussion

This study sought to develop a spatially explicit model that integrates climatic, environmental and remotely sensed variables in assessing the spread of MSV disease on maize farms at the Ofcolaco Farms in Tzaneen, South Africa as an experimental site. To achieve this objective, we integrated environmental and climatic variables together with Landsat 8 derived vegetation indices to predict the probability of MSV occurrence at Ofcolaco maize Farms in Limpopo, South Africa.

The results of this study have shown that the MSV disease occurrence had a positive relationship with remotely sensed variables. For instance, an increase in NDVI, the probability of MSV disease occurrence also increases exponentially. Similarly, for MSAVI, NBR2, NGRDI, GDVI, EVI, and NBR the same pattern was observed except for CI, which had a negative correlation. Maize crops with low NDVI values had a low probability of disease occurrence while those with high NDVI had a higher probability of occurrence. Higher NDVI values might mean more hosts for MSV vectors (Shepherd et al., 2010). These results were further confirmed by Pearson's correlation coefficient results which indicated that most vegetation indices had a positive correlation with the area affected by maize streak virus disease except for CI with a negative correlation.

Furthermore, results have shown that temperature variations had an influence on the occurrence of MSV disease occurrence on maize crops at Ofcolaco farms, Tzaneen in South Africa. An increase in temperature triggered an increase in the probability of MSV disease occurrence. For instance, a one degree Celsius increase in temperature triggered an increase in MSV disease by a magnitude of $\pm 60\%$ that is from 0.2 to 0.6 probability of occurrence and the reverse is true for rainfall. This is in congruent with other studies which found out that global warming had a strong influence on disease occurrence (Shepherd et al., 2010, Matsukura et al., 2012). Temperature variations have shown an effect on the growth and survival of vegetation species which hosts both the disease and the vectors, leafhoppers. The study by (Patz and Olson, 2006) demonstrated that most diseases are sensitive to a climate which included heat-related changes or increases. The authors further highlighted that land cover change can influence micro-climatic conditions, such as

evapotranspiration and temperature, which are key determinants in the emergence of many diseases. This assertion is further supported by (Patz et al., 2000) who reported that environmental changes do affect the breeding, development, and proliferation of parasite species and their hosts, and their capacity to spread rapidly.

However, rainfall, elevation, and slope have shown no effect on the probability of MSV disease occurrence. Areas that are likely to be affected by MSV disease are in the low lying areas, while areas with high altitude are less likely to be affected by maize streak virus. This study has demonstrated the strengths and advantages of predicting the probability of occurrence of MSV disease from remote sensing data coupled with spatially environmental variables and spatial modelling techniques (Nutter Jr et al., 2010, Hill and Mayo, 1980). From the spatial analysis, the significance of all vegetation indices, and temperature variables for MSV disease presence suggests that monitoring of these variables may be vital in an effort to predict MSV outbreaks under the changing climatic conditions (Nutter Jr et al., 2010).

7.5.1 Implications of the study on maize crop disease control

The findings of this study may aid in the selection of additional information, which could be used to assist farmers to determine crop areas that likely to be affected by MSV disease. This information will help farmers to prioritize the spraying and other-related disease control mechanisms and minimize costs by avoiding the spraying of the entire farm. However, these findings should be treated with caution, as they were derived based on a single-date data, therefore some slightly different findings may be obtained during other times of the year and in different climatic conditions. Therefore, there is a need for further research incorporating all the different farm practices and crop age, as variables that may influence the probability of the MSV disease occurrence. Further studies should also investigate the role of another potential spatially derived environmental variables, such as soil moisture, crop varieties, etc., in monitoring and predicting of MSV disease. Overall, the results presented in this study provide an evidence that selected environmental factors and their interaction do influence the presence of MSV disease. There is, therefore, a

need to up-scale and replicate the model employed in this study to other MSV disease prone areas in Southern Africa.

7.6 Conclusion

The results from this study indicate that local factors, such as vegetation indices, and temperature, are critical in predicting the occurrence of MSV disease. In particular, the results reveal that under some circumstances when temperature or NDVI increases, the probability of MSV disease occurrence increases. In this regard, the results of this study can be used to better understand factors influencing the spatial distribution of MSV disease and contribute towards the development of site-specific control strategies. Since this study on maize disease monitoring, utilized medium spatial resolution data and environmental variables for maize streak virus disease, the researcher, therefore, recommend future studies to assess the applicability of high-resolution data, with a large foot-print in detecting and monitoring other complex maize diseases, such as Grey leaf spot.

Acknowledgments

The author would like to thank the Ofcolaco communities and farmers for their support during field data collection as well as for allowing us to use their farms as experimental sites. The author would also like to thank the South African weather services for the provision of climatic data. NASA is also appraised for providing Landsat and DEM data. A word of appreciation goes to VLIR-IUC Programme for providing logistical support.

7.8 References

- BOSQUE-PÉREZ, N. A. 2000. Eight decades of maize streak virus research. *Virus research*, 71, 107-121.
- DABROWSKI, Z., NWILENE, F. & KUMAR, R. 1991. First regular observations on leafhoppers, Cicadulina spp., vectors of maize streak virus (MSV) in southeastern Nigeria. *International Journal of Tropical Insect Science*, 12, 249-261.
- DONATELLI, M., MAGAREY, R., BREGAGLIO, S., WILLOCQUET, L., WHISH, J. & SAVARY, S. 2017. Modelling the impacts of pests and diseases on agricultural systems. *Agricultural Systems*.
- FAJEMISIN, J., KIM, S., EFRON, Y. & ALAM, M. 1984. Breeding for durable disease resistance in tropical maize with special reference to maize streak virus. *FAO Plant Production and Protection Paper*, 55, 49-71.
- HAMERSCHLAG, K. & KAPLAN, J. 2007. More integrated pest management please, how USDA could deliver greater environmental benefits from Farm Bill conservation programs. *Natural Resources Defense Council, New York*.
- HASAN, M., AHMAD, M., RAHMAN, M. & HAQUE, M. 2009. Aphid incidence and its correlation with different environmental factors. *Journal of the Bangladesh Agricultural University*, 7, 15-18.
- HESS, G., RANDOLPH, S., ARNEBERG, P., CHEMINI, C., FURLANELLO, C., HARWOOD, J., ROBERTS, M. & SWINTON, J. 2002. Spatial aspects of disease dynamics. *The ecology of wildlife diseases*, 102-118.
- HILL, R. E. & MAYO, Z. 1980. Distribution and Abundance of Corn Rootworm Species 1 as Influenced by Topography and Crop Rotation in Eastern Nebraska 2. *Environmental Entomology*, 9, 122-127.
- IQBAL, J., ASHFAQ, M., MANSOOR, U. H., SAGHEER, M. & NADEEM, M. 2010. Influence of abiotic factors on population fluctuation of leaf hopper, Amrasca biguttula biguttula (Ishida) on Okra. *Pakistan J. Zool*, 2, 615-621.
- MAHUKU, G., LOCKHART, B. E., WANJALA, B., JONES, M. W., KIMUNYE, J. N., STEWART, L. R., CASSONE, B. J., SEVGAN, S., NYASANI, J. O. & KUSIA,

- E. 2015. Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. *Phytopathology*, 105, 956-965.
- MATSUKURA, K., YOSHIDA, K. & MATSUMURA, M. 2012. Estimation of climatic factors relating to occurrence of the maize orange leafhopper, *Cicadulina bipunctata*. *Population ecology*, 54, 397-403.
- MONJANE, A. L., HARKINS, G. W., MARTIN, D. P., LEMEY, P., LEFEUVRE, P., SHEPHERD, D. N., OLUWAFEMI, S., SIMUYANDI, M., ZINGA, I. & KOMBA, E. K. 2011. Reconstructing the history of maize streak virus strain a dispersal to reveal diversification hot spots and its origin in southern Africa. *Journal of virology*, 85, 9623-9636.
- NELSON, M. R., ORUM, T. V., JAIME-GARCIA, R. & NADEEM, A. 1999. Applications of geographic information systems and geostatistics in plant disease epidemiology and management. *Plant disease*, 83, 308-319.
- NUTTER JR, F. W., VAN RIJ, N., EGGENBERGER, S. K. & HOLAH, N. 2010. Spatial and temporal dynamics of plant pathogens. *Precision Crop Protection- the Challenge and Use of Heterogeneity*. Springer.
- PATZ, J. A., GRACZYK, T. K., GELLER, N. & VITTOR, A. Y. 2000. Effects of environmental change on emerging parasitic diseases. *International journal for parasitology*, 30, 1395-1405.
- PATZ, J. A. & OLSON, S. H. 2006. Climate change and health: global to local influences on disease risk. *Annals of Tropical Medicine & Parasitology*, 100, 535-549.
- ROSE, D. 1978. Epidemiology of maize streak disease. *Annual Review of Entomology*, 23, 259-282.
- SEAGLE, S. W. & MCNAUGHTON, S. 1992. Spatial variation in forage nutrient concentrations and the distribution of Serengeti grazing ungulates. *Landscape ecology*, 7, 229-241.
- SHEPHERD, D. N., MARTIN, D. P., VAN DER WALT, E., DENT, K., VARSANI, A. & RYBICKI, E. P. 2010. Maize streak virus: an old and complex 'emerging' pathogen. *Molecular plant pathology*, 11, 1-12.
- THOTTAPPILLY, G., BOSQUE-PÉREZ, N. & ROSSEL, H. 1993. Viruses and virus diseases of maize in tropical Africa. *Plant Pathology*, 42, 494-509.

Chapter 8

TESTING THE POTENTIAL OF MULTISPECTRAL SENSORS IN DETECTING GREY LEAF SPOT DISEASE OF MAIZE

This Chapter is based on:

Inos Dhau, Elhadi Adam, Onesimo Mutanga, Kwabena Ayisi, Elfatih Mohamed Abdel-Rahman, John Odindi & Mhosisi Masocha (2017) Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop, Geocarto International,

DOI: [10.1080/10106049.2017.1343391](https://doi.org/10.1080/10106049.2017.1343391) Pages 1-14

Abstract

Development of techniques for early detection of maize grey leaf spot (GLS) infection is valuable in preventing crop damage and minimising yield loss. In this study, the researcher tested whether GLS field symptoms on maize can be detected using hyperspectral field spectra resampled to different sensor resolutions. First, field spectra were acquired from healthy, moderate and severely infected maize leaves during the 2013 and 2014 growing seasons. The spectra were then resampled to four sensor spectral resolutions -WorldView-2, Quickbird, RapidEye, and Sentinel-2. In each case, the Random Forest algorithm was used to classify the 2013 resampled spectra to represent the three identified disease severity categories. Classification accuracy was evaluated using an independent test dataset obtained during the 2014 growing season. Results showed that Sentinel-2, with 13 spectral bands, achieved the highest overall accuracy and a kappa value of 84% and 0.76, respectively while the WorldView-2, with 8 spectral bands, yielded the second highest overall accuracy and a kappa value of 82% and 0.73, respectively. Results also showed that the 705 and 710nm red edge bands were the most valuable in detecting the GLS for Sentinel-2 and RapidEye, respectively. On the resampled WorldView 2 and Quickbird sensor resolutions, the respective 608 nm and 660 nm in the yellow and red bands were identified as the most valuable for discriminating all categories of infection. Overall, the results imply that opportunities exist for developing operational remote sensing systems based on multispectral sensors, especially Sentinel-2 and WorldView-2 for early detection of GLS. Adoption of such datasets is particularly valuable for minimizing crop damage and improving yield.

Key Words: field spectroscopy, grey leaf spot, spectral resampling, multispectral remote sensing

8.0 Introduction

Maize is the most important cereal crop in Sub-Saharan Africa (SSA) and contributes 15 - 50% of energy in human diets in the region (Kagoda et al. (2011), Archetti et al., 2009a, Archetti et al., 2009b). The current consumption of maize is estimated to be about 112 kg annually per capita which is equivalent to 308 g per day per capita (Degraeve et al., 2016). Annual demand for maize in the region is projected to increase at a rate of 2.4 % per annum up to 2025 (Pinstrup-Andersen et al., 1999). Currently, house-hold and per capita maize consumption in the SSA region is the highest amongst developing countries (Pingali and Pandey, 2001) consequently measures to improve maize production are considered essential for food security in the region (Watkins and Von Braun, 2003). Although studies in SSA have shown that there is an increase in total maize production over the years (Beyene and Kassie, 2015) most of this increase has been attributed to area expansion rather than optimisation of production techniques and therefore increased yield on existing acreage (Beyene and Kassie, 2015). Beyene and Kassie (2015) attribute the low yield in the region to the slow adoption of precision agriculture and associated technologies.

In the SSA, an interplay of biotic and abiotic factors is known to cause up to 80% maize yield losses (DeVries and Toenniessen, 2001, Pingali and Pandey, 2001). Among the major diseases that threaten the stability of maize production in the region, particularly South Africa, are grey leaf spot (GLS) (Derera et al., 2008), Northern Corn Leaf Blight (NCLB) (Degefu et al., 2004, Welz and Geiger, 2000) and *Phaeosphaeria* leaf spot (PLS) (Sibiya et al., 2011). The GLS caused by *Cercospora zea-maydis* was first identified from specimens collected in 1924 by Tehon and Daniels in Alexander County in southern Illinois near the Mississippi River (Ward et al., 1999). In South Africa, GLS was first noted in the KwaZulu-Natal province in 1988 and has since spread rapidly to neighbouring provinces and other Southern Africa countries (Ward et al., 1997b). More recently molecular techniques have allowed a detailed analysis of the genetic variability of the pathogen population and the existence of two very distinct groups have been revealed (Meisel et al., 2009). A review of the literature suggests that there are two possible species complexes

associated with grey leaf spot, namely the *C. sorghi* complex (*C. sorghi* and *C. sorghi* var. *maydis*), and the *C. zea-maydis* complex (Groups I and II). The description of *C. zeina* has now resolved some of this taxonomic uncertainty by demonstrating that Group II is, in fact, a distinct species (*C. zeina*) and that Group I to which the name *C. zea-maydis* applies apparently does not occur in South Africa (Crous et al., 2006, Meisel et al., 2009, Dunkle and Levy, 2000).

It is currently recognised as one of the most severe maize damaging diseases in the region, especially in areas with warm temperatures and prolonged relative humidity (Wegary et al., 2003, Derera et al., 2008, Ward et al., 1999, Lyimo et al., 2012, Meisel et al., 2009, Paul and Munkvold, 2005). Recently a number of countries including, Kenya, Zimbabwe, and South Africa have reported 30-60% maize yield losses due to GLS, and hence considered a serious threat to food security (Muriithi and Gathama, 1998, Ward et al., 1997a). The disease reduces maize yield by damaging photosynthetic tissue and increasing stem and root lodging (Derera et al., 2008). In maize, the GLS symptoms first appear on the lower leaves as small tan spots typically long rectangular or irregular 1 to 3 mm shapes (Ward et al., 1999). At the early stage, immature GLS lesions are not easily distinguishable from lesions caused by other foliar maize pathogens (Ward et al., 1999). However mature foliar lesions symptomatic of GLS can readily be distinguished from other maize foliar diseases; grey to tan rectangular shape (5 to 70 mm long by 2 to 4 mm wide) running parallel to the leaf veins (Ward et al., 1999). The latent period for GLS is long compared to other foliar pathogens and can take as long as 14 to 28 days after infection for lesions to sporulate (Beckman and Payne, 1982). Under severe epidemics, lesions may coalesce and blight the entire leaves (Ward et al., 1999, Beckman and Payne, 1982). An integrated approach to planning and management of GLS in maize is critical for sustainable production and improvement in yield quality and quantity. This approach requires, inter alia that maize disease information is available in ways that allow for the timely adoption of relevant management practice at the right place (Geerts et al., 2006, de Bie, 2000, Mulla, 2013), Pingali & Pandey, 2001). A basic principle of precision agriculture is that the presence, distribution, and intensity of a specific stress factor within a field must be identifiable (Hillnhütter et al., 2011). The distinctive patchy appearance on maize canopy due to GLS

infestation makes the disease suitable for adoption of precision agriculture tools such as geo-information and remote sensing techniques (Hillnhütter et al., 2011).

Visual field-based observations have been used to determine GLS severity and incidences (Nilsson, 1995, Munkvold et al., 2001, Elwinger et al., 1990, Bubeck et al., 1993), however, such visual protocols are often expensive, time-consuming and prone to human error. The assessment of disease severity and its real-time field distribution could be valuable for adoption and timing of relevant mitigation measures like fungicide applications. Since the GLS is a foliar fungal disease, and plant leaves have a well-known spectral signature, remote sensing technologies can be used to complement field-based protocols in determining infection onset and severity (Nutter Jr and Schultz, 1995).

The use of remote sensing in plant pathology and crop protection is well established (Bock et al., 2010, Hatfield and Pinter, 1993, Jackson, 1986, Nilsson, 1995, West et al., 2003, Mahlein et al., 2012). For example, field spectroscopy has been used to detect maize dwarf mosaic viral disease (Ausmus and Hilty, 1973), wheat yellow rust (Bravo et al., 2003, Moshou et al., 2004), powdery mildew in winter wheat (Zhang et al., 2012), thrips in green-peel citrus (Dong et al 2014) and bacterial leaf blight in rice (Yang, 2010). Several studies have described the views of sensing leaf reflectance in the visible (VIS, 400-700 nm), near infrared (NIR, 700-1000 nm) and short wave infrared (SWIR, 1000-2500 nm) for detecting changes in plant vitality with emphasis on fungal plant diseases using non-imaging spectroradiometers (Delalieux et al., 2007, Mahlein et al., 2010, Rumpf et al., 2010, Steddom et al., 2005). Whereas field spectroscopy has proved to be useful in disease detection, up-scaling the developed models to the current available spaceborne or airborne sensors remains a common challenge. This is due to the fact that most of the currently operational satellites lack the necessary fine spectral resolution sensors, hence the application of satellite remote sensing for crop disease detection is still in its infancy.

Recently, some researchers have suggested that this limitation could be overcome by resampling field spectral data to band settings of existing sensors. For mapping relevance, this can be followed by testing their capability in existing coarser spectral resolution formats for different remote sensing applications, where hyperspectral

data are not available (Slaton et al., 2001, Cho and Skidmore, 2009, Crawford et al., 2003). Indeed, several studies (Petropoulos et al., 2012, Adam et al., 2012b, Mansour et al., 2012a, Mutanga and Skidmore, 2005) have resampled field spectra data to band settings of existing or planned sensors to broaden their remote sensing applications. Whereas these studies have been able to demonstrate the possibility to up-scale field spectra to airborne or satellite sensors, a major challenge of such up-scaling remains the field spectrometer's higher signal to noise ratio than satellite or airborne images, due to the former's shorter signal path length (Milton et al., 2009). Nevertheless, Mutanga et al. (2015) recently provided an insight on the magnitude of errors expected when up-scaling field spectral models to airborne or satellite image. They note that the minor difference between the model developed on the resampled field spectra in comparison to the actual image indicates the relevance of field spectroscopy and spectral resampling in mapping (Mutanga et al., 2015). The advent of new multispectral sensors such as RapidEye, Sentinel, and WorldView-series is seen as a trade-off between benefits offered by multispectral and hyperspectral remotely sensed data. This is mainly due to their fine spatial resolution and a reasonable number of spectral bands, which are configured within unique portions of the electromagnetic spectrum such as the red edge. These new sensors provide an opportunity for more crop monitoring and precision agriculture.

In this study, the researcher sought to determine whether field spectrometry measurements resampled to different multi-spectral sensor resolutions can be used to detect the severity of GLS infection in maize. The specific objectives were: (i) to test if the severity of GLS can be detected by hyperspectral data resampled to Sentinel-2, Worldview-2, RapidEye, and Quickbird sensor resolutions and (ii) to determine the best spectral bands that are important for detecting GLS. Reliable results from the study offer an opportunity for mapping and monitoring GLS infestations in maize crop using sensors on aerial and satellite platforms.

8.1 Material and methods

8.1.1 Study Area

This study was carried out at Cedara Agricultural Research Farm, located at 1076 metres above sea level and approximately 100 km in-land of South Africa's major port city of Durban in KwaZulu-Natal province (30°16 East, 29°32South). GLS symptoms were carefully identified in the field over two seasons in March 2013 and 2014 for spectral data acquisition. The GLS stages were then assessed with the help of the plant pathologist based on visual leaf assessment (Bock et al., 2010, Sibiya et al., 2011, Bubeck et al., 1993, Elwinger et al., 1990, Munkvold et al., 2001). Disease severity was assessed by estimating percent leaf area affected by disease (PLAA). Ratings were recorded on leaves using a modified scale of 0-5 as described by Paul and Munkvold, (2004) where (0= no disease; 1 =10-20%; 2 = 21-40%; 3= 41-60%; 4= 61-80% and 5 = 81-100%). As recommended by Nutter et al. (1991), leaves with less than 20% GLS surface area were considered to be in an early stage of infestation while leaves with 21 to 60% GLS infestation were considered to be in the moderate stage of infestation. Leaves with 61% to 100 % were considered to be severely infested (Figure 8.1). The visual leaf assessment was benchmarked by plotting Normalised Difference Vegetation Index (NDVI) box plots of the percentage lesion area to show the spread of data that was collected over a period of two years as shown on Figure 8.1 below. The boxplots reveal that the spread of data during the two seasons was nearly the same and statistical analysis revealed that there were no significant differences between the means of the same classes for the two seasons.

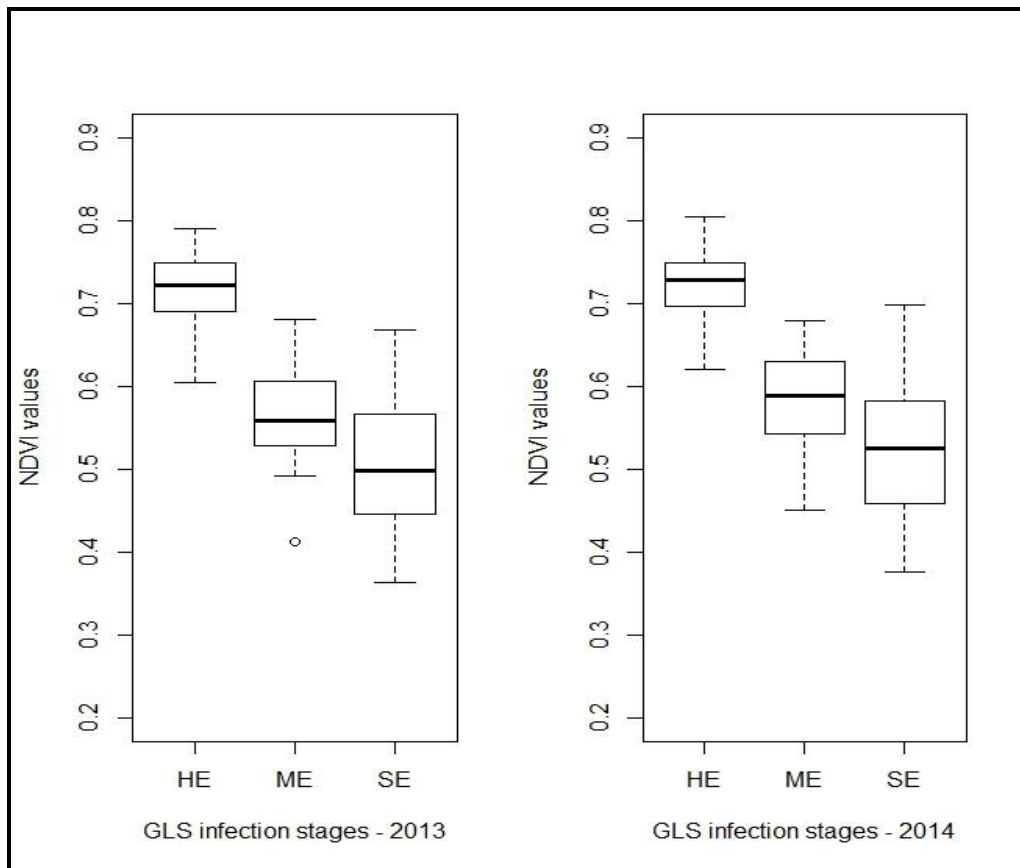


Figure 8.1: Boxplots showing the median, quartile data and the spread of data for healthy, moderate and severely infected maize with grey leaf spot for the two seasons. NDVIs were calculated from the field spectroradiometer measurements.

8.1.2 Field spectral measurements and processing

Three stages; healthy, moderate and severely infested maize leaves were first identified with the help of a plant pathologist in the field. Spectral measurements from the three stages were then obtained from 10:00 to 14:00 local time (GMT+2) under clear sunny and cloud-free sky using a non-imaging Spectroradiometer, Analytical Spectral Devices (ASD) FieldSpec® 3 optical sensor (Analytical Spectral Devices, Inc., Boulder, CO, USA). The ASD FieldSpec® 3 spectrometer has a 350 - 2,500 nanometers (nm) spectral range, with 1.4 nm and 2 nm sampling intervals for the ultraviolet to visible-near infrared region (350–1,000 nm) and the short-wave infrared region (1,000–2,500nm), respectively.

Plant and leaf samples representing each of the GLS infestation stages were chosen and six measurements per leaf acquired. The six measurements were then averaged to derive the representative reflectance spectra for the leaf. A white reference spectral measurement on the calibration panel was performed every 10–20 measurements to offset any change in the atmospheric condition and sun irradiance spectrum. Field spectral measurements were done over two (2013 and 2014) growing seasons. The field spectra for the year 2013 were used for training the random forest classifier model while the 2014 field spectra were used to test the stability and the accuracy of the random forest model. The sample size for each class is shown in Figure 8.2.


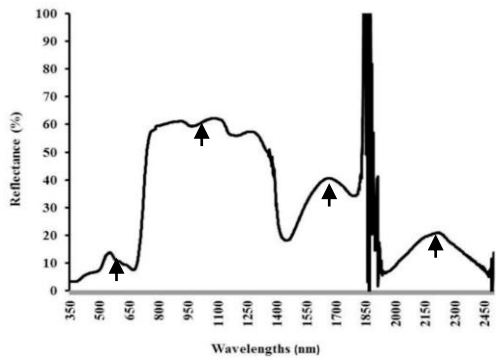

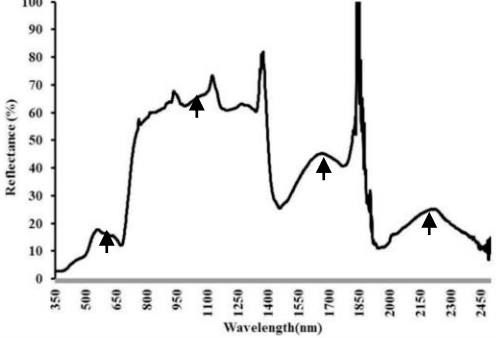

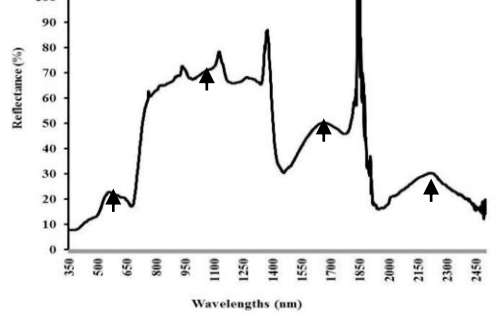
Disease Status	Images	Mean reflectance	Training Samples (2013)	Testing samples (2014)
Healthy			88	68
Moderate			92	70
Severe			86	75

Figure 8.2: Disease status, mean reflectance and number of samples for healthy, moderate and severe maize grey leaf spot infestation for the spectroradiometer measurements and multispectral sensors. The arrows show where major differences between the three disease stages are located

The spectral measurements from the three GLS infestation stages (Figure 8.3) were then resampled to the Sentinel-2, WorldView-2, Quickbird and RapidEye spectral resolutions using a Gaussian model in ENVI 4.7 image processing software (ENVI, 2009). The method uses a Gaussian model with a full width at half maximum (FWHM) equal to the band centres provided (Table 8.1). The resampled spectra were then used for the further analysis.

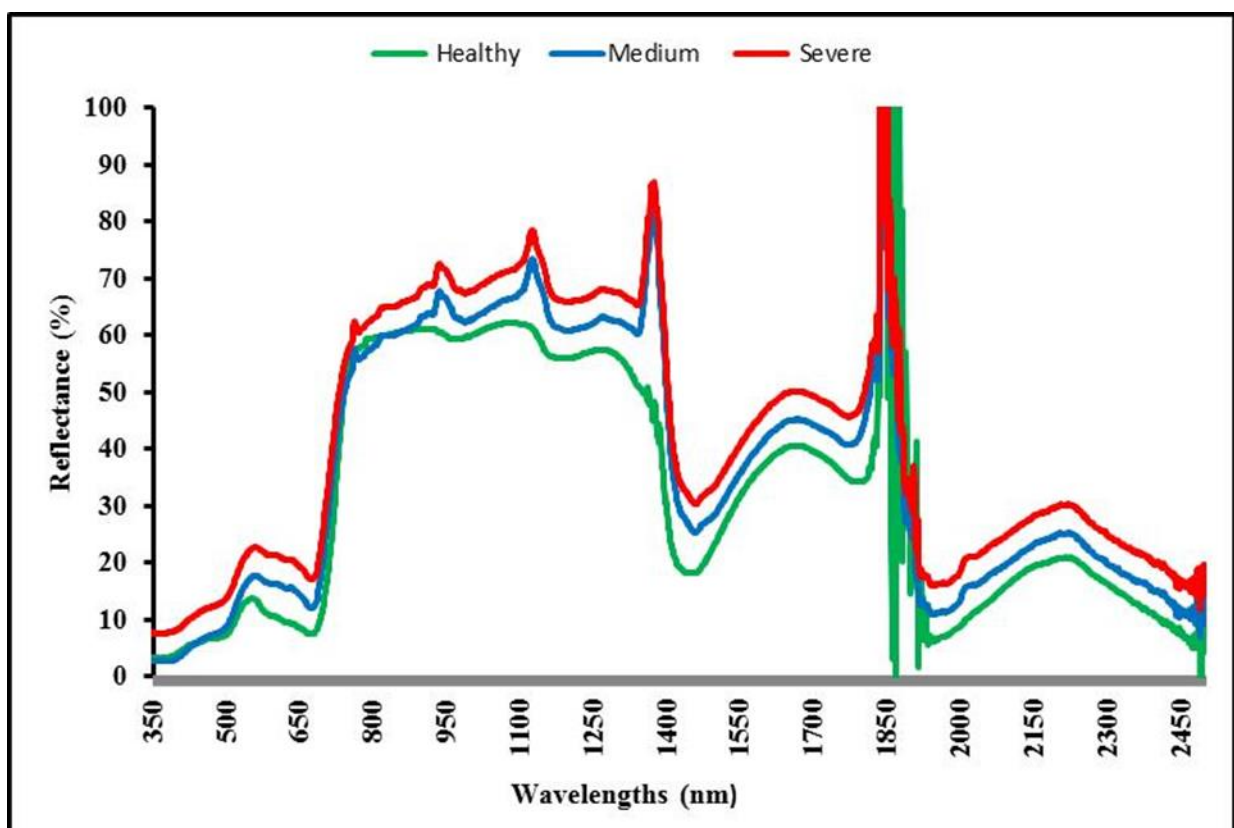


Figure 8.3: Comparison between the mean reflectance of the three GLS infestation stages (healthy, moderate and severely).

Table 8.1: Multispectral sensors and their spectral properties.

Sensor	Band description	Band centre (nm)	bandwidth (nm)
Sentinel 2	B1	443	20
	B2	490	65
	B3	560	35
	B4	665	30
	B5	705	15
	B6	740	15
	B7	783	20
	B8	842	115
	B8b	865	20
	B9	945	20
	B10	1380	30
	B11	1610	90
Worldview-2	B12	2190	120
	Coastal	427	50
	Blue	478	60
	Green	546	70
	Yellow	608	40
	Red	659	60
	Red-edge	725	40
	NIR-1	831	125
NIR-2	908	180	
Quickbird	Blue	485	70
	Green	560	80
	Red	660	60
	NIR	830	140
RapidEye	Blue	475	70
	Green	555	70
	Red	658	55
	Red-edge	710	40
	NIR	805	90

8.2 Random forest classifier and accuracy assessment

The random forest (RF) algorithm was used to classify the resampled spectra to represent the three identified disease severity categories. Random forest is an ensemble learning technique developed by Breiman (2001) to improve the classification and regression of trees (CART) by combining a large set of decision trees. RF grows multiple unpruned trees (*n_{tree}*) on bootstrap samples of the original data (Breiman 2001). Each tree is grown on a bootstrap sample (2/3 of the original data known as “in-bag” data) taken with replacement from the original data (Breiman, 2001a). The samples that are not in the bootstrap sample are referred to as the out-of-bag (OOB) sample. The OOB sample (~37% of the total data) can be used to estimate the misclassification error and variable importance (Breiman 2001). Trees are split to many nodes using random subsets of variables (*mtry*); the default *mtry* value is the square root of the total number of variables. From the *mtry* selected variables, the variable that yields the highest decrease in impurity is chosen to split the samples at each node (Breiman, 2001). A tree is grown to its maximum size without pruning until the nodes are pure. This means that the nodes hold samples of the same class or contain a certain number of samples. A prediction of the response variable (e.g., GLS stages) is made by combining the prediction over all trees. In a classification, a majority vote from all the trees in the ensemble determines the final prediction (Breiman, 2001). A more detailed description of RF can be found in among others Breiman (2001) and Touw et al. (2012).

RF also provides variable importance measurement, which shows the contribution of each variable (band) in the final predictive model. This is a measure of how much classification accuracy would decrease if data of a particular variable (band) were removed while all variables remained the same (Breiman, 2001a, Verikas et al., 2011). In this study, variable importance measurement was used to determine the importance of each band from the different sensors in detecting GLS infestation. The choice of the random forest algorithm was motivated by its superiority over competing for statistical methods like the discriminant analysis, classification trees, support vector machines, and principal component analysis (Adam and Mutanga, 2009a, Cochrane, 2000, Mutanga and Skidmore, 2004).

In RF, only two factors (i.e., *mtry* and *ntree*) have to be optimized. The default number of trees (*ntree*) is usually 500, while the default value for the number of variables (*mtry*) is \sqrt{p} where P equals the number of predictor variables within a data set. In this study, a grid-search approach based on the OOB estimate of error was used to find the optimal combination of these two parameters (*ntree* and *mtry*) (Adam et al., 2014). The grid-search value for *mtry* was varied from 1 to the total number of bands for each sensor with a single value interval, while the range of the grid search value for the *ntree* parameter was varied from 500 (default value) to 10,000 with an interval of 500 (20 steps). For example, the grid search yielded a total of 160 and 80 combinations of *ntree* and *mtry* values for WV-2 and Quickbird, respectively. The dataset collected in 2013 was used to optimize and train a random forest model.

The independent test data (sample size is shown on figure 2) which was collected in 2014 was used to assess the accuracy of the random forest classifier. A confusion matrix was constructed to compute the overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) as criteria for evaluating the generalization ability (accuracy) of the RF classifiers (Mather and Tso 2003). OA is a ratio (%) between the number of correctly classified samples and the number of test samples, while UA represents the likelihood that a sample belongs to the specific class and the classifier accurately assigns it to such a class. PA expresses the probability of a certain class being correctly classified. Kappa analysis which uses the K statistic was also calculated to determine if one error matrix is significantly different from another. The kappa coefficient provides a measure of the actual agreement between reference data and a random classifier. If the kappa coefficients are equal or close to one, then there is perfect agreement (Congalton and Green 2008).

8.3 Results

8.3.1 Optimizing random forest parameters

Results from the grid search indicated that the OOB error rate is sensitive to *n*tree and *m*try parameters. Different combinations of *n*tree and *m*try produced the lowest OOB error rates for the different sensors. The default settings for both *n*tree and *m*try did not produce better results for the different sensors. For example, the combination of *m*try value of 4 and *n*tree value of 500 produced lowest OOB error for Quickbird, while the combination of *m*try value of 10 with different *n*tree values (3000, 4500 and 9500) produced lowest OOB error (0.08) for Sentinel-2 (Figure 8.4). The highest OOB error rates (0.13 to 0.16 %) were produced with the lowest *m*try values across different sensors (Figure 8.4).

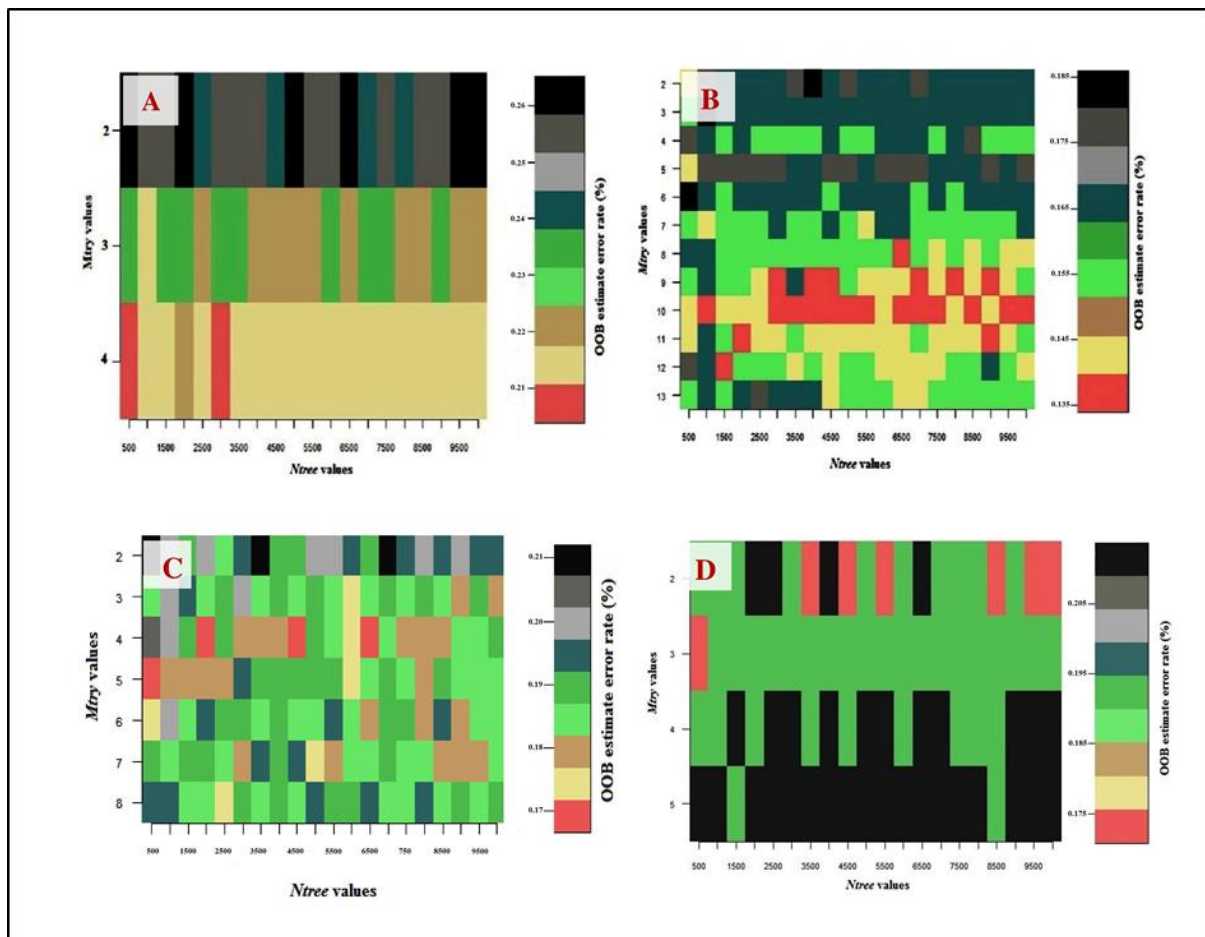


Figure 8.4: Optimization of the random forest parameters (*m*try and *n*tree) using the grid search procedure. The OOB method was used to determine the error rates for the different combinations; 60 combinations for Quickbird (A), 240 combination for Sentinel-2 (B), 160 combinations for Worldview-2 (C) and 80 combinations for RapidEye (D).

8.3.2 Classification results

The classification accuracies obtained using field spectra resampled to Worldview-2, Quickbird, RapidEye and Sentinel 2 sensor resolutions for both training and testing datasets are shown in Table 8.2. RF produced the highest overall accuracy (84.04 %) and a kappa value of 0.76 for Sentinel-2. On the other hand, the lowest overall accuracy (76.77 %) was achieved when Quickbird was used. It is also worth noting that the difference between the results of the training dataset and the test dataset for all sensors is about ± 2 %, which can be considered insignificant. Table 8.3 provides further details about the confusion matrix, producer and user accuracies achieved for Sentinel-2, the sensor with the highest accuracy.

Table 8.2: Overall accuracy (OA) and Kappa index of agreement achieved by field spectroradiometer and four multispectral sensors calculated using the 2013 training and an independent 2014 testing datasets

Sensor	Training data		Testing data	
	OA	KAPPA	OA	KAPPA
Field Spectroradiometer	92.8	0.87	89.2	0.83
Worldview-2	84.7	0.77	82.16	0.73
Quickbird	78.65	0.68	76.77	0.66
RapidEye	82.77	0.74	81.22	0.72
Sentinel-2	86.47	0.80	84.04	0.76

Table 8.3: Confusion matrix used to develop producer's and user's accuracy (%) for Sentinel-2 for the different levels of maize grey leaf spot (H- Healthy, M-Moderate and S-Severe). OOB method was used to evaluate the accuracy on the independent test dataset acquired in 2014.

Class	Testing data 2014			
	H	M	S	Total
H	59	6	3	68
M	2	64	4	70
S	5	14	56	75
Total	66	84	63	213

Overall Accuracy = 84.04

Kappa = 0.76

Producer Accuracy = 89.39

User Accuracy = 86.76

8.3.3 Measuring the importance of variables (bands) in detecting the GLS stages in maize

Figure 8.5 shows the most important bands in each sensor for detecting GLS in maize leaves. The importance of each band was measured using the mean decrease in accuracy in random forest algorithm (Figure 8.5). The most important bands are located at different positions for each sensor. For example, the red edge (705 nm) was found to be the most important (highest mean decrease in accuracy) band for Sentinel-2. Using the Worldview-2, the yellow band located at 608 nm was the most important. The most important band for detecting the GLS using Quickbird was the red (660nm) while red (658nm) and the red edge bands (710 nm) for RapidEye were the most important bands (Figure 8.5).

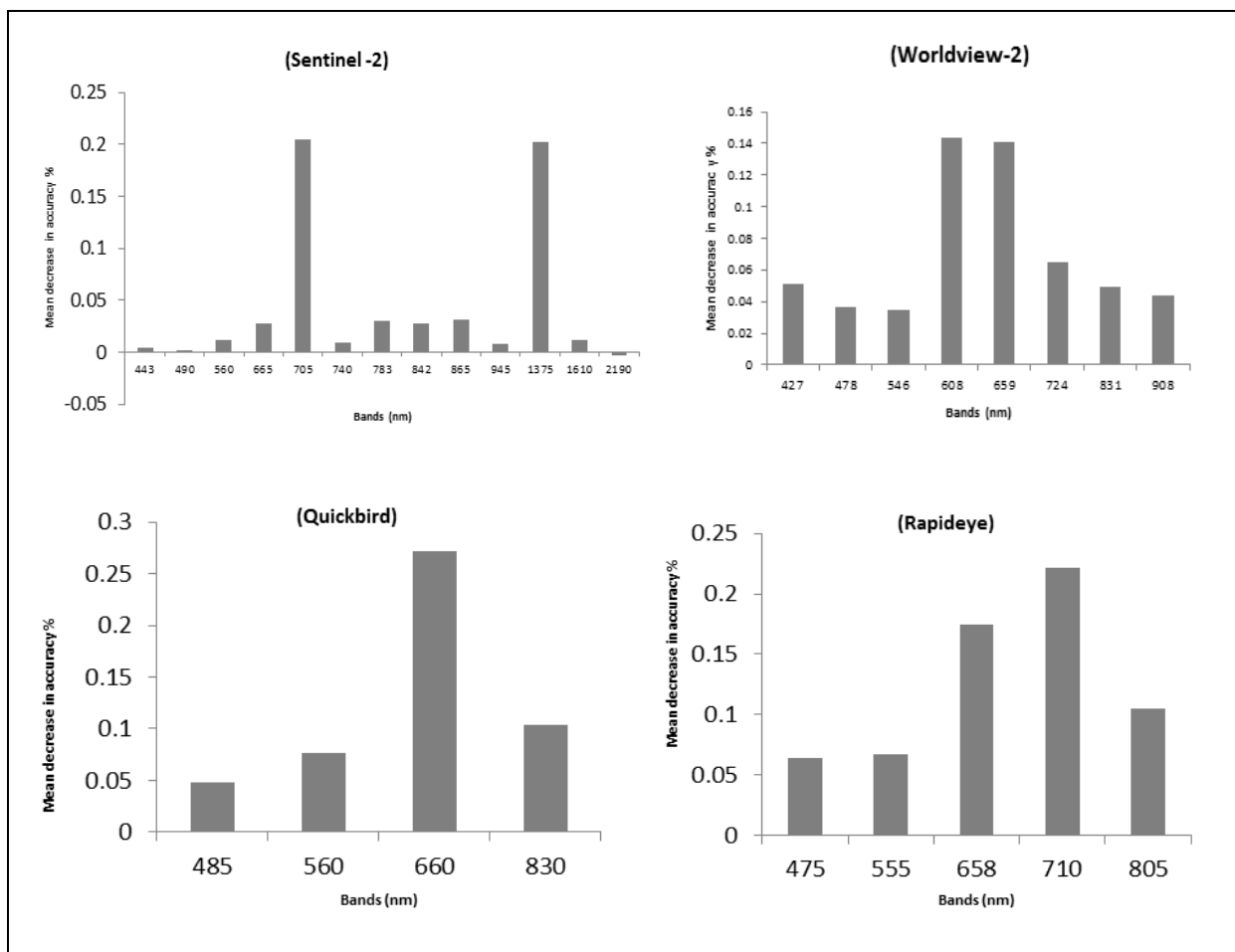


Figure 8.5: The importance of each band for different sensors used in this study for distinguishing the three stages of GLS infestation in maize using random forest (RF).

8.4 Discussion

The GLS is considered one of the most serious diseases affecting maize production in many countries, particularly in North America and Africa. Timely detection and control of GLS require intensive spatial data collection for designing optimal mitigation strategies (Mulla, 2013). A number of studies have demonstrated the importance of remote sensing data in defining the spatial variability of pests and diseases in crops (Falkenberg et al., 2007, Qin and Zhang, 2005, Yang, 2012, Zhang et al., 2003). The main objective of this study was to test the use of the spectral resolution of the newly developed multispectral sensors in detecting the severity of GLS in maize. Results indicate that different stages of GLS can be reliably discriminated using the newly developed multispectral sensors such as Worldview-2, Sentinel-2, and RapidEye.

The spectral and spatial resolutions of the newly developed multispectral sensors offer significant potential in real time data provision, necessary for precision agriculture. The newly developed multispectral sensors can be considered a compromise between multispectral and hyperspectral sensor limitations, useful for the adoption of improved agricultural management practices. Hyperspectral remotely sensed data is costly and difficult to process. However, it provides the ability to investigate spectral response of different vegetation stresses in narrow spectral bands (10 nm wide) within a wide spectral range, which is otherwise masked by relatively cheap, broad, and limited bands inherent in existing multispectral sensors.

Results in this study suggest that the new generation Sentinel-2's 705nm red edge band, the Worldview-2's 608 nm yellow band, the Quickbird's 660nm red and RapidEye's 710nm red edge bands are all suitable for detecting GLS disease. However, the resampled Sentinel-2 dataset can discriminate healthy leaves from GLS infested leaves with the highest OA (84%) and a Kappa value over 0.76. Results in this study also suggest that the Worldview-2 is valuable for detecting GLS under field conditions (OA of 82% and Kappa of 0.73). These results demonstrate for the first time that the Sentinel-2 wavebands are strategically positioned for agricultural applications, suitable for discriminating healthy maize from GLS infested

maize as well as different levels of infestation. Conversely, the OA achieved in this study suggests that existing multispectral sensors such as QuickBird, RapidEye, and Worldview-2, in comparison to Sentinel-2, are less optimally positioned, and therefore less suited for detecting GLS infestation. Overall, the results highlight the value of multispectral image data in detecting and mapping plant diseases. Although the Sentinel-2 achieved the highest accuracy, due to its high spectral resolution (13bands), its spatial resolution (10m) is coarser than the Worldview-2 (2m). Thus, Worldview-2 provides a viable up-scaling option for mapping GLS as it is characterised by a higher spatial resolution and an acceptable level of classification accuracy achieved with fewer bands than Sentinel-2.

The results concur with Delwiche and Kim (2000) who assessed the spectral characteristics of Fusarium head blight disease in winter wheat and established that the reflectance at 550, 568, 605, 623, 660, 697, 715 and 733 nm are the best disease presence indicators. (Moshou et al., 2004) reported that the reflectance at 680, 725 and 750 nm could be used to detect yellow rust in winter wheat while (Yang et al., 2005) established that the band centered at 694 nm can be used to detect green bug infestation. (Graeff et al., 2006) successfully distinguished powdery mildew infested from healthy wheat leaves with reflectance at 490, 510, 516, 540, 780, and 1300nm during early infestation.

The high accuracy achieved in this study contributes towards the development of operational remote sensing systems on advanced multispectral sensors. This approach is valuable for mapping crop diseases, particularly in Africa where the high cost of hyperspectral image datasets has impeded wide adoption of remotely sensed imagery. These findings may be useful in assessing the spectral configurations of new sensors under development and inform spectral adjustments and optimization for specific applications on the up-scaled image data (Adjorlolo et al., 2012b, Mutanga et al., 2005).

8.5 Conclusions

This paper aimed at discriminating GLS infestation severity on maize using field spectrometry data resampled to different sensor resolutions.

The results demonstrate that:

1. Field symptoms of GLS on maize can be detected by hyperspectral field spectra resampled to different sensor resolutions; and,
2. Using the RF algorithm, Sentinel-2 and Worldview-2 have a higher and acceptable accuracy level, respectively, in discriminating GLS using the red, red edge, and SWIR regions of the spectrum.

Overall, the results open up opportunities for discriminating GLS in maize using sensors such as Sentinel-2 and Worldview-2. The study, however, recommends that more studies be done to investigate the use of both spectral and spatial resolution to detect the spatial distribution of the GLS infestations on maize. The use of new generation multispectral sensors offers a great opportunity for detection and mapping of GLS from space, offering real-time operational data, ultimately improving strategies for mitigating crop damage and yield loss.

8.6 References

- ADAM, E. & MUTANGA, O. 2009. Spectral discrimination of papyrus vegetation (i Cyperus papyrus L.i) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 612-620.
- ADAM, E., MUTANGA, O., ODINDI, J. & ABDEL-RAHMAN, E. M. 2014. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458.
- ADAM, E. M., MUTANGA, O., RUGEGE, D. & ISMAIL, R. 2012. Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing*, 33, 552-569.
- ADJORLOLO, C., MUTANGA, O., CHO, M. & ISMAIL, R. 2012. Challenges and opportunities in the use of remote sensing for C3 and C4 grass species discrimination and mapping. *African Journal of Range & Forage Science*, 29, 47-61.
- ARCHETTI, M., D RING, T. F., HAGEN, S. B., HUGHES, N. M., LEATHER, S. R., LEE, D. W., LEV-YADUN, S., MANETAS, Y., OUGHAM, H. J. & SCHABERG, P. G. 2009a. Unravelling the evolution of autumn colours: an interdisciplinary approach. *Trends in ecology & evolution*, 24, 166-173.
- ARCHETTI, M., DÖRING, T. F., HAGEN, S. B., HUGHES, N. M., LEATHER, S. R., LEE, D. W., LEV-YADUN, S., MANETAS, Y., OUGHAM, H. J. & SCHABERG, P. G. 2009b. Unravelling the evolution of autumn colours: an interdisciplinary approach. *Trends in Ecology & Evolution*, 24, 166-173.
- AUSMUS, B. S. & HILTY, J. W. 1973. Reflectance studies of healthy, maize dwarf mosaic virus-infected, and i Helminthosporium maydis i infected corn leaves. *Remote Sensing of Environment*, 2, 77-81.
- BECKMAN, P. M. & PAYNE, G. A. 1982. External growth, penetration, and development of Cercospora zae-maydis in corn leaves. *Phytopathology*, 72, 810-815.

- BEYENE, A. D. & KASSIE, M. 2015. Speed of adoption of improved maize varieties in Tanzania: An application of duration analysis. *Technological Forecasting and Social Change*, 96, 298-307.
- BOCK, C., POOLE, G., PARKER, P. & GOTTWALD, T. 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29, 59-107.
- BRAVO, C., MOSHOU, D., WEST, J., MCCARTNEY, A. & RAMON, H. 2003. Early disease detection in wheat fields using spectral reflectance. *Biosystems Engineering*, 84, 137-145.
- BREIMAN, L. 2001. Random forests. *Machine learning*, 45, 5-32.
- BUBECK, D., GOODMAN, M., BEAVIS, W. & GRANT, D. 1993. Quantitative trait loci controlling resistance to gray leaf spot in maize. *Crop Science*, 33, 838-847.
- CHO, M. & SKIDMORE, A. 2009. Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy. *International Journal of Remote Sensing*, 30, 499-515.
- COCHRANE, M. 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, 21, 2075-2087.
- CRAWFORD, M. M., HAM, J., CHEN, Y. & GHOSH, J. Random forests of binary hierarchical classifiers for analysis of hyperspectral data. *Advances in Techniques for Analysis of Remotely Sensed Data*, 2003 IEEE Workshop on, 2003. IEEE, 337-345.
- CROUS, P. W., GROENEWALD, J. Z., GROENEWALD, M., CALDWELL, P., BRAUN, U. & HARRINGTON, T. C. 2006. Species of *Cercospora* associated with grey leaf spot of maize. *Studies in Mycology*, 55, 189-197.
- DE BIE, C. A. J. M. 2000. *Comparative Performance of Agro-Ecosystems*. PhD, Wageningen.
- DEGEFU, Y., LOHTANDER, K. & PAULIN, L. 2004. Expression patterns and phylogenetic analysis of two xylanase genes (htxyl1 and htlyl2) from *Helminthosporium turcicum*, the cause of northern leaf blight of maize. *Biochimie*, 86, 83-90.
- DEGRAEVE, S., MADEGE, R. R., AUDENAERT, K., KAMALA, A., ORTIZ, J., KIMANYA, M., TIISEKWA, B., DE MEULENAER, B. & HAESAERT, G. 2016. Impact of local pre-harvest management practices in maize on the occurrence

- of *Fusarium* species and associated mycotoxins in two agro-ecosystems in Tanzania. *Food Control*, 59, 225-233.
- DELALIEUX, S., VAN AARDT, J., KEULEMANS, W., SCHREVEENS, E. & COPPIN, P. 2007. Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. *European Journal of Agronomy*, 27, 130-143.
- DERERA, J., TONGOONA, P., PIXLEY, K. V., VIVEK, B., LAING, M. D. & VAN RIJ, N. C. 2008. Gene action controlling gray leaf spot resistance in Southern African maize germplasm. *Crop Science*, 48, 93-98.
- DEVRIES, J. & TOENNIESSEN, G. H. 2001. *Securing the harvest: biotechnology, breeding, and seed systems for African crops*, CABI.
- DUNKLE, L. D. & LEVY, M. 2000. Genetic relatedness of African and United States populations of *Cercospora zea-maydis*. *Phytopathology*, 90, 486-490.
- ELWINGER, G., JOHNSON, M., HILL, R. & AYERS, J. 1990. Inheritance of resistance to gray leaf spot of corn. *Crop Science*, 30, 350-358.
- ENVI 2009. *ENVI 4.7: environment for visualizing images*. Exelis Visual Information Solutions., ITT Industries, Colorado.
- FALKENBERG, N. R., PICCINNI, G., COTHREN, J. T., LESKOVAR, D. I. & RUSH, C. M. 2007. Remote sensing of biotic and abiotic stress for irrigation management of cotton. *Agricultural Water Management*, 87, 23-31.
- GEERTS, S., RAES, D., GARCIA, M., CASTILLO, C. D. & BUYTAERT, W. 2006. Agro-climatic suitability mapping for crop production in the Bolivian Altiplano: A case study for quinoa. *Agricultural and Forest Meteorology* 139, 399-412.
- GRAEFF, S., LINK, J. & CLAUPEIN, W. 2006. Identification of powdery mildew (*Erysiphe graminis* sp. *tritici*) and take-all disease (*Gaeumannomyces graminis* sp. *tritici*) in wheat (*Triticum aestivum* L.) by means of leaf reflectance measurements. *Central European Journal of Biology*, 1, 275-288.
- HATFIELD, P. & PINTER, P. 1993. Remote sensing for crop protection. *Crop Protection*, 12, 403-413.
- HILLNHÜTTER, C., MAHLEIN, A. K., SIKORA, R. A. & OERKE, E. C. 2011. Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields. *Field Crops Research*, 122, 70-77.
- JACKSON, R. D. 1986. Remote sensing of biotic and abiotic plant stress. *Annual review of phytopathology*, 24, 265-287.

- KAGODA, F., DERERA, J., TONGOONA, P., COYNE, D. L. & TALWANA, H. L. 2011. Grain yield and heterosis of maize hybrids under nematode infested and nematicide treated conditions. *Journal of nematology*, 43, 209.
- LYIMO, H. J. F., PRATT, R. C. & MNYUKU, R. S. O. W. 2012. Composted cattle and poultry manures provide excellent fertility and improved management of gray leaf spot in maize. *Field Crops Research*, 126, 97-103.
- MAHLEIN, A.-K., OERKE, E.-C., STEINER, U. & DEHNE, H.-W. 2012. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, 133, 197-209.
- MAHLEIN, A.-K., STEINER, U., DEHNE, H.-W. & OERKE, E.-C. 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture*, 11, 413-431.
- MANSOUR, K., MUTANGA, O., EVERSON, T. & ADAM, E. 2012. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS journal of photogrammetry and remote sensing*, 70, 56-65.
- MEISEL, B., KORSMAN, J., KLOPPERS, F. J. & BERGER, D. K. 2009. *Cercospora zeina* is the causal agent of grey leaf spot disease of maize in southern Africa. *European journal of plant pathology*, 124, 577-583.
- MILTON, E., SCHAEPMAN, M., ANDERSON, K., KNEUBÜHLER, M. & FOX, N. 2009. Progress in field spectroscopy. *Remote Sensing of Environment*, 113, S92-S109.
- MOSHOU, D., BRAVO, C., WEST, J., WAHLEN, S., MCCARTNEY, A. & RAMON, H. 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and electronics in agriculture*, 44, 173-188.
- MULLA, D. J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 358-371
- MUNKVOLD, G., MARTINSON, C., SHRIVER, J. & DIXON, P. 2001. Probabilities for profitable fungicide use against gray leaf spot in hybrid maize. *Phytopathology*, 91, 477-484.
- MURIITHI, L. & GATHAMA, S. 1998. Gray leaf spot of maize: a new disease on increase. *Crop Protection Newsletter*.

- MUTANGA, O., ADAM, E., ADJORLOLO, C. & ABDEL-RAHMAN, E. M. 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. *International Journal of Applied Earth Observation and Geoinformation*, 34, 178-187.
- MUTANGA, O. & SKIDMORE, A. 2004. Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment*, 90, 104-115.
- MUTANGA, O. & SKIDMORE, A. 2005. Discriminating tropical grass canopies grown under different nitrogen treatments using spectra resampled to HYMAP. *International Journal of Geoinformatics*, 1, 21-32.
- MUTANGA, O., SKIDMORE, A., KUMAR, L. & FERWERDA, J. 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, 26, 1093-1108.
- NILSSON, H.-E. 1995. Remote sensing and image analysis in plant pathology. *Canadian Journal of Plant Pathology*, 17, 154-166.
- NUTTER JR, F. W. & SCHULTZ, P. M. 1995. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Canadian Journal of Plant Pathology*, 17, 174-184.
- PAUL, P. & MUNKVOLD, G. 2005. Influence of temperature and relative humidity on sporulation of *Cercospora zeae-maydis* and expansion of gray leaf spot lesions on maize leaves. *Plant Disease*, 89, 624-630.
- PAUL, P. A. & MUNKVOLD, G. 2004. A model-based approach to preplanting risk assessment for gray leaf spot of maize. *Phytopathology*, 94, 1350-1357.
- PETROPOULOS, G. P., KALAITZIDIS, C. & PRASAD VADREVU, K. 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, 41, 99-107.
- PINGALI, P. & PANDEY, S. 2001. World maize needs meeting: technological opportunities and priorities for the public sector. En Pingali PL (Ed.) CIMMYT 1999–2000 World Maize Facts and Trends. Meeting World Maize Needs: Technological Opportunities and Priorities for the Public Sector. CIMMYT. Mexico, DF Part 1.

- PINSTRUP-ANDERSEN, P., PANDYA-LORCH, R. & ROSEGRANT, M. W. 1999. World food prospects: Critical issues for the early twenty-first century.
- QIN, Z. & ZHANG, M. 2005. Detection of rice sheath blight for in-season disease management using multispectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 7, 115-128.
- RUMPF, T., MAHLEIN, A.-K., STEINER, U., OERKE, E.-C., DEHNE, H.-W. & PLÜMER, L. 2010. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74, 91-99.
- SIBIYA, J., TONGOONA, P., DERERA, J., VAN RIJ, N. & MAKANDA, I. 2011. Combining ability analysis for *Phaeosphaeria* leaf spot resistance and grain yield in tropical advanced maize inbred lines. *Field crops research*, 120, 86-93.
- SLATON, M. R., HUNT, E. R. & SMITH, W. K. 2001. Estimating near-infrared leaf reflectance from leaf structural characteristics. *American Journal of Botany*, 88, 278-284.
- STEDDOM, K., BREDEHOEFT, M., KHAN, M. & RUSH, C. 2005. Comparison of visual and multispectral radiometric disease evaluations of *Cercospora* leaf spot of sugar beet. *Plant Disease*, 89, 153-158.
- VERIKAS, A., GELZINIS, A. & BACAUSKIENE, M. 2011. Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44, 330-349.
- WARD, J., LAING, M. & NOWELL, D. 1997a. Chemical control of maize grey leaf spot. *Crop protection*, 16, 265-271.
- WARD, J. M. J., LAING, M. D. & NOWELL, D. C. 1997b. Chemical control of maize grey leaf spot. *Crop Protection*, 16, 265-271.
- WARD, J. M. J., STROMBERG, E. L., NOWELL, D. C. & NUTTER JR, F. W. 1999. Gray leaf spot: a disease of global importance in maize production. *Plant disease*, 83, 884-895.
- WATKINS, K. & VON BRAUN, J. 2003. Time to stop dumping on the world's poor. *Trade Policies and Food Security*, 1-18.
- WEGARY, D., HABTAMU, Z., SINGH, H. & HUSIEN, T. 2003. Inheritance of grey leaf spot resistance in selected maize inbred lines. *Afr. Plant Prot*, 9, 53-54.
- WELZ, H. & GEIGER, H. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant breeding*, 119, 1-14.

- WEST, J. S., BRAVO, C., OBERTI, R., LEMAIRE, D., MOSHOU, D. & MCCARTNEY, H. A. 2003. The potential of optical canopy measurement for targeted control of field crop diseases. *Annual review of Phytopathology*, 41, 593-614.
- YANG, C.-M. 2010. Assessment of the severity of bacterial leaf blight in rice using canopy hyperspectral reflectance. *Precision Agriculture*, 11, 61-81.
- YANG, C. 2012. A high-resolution airborne four-camera imaging system for agricultural remote sensing. *Computers and Electronics in Agriculture*, 88, 13-24.
- YANG, Z., RAO, M., ELLIOTT, N., KINDLER, S. & POPHAM, T. 2005. Using ground-based multispectral radiometry to detect stress in wheat caused by greenbug (Homoptera: Aphididae) infestation. *Computers and electronics in agriculture*, 47, 121-135.
- ZHANG, M., QIN, Z., LIU, X. & USTIN, S. L. 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 4, 295-310.

Chapter 9

IDENTIFICATION, DETECTION, AND MAPPING OF MAIZE STREAK VIRUS AND GREY LEAF SPOT DISEASES:

A SYNTHESIS

9.0 Introduction

The rapid expansion in the geographical distribution of maize diseases namely; Maize Streak Virus disease (MSV) and Grey Leaf Spot (GLS) has caused a severe and drastic decrease in grain yield and quality in sub-Saharan Africa (Fajemisin, 1986, Flett et al., 1996, Payak and Renfro, 1974, Thottappilly et al., 1993, Ward et al., 1999). Although, the ripple effects of climate change, drought, and pests cannot be downplayed, the continued spread of aforementioned crop diseases in both commercial and small-holder maize farms has extensively contributed to the current scourge of food insecurity and malnutrition in countries where maize crops are one of the key staple foods (Anderson et al., 2004, Chakraborty and Newton, 2011, Schmidhuber and Tubiello, 2007, Cairns et al., 2012). For instance, in most sub-Saharan African countries maize is the main staple food, especially for the rural populace (Smale et al., 2013, Byerlee and Eicher, 1997, Nelson et al., 2009, Heisey and Mwangi, 1996). Previous studies have shown that maize production is already in the decline and the observed trends and patterns, demonstrate the likelihood of further food deficit in most African countries (Shiferaw et al., 2011, de Graaff et al., 2011, Pingali and Heisey, 2001). Crop diseases have been found to account for about 10% loss of the world's maize production (Jones and Thornton, 2003, Christou and Twyman, 2004). Despite the rate at which these diseases spread and affect maize crop production, yield losses associated with these diseases are not yet fully documented and this is the case for most small-scale farmers in African countries (Cairns et al., 2012, Chakraborty and Newton, 2011, Fajemisin, 1986). The history of these diseases in Africa dates back to the early 1990s and in some countries like Zimbabwe and even South Africa, they are now recognized as economically

important pathogens (Chakraborty and Newton, 2011, Fajemisin, 1986, Smale et al., 2013).

The diverse nature of farming approaches in these countries further complicates disease control mechanisms and efforts. In sub-Saharan African countries, for example, there are two distinct groups of farmers: large-scale maize producers and small-scale subsistence farmers (also referred to as small-holder farmers). These farmers have different coping strategies depending on resource availability. For example, large-scale maize farmers are usually supported by sophisticated market infrastructure, which always helps to stabilize their revenues. On the other hand, the majority of farmers, i.e. small scale-subsistence maize farmers, are more vulnerable to these disease epidemics as they have limited knowledge, with no land rights, as they operate under a communal land tenure system. Therefore, for these farmers, the impacts of these diseases on their crops is devastating as they have no capital to access sophisticated disease management options. In addition, small scale farmers, a farm in small areas, less 10ha, making the use of pesticides or fungicides costly. Although some countries have intervention mechanisms in place, most of them are not as efficient as the spatial distribution and information on the areas affected by these diseases remain rudimentary. This is further exacerbated by the lack of up-to-date farm specific spatial information, pathological expertise, and resources, especially in developing countries, as well as the patchy-nature of disease occurrence. As a result, of these challenges, most countries have been slow in curbing and effectively controlling these diseases, and the possible threats to maize production.

To overcome the challenges and to ensure effective control of these diseases, as well as ensure food security, there is a need to embrace new remote sensing technologies with a global-foot print, fine spatial resolution and a repeated coverage which can be used to identify, detect and map lethal maize diseases, such as Maize Streak Virus disease and Grey Leaf Spot. Information derived will help in devising sustainable and robust control measures and mitigation measures at low costs. To help address the above challenges, this study sought to identify, detect, and map the spatial distribution of two maize diseases: maize streak virus and Grey leaf spot using remotely sensed data

Specifically, the objectives of this work were to:

- compare Random forest's forward variable selection and guided regularized random forest methods for optimum variable selection;
- detect the severity of maize streak virus infestations in maize hybrid lines using in-situ hyperspectral data;
- detect and map maize streak virus using RapidEye satellite imagery;
- investigate the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations
- to determine environmental factors that explain the probability of maize streak virus disease occurrence; and
- Test the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot infection in maize hybrid line.

9.1 To compare Random forest's forward variable selection and guided regularized random forest methods for optimum variable selection

The recent advancement in remote sensing technologies has resulted in the introduction of new sensors with multiple or hyper bands. Hence the use of such datasets in disease monitoring requires intensive filtering or screening of important spectral variables that can optimally detect and map maize diseases. For instance, Hyperspectral data has numerous spectral bands ranging between 350 nm and 2500 nm, with bandwidths of less than 2 nm (Demarchi et al., 2014). The use of a large number of variables thus compromises accuracy besides being laborious and enhances modelling complexity (Hsu, 2007, Pal, 2009). Therefore, methods that can reduce data dimensionality without discarding important information are critical especially when dealing with hyperspectral data. So far, there are two categories of variable reduction methods namely; feature selection and feature extraction. Feature selection methods reduce the dimensionality by selecting a subset of features capturing the relevant properties of the entire data set whereas, the feature extraction approach provides new features based on a linear or nonlinear transformation of the original feature sets (Pal, 2009). Thus, before mapping the spatial distribution of maize streak virus in small-holder farms at Ofcolaco Farms, in Tzaneen, South Africa, the strengths of two unique methods i.e. guided regularized random forest and traditional random forest's variable selection methods were tested to select a subset of resampled Hymap and AISA Eagle variables that can accurately classify four stages of MSV infection.

The results of this work showed that 502.4, 636.3, 669.7, 683, 729.8 and 850.4 bands derived from the resampled AISA Eagle were selected as the best using GRRF yielded. For example, based on these bands in classifying maize streak virus, a lower OOB error of 8.4 % was obtained. On the other hand, from the resampled HyMap dataset, 480.4, 571.9, 633.5 and 708.9 were selected as the optimal bands with an OOB error of 7.25%. Following this process, the selected best subsets of bands were then used as input variables in random forest classifier to map MSV disease. The results obtained demonstrated that the four resampled Hymap selected bands yielded an overall accuracy of 91.67% and a Kappa value of 0.89 whereas the

six bands selected from resampled AISA, using the GRRF produced a slightly lower accuracy of 89.17% and a Kappa value of 0.86. The results have shown that the GRRF algorithm has the potential to select compact feature sub sets, and the accuracy performance is better than that of RF's variable selection method.

9.2 Detect the severity of maize streak virus infestations in maize hybrid lines using in-situ hyperspectral data

Continuous monitoring of maize crop health status is critical for developing early warning systems, effective control mechanisms, and appropriate modelling of crop yield and associated projections. Also, determining farm patches affected with diseases, such as MSV helps to avoid indiscriminate and costly applications of agrochemicals throughout the entire field. Crop information on the spatial extent and real-time distribution and disease-related damages have largely depended on visual surveys, which are often costly and untimely. Satellite data, both in-situ and airborne therefore hypothesized to provide spatially explicit information about the damage, and real-time spatial distribution of disease infestation over small and large scale areas (Raji et al., 2016). Although hyperspectral data has been extensively applied, its application in crop yield and other related studies, its performance in disease monitoring remains unclear. This study for the first time sought to test whether field spectrometry measurements could accurately discriminate between various stages of MSV maize infection (healthy, early, moderate, and severely), using the GRRF machine learning algorithm. The results have shown that various stages of MSV infection (healthy, early, moderate, and severely) can accurately be discriminated based on hyperspectral bands positioned at 552 nm, 603 nm, 683 nm, 881 nm, and 2338 nm with an overall accuracy of 95.83 % and a Kappa of 0.94. The results underscored the relevance of the GRRF algorithm in determining the optimal wavelengths that can detect and discriminate different levels of MSV crop infections at the plot level.

9.3 Detect and map maize streak virus using RapidEye satellite imagery

The need to develop robust dataset and methodologies that can quickly identify and map croplands infested with diseases, such as MSV has remained one of the major research endeavors by scientists to help curb crop loss in both commercial and non-commercial farms. Although some progress has been made in the developed countries, in sub-Saharan Africa and other parts of the developing world the spread of diseases has continued to spread unabated, due to the lack of financial, and technical expertise, as well as climate change effects. Poor financial resources have hampered the acquisition of high-resolution data i.e. LIDAR data that could map crop diseases at plant or plot level. However, although still expensive the provision of RapidEye at 5m resolution can aid by providing insights on the rate of crop disease infestation and potentially vulnerable areas at the farm level. The aim of this work was to, therefore, detect and map maize streak virus using the 5m spatial resolution RapidEye multispectral sensor at farm level in Ofcolaco, Tzaneen, South Africa. It is hypothesized that the findings from this work would greatly help both commercial and non-commercial farmers, agricultural departments and related research institutes in embracing new techniques to identify, map and quantify affected croplands at the farm level. This has remained a challenge, especially from the use of hyperspectral data, although considered to be optimal for plot level map. Hyperspectral is not recommended for farm level mapping due to the lack of a spatial representation. The findings of the study have shown that raw spectral data derived from the 5m spatial resolution RapidEye data can assist in mapping MSV diseases with reasonably high accuracy (OA = 82.75%). The results further improved by 3.4% using advanced RapidEye derivatives (vegetation indices). Two visible spectral bands (blue and red-edge), NIR, the SAVI, EVI, RI, and NDVI were identified as central in disease monitoring.

9.4 Investigating the potential of the Landsat-8 data in detecting and mapping maize streak virus infestations

Notwithstanding the plausible application and performance of both in-situ hyperspectral data and the resampled hyperspectral data to high spatial and spectral resolution sensors (RapidEye, Sentinel 2, and Worldview-2) in detecting MSV and GLS maize diseases, their application remains a daunting task. For example, Worldview 2, and RapidEye are commercial sensors, their accessibility and utility comes at a cost. Moreover, although they have an appealing spatial resolution $\pm 2\text{m}$, with strategically-position bands (Red-Edge, Yellow wavelengths), they cannot easily be used in crop disease monitoring, over time, as they do not have archival data when compared to the Landsat series data, which dates back to the 1970s. Besides, the two sensors cannot be used in operational studies, due to high acquisition costs, small area coverage, multi-collinearity, and limited availability. These challenges hinder their successful use in regional scale crop diseases monitoring. Because of these limitations with the “best” performing sensors the study further examined the capability of the newly launched medium-resolution Landsat-8, with more appealing spatial, spectral, radiometric and temporal characteristics, in detecting and mapping maize streak virus in farms at Ofcolaco. The findings of the study showed that the use of Landsat 8 spectral bands in the detection and mapping of maize streak virus yielded classification results with an overall accuracy of 50%. The inclusion of the vegetation indices, in the classification process slightly improved classification accuracies by a magnitude of 1.29%. The study shows promising findings, hence the need to test such data sets together with other high-resolution multispectral information in mapping complex crop diseases, such as the grey leaf spot.

9.5 Spatial modelling of maize streak virus disease using environmental variables

Environmental variables that can influence the presence of MSV disease at a local scale were examined. Climatic, topographical variables together with Landsat 8 derived vegetation indices were integrated to predict the probability of occurrence of MSV infestation in Ofcolaco maize fields. Correlation analysis and logistic regression to relate vegetation indices, climatic and topographical variables to incidences of maize streak virus disease were used in this study. The results indicate the potential of integrating Landsat 8 derived vegetation indices and environmental variables to improve the prediction of MSV disease outbreaks in maize fields. The result is critical for maize crop health monitoring using remote sensing data and environmental variables.

9.6 Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot infection in maize hybrid line.

Although the field spectroscopy is renowned for its utility and robustness in plant health monitoring, its application remains unsatisfactory as it is largely restricted to plot level mapping. Given that most countries are becoming more and more food insecure despite increasing population, regional scale or “wall-to-wall” mapping and monitoring of crop disease is urgently required especially for developing countries like South Africa where maize crop remains a staple food, as well as the cash crop. Thus, after realizing the limitations associated with hyperspectral data, both technical and pre-processing requirements in detecting maize streak virus, I was prompted to test the potential of new generation sensors namely; WorldView-2, Quickbird, RapidEye, and Sentinel-2 resampled from hyperspectral data in mapping the most complex maize disease; Grey Leaf Spot (GLS) in maize fields. Comparatively the results showed that Sentinel-2 image can outperform both Worldview 2 and RapidEye images in detecting and mapping GLS disease in maize crops. For example, Sentinel 2 data resampled from hyperspectral data yielded the highest

overall accuracy of 84% and a kappa value of 0.76, followed by WorldView-2, which yielded slightly lower accuracies with 82% overall accuracy and 0.73 Kappa. On the other hand, classification results obtained using the RapidEye spectral bands yielded 82.75%, with 3.4% improvement based on the use of vegetation indices.

9.7 Conclusions and future perspectives

The main aim of this study was to identify, detect, and map the spatial distribution of two maize diseases: maize streak virus and grey leaf spot using remotely sensed data. The results have shown that the integration of different remote sensing and machine learning algorithms, can help in detecting and mapping maize diseases (MSV and GLS) at both farm and regional scales. This information is also helpful in determining farm portions affected and in making decisions regarding site-specific applications of chemicals and fungicides and thus contribute to the concept of precision farming. Overall, the results imply that opportunities exist for developing operational remote sensing systems for detection of maize diseases. Adoption of such remote sensing techniques is particularly valuable for minimizing crop damage, improving yield and ensuring food security. Despite the moderate performance of Landsat 8 in mapping maize diseases, when compared to its counterparts, hyperspectral data and hyperspectral data resampled to RapidEye, Sentinel 2 and Worldview 2. Its performance presents a unique potential for future related studies, especially at national or regional scale. There is a need for future studies to further investigate the use of Landsat 8 and Sentinel 2 in disease mapping and monitoring at national and even regional scales. Such findings can help governments, farmers and policy and decision makers to ultimately improve maize crop disease mitigation strategies and yield loss, so as to enhance maize production by minimizing losses. While the current study has proven that maize diseases, such as MSV and GLS can accurately be detected and mapped using satellite data, however, for precise diseases control and mitigation, there is a need for future studies to further conduct temporal analysis on their spread along phenological growth stages (young, intermediate and mature). This information is critical because if such diseases can be detected at any early stage, they can be controlled before too much damage has occurred as compared to when they have matured.

9.8 References

- ANDERSON, P. K., CUNNINGHAM, A. A., PATEL, N. G., MORALES, F. J., EPSTEIN, P. R. & DASZAK, P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change, and agrotechnology drivers. *Trends in Ecology & Evolution*, 19, 535-544.
- BYERLEE, D. & EICHER, C. K. 1997. *Africa's emerging maize revolution*, Lynne Rienner Publishers.
- CAIRNS, J. E., SONDER, K., ZAIDI, P., VERHULST, N., MAHUKU, G., BABU, R., NAIR, S., DAS, B., GOVAERTS, B. & VINAYAN, M. 2012. 1 Maize Production in a Changing Climate: Impacts, Adaptation, and Mitigation Strategies. *Advances in agronomy*, 114, 1.
- CHAKRABORTY, S. & NEWTON, A. C. 2011. Climate change, plant diseases, and food security: an overview. *Plant Pathology*, 60, 2-14.
- CHRISTOU, P. & TWYMAN, R. M. 2004. The potential of genetically enhanced plants to address food insecurity. *Nutrition research reviews*, 17, 23-42.
- DE GRAAFF, J., KESSLER, A. & NIBBERING, J. W. 2011. Agriculture and food security in selected countries in Sub-Saharan Africa: diversity in trends and opportunities. *Food Security*, 3, 195-213.
- DEMARCHI, L., CANTERS, F., CARIOU, C., LICCIARDI, G. & CHAN, J. C.-W. 2014. Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 166-179.
- FAJEMISIN, J. Maize diseases in Africa and their role in the varietal improvement process. 1. Eastern, Central, and Southern Africa Regional Maize Workshop, Lusaka (Zambia), 10-17 Mar 1985, 1986. CIMMYT.
- FLETT, B., BENSCH, M., SMIT, E. & FOURIE, H. 1996. A field guide for identification of maize diseases in South Africa.
- HEISEY, P. W. & MWANGI, W. 1996. Fertilizer use and maize production in sub-Saharan Africa.
- HSU, P. 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 78-92.
- JONES, P. G. & THORNTON, P. K. 2003. The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global environmental change*, 13, 51-59.
- NELSON, G. C., ROSEGRANT, M. W., KOO, J., ROBERTSON, R., SULSER, T., ZHU, T., RINGLER, C., MSANGI, S., PALAZZO, A. & BATKA, M. 2009. *Climate change: Impact on agriculture and costs of adaptation*, Intl Food Policy Res Inst.
- PAL, M. 2009. Margin-based feature selection for hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, 11, 212-220.
- PAYAK, M. & RENFRO, B. 1974. A decade of research on maize diseases-impact on production and its international cooperative outreach.
- PINGALI, P. L. & HEISEY, P. W. 2001. Cereal-crop productivity in developing countries: past trends and future prospects. *Agricultural science policy: Changing global agendas*.

- RAJI, S. N., SUBHASH, N., RAVI, V., SARAVANAN, R., MOHANAN, C. N., MAKESHKUMAR, T. & NITA, S. 2016. Detection and Classification of Mosaic Virus Disease in Cassava Plants by Proximal Sensing of Photochemical Reflectance Index. *Journal of the Indian Society of Remote Sensing*, 44, 875-883.
- SCHMIDHUBER, J. & TUBIELLO, F. N. 2007. Global food security under climate change. *Proceedings of the National Academy of Sciences*, 104, 19703-19708.
- SHIFERAW, B., PRASANNA, B. M., HELLIN, J. & BÄNZIGER, M. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, 3, 307.
- SMALE, M., BYERLEE, D. & JAYNE, T. 2013. Maize revolutions in sub-Saharan Africa. *an African green revolution*. Springer.
- THOTTAPPILLY, G., BOSQUE-PÉREZ, N. & ROSSEL, H. 1993. Viruses and virus diseases of maize in tropical Africa. *Plant Pathology*, 42, 494-509.
- WARD, J. M., STROMBERG, E. L., NOWELL, D. C. & NUTTER JR, F. W. 1999. Gray leaf spot: a disease of global importance in maize production. *Plant disease*, 83, 884-895.