

---

**MODELLING SPATIAL VARIABILITY OF COFFEE (*COFFEA ARABICA L.*)  
CROP CONDITION WITH MULTISPECTRAL REMOTE SENSING DATA**

Chemura Abel



A thesis submitted to the College of Agriculture, Engineering and Science, at the University of  
KwaZulu-Natal, in fulfilment of the academic requirements for the degree of Doctor of  
Philosophy in Environmental Science

Academic Supervisor: Professor Onesimo Mutanga

Academic Advisors: Dr. John Odindi

Dr. Timothy Dube

Pietermaritzburg,

South Africa

November 2017

---

## ABSTRACT

Coffee (*Coffea arabica* L.) is an important economic crop produced by over 25 million farmers, across four continents. Coffee production is not just important for livelihoods of farmers, but also provide ecosystem services. As a perennial crop, coffee requires a robust, reliable and cost-effective monitoring strategy for diseases, pests, water stress, soil fertility and other crop stressors, to ensure long-term productivity and to safeguard investments and other related ecosystem services. The current monitoring methods largely rely on spontaneous field inspections and sampling, which are not only labour intensive, but also conclusive once economic damage has been inflicted on the crop. The coffee crop, due to its architecture, production system, planting arrangement and cropping cycle presents many challenges that make traditional remote sensing not applicable. Vegetation monitoring data and algorithms developed for either natural ecosystems or annual agricultural crops cannot be applied to coffee production because its distinct characteristics. In addition, since the majority of the coffee production is in developing countries, easily available multispectral level approaches are required for crop condition assessments of a unique and challenging but yet economically and ecologically important cropping system.

This study presents an integrated system for crop condition assessment using available multispectral remote sensing data for landscape scale modelling and mapping of crop health in coffee plantations. First, as a perennial crop, coffee is planted in various age groups across the landscape for continuity. However, age of the coffee plants significantly affects remote sensing potential due to its relationship with canopy cover and soil background effect. An approach to develop reliable and detailed age-specific thematic maps for areas producing coffee was developed by combining the machine learning robustness of the random forests algorithm with the improved sensor design of Landsat 8 OLI data. Results showed that higher overall classification accuracy was achieved when coffee was classified as a single class (90.3% for OLI and 86.8% for ETM+) than the three age-based coffee classes (86.2% for OLI and 81.0% for ETM+). It was concluded that disaggregating coffee classes to produce age-specific maps reduce overall accuracy, but the usefulness of a thematic map with age-specific classes is more than the value of the marginal decrease in accuracy. Thus, this research achieved inter and intra-class discrimination of landcover classes in a heterogenous coffee producing area. A reliable coffee age-mask was produced, thereby solving the problem of age variability in coffee plantations, which is one of the key limitations to crop condition assessments with remote sensing in perennial crops.

Next, the age specific thematic maps were used as age masks to develop an NDVI based method to detect and quantify incongruous patches based on their deviation from the age-expected mean values. Incongruous patches represent areas experiencing poor crop growth, pest infestation, disease outbreaks, soil fertility problems and/or responding or not responding to management interventions. Results showed that using the age-adjusted anomalies performed better in separating incongruous and healthy patches than using the global mean for both normalized difference vegetation index (NDVI) (Overall accuracy=77.2% and 66.4% respectively) and for Land Surface Water Index (LSWI) (Overall accuracy=66% and 49.2% respectively). When applied to other Landsat 8 OLI scenes, the results showed that the proportions of coffee fields that were modelled incongruent decreased with time for the young age category while it increased for the mature and

old age classes with time. Using the distribution of transformed age-adjusted NDVI values around known health values with the inverse cumulative distribution function (ICDF) is an innovative way of detecting and mapping anomalous areas in plantations both spatially and temporally.

Having identified anomalous areas, more specific remote sensing approaches could now be used to determine the cause of the identified anomalies in the coffee fields. To achieve this, the effect of plant water stress and coffee leaf rust (CLR), a major coffee disease was explored as potential abiotic and biotic stressors that influence the spectral response of coffee at field scale. In both cases the random forest algorithm successfully classified and quantitatively predicted stress levels in coffee leaves. In plant water stress modelling, the results showed that the bands selected through reflectance sensitivity performed best in water stress detection ( $r = 0.87$ ,  $RMSE = 4.91\%$  and  $pBias = 0.9\%$ ). In CLR discrimination, the results showed that simulated Sentinel-2 MSI derived vegetation indices achieved relatively high overall accuracy of 76.2% when compared to 69.8% obtained using raw spectral bands. The RBF-PLS model satisfactorily modelled CLR severity ( $R^2=0.92$  and  $RMSE=6.1\%$ ) using all simulated Sentinel-2 MSI bands. These results indicate that it is possible to reliably predict plant water content (PWC) and CLR using wavebands in the VIS/NIR range that correspond with many of the available multispectral remote sensing data. This is an important development as previous studies mainly used hyperspectral data and data beyond the VIS/NIR range that is not available in many of the commonly accessible multispectral sensors. However, further research at field and landscape scale is required to operationalize these findings as problems are encountered in transferring models from leaf level to canopy level.

The successful isolation of abiotic and biotic stressors that influence spectral signatures of coffee plantations set the basis to use multispectral remote sensing data to model the spatial variability of chlorophyll and nitrogen with Sentinel 2 MSI data. Results showed that the best modelling results ( $R^2=0.77$ ,  $RMSE=5.9$ ) were achieved when all the bands at 10m spatial resolution were used in modelling coffee leaf chlorophyll for mature coffee stands. It was also concluded that the interaction between spectral settings and spatial resolution are important in chlorophyll estimation with Sentinel-2 MSI data. These findings are imperative in that Sentinel 2 MSI data comes with variable spatial and spectral resolution. Therefore, our finding that the spatial resolution, together with the spectral settings, cast aspersions on spectroscopic simulation studies that did not consider spatial resolution of Sentinel 2 MSI data in biochemical and biophysical vegetation characterisation. A combination of optimized bands and vegetation indices produced the best results for coffee foliar nitrogen modelling ( $R^2=0.78$ ,  $RMSE=0.23$ ) at a landscape scale. The obtained nitrogen distribution maps can be used in precision coffee plantation management, insurance assessment and yield forecasting which were previously daunting tasks. This study therefore extended the frontiers of knowledge by developing an integrated multi-sensor approach for landscape scale modelling of coffee plant health. The use of multispectral remotely sensed data that is reliable and available at little to no cost as an added advantage to coffee producing countries in developing countries particularly in Sub-Saharan Africa.

## **PREFACE**

The research work described in this thesis was carried out in the School of Agricultural, Earth and Environmental Sciences (SAES), University of KwaZulu-Natal, Pietermaritzburg, from January 2015 to August 2017, under the supervision of Prof. Onesimo Mutanga (School of Agricultural, Earth and Environmental Sciences, University of KwaZulu Natal, South Africa).

I would like to declare that the research work reported in this thesis has never been submitted in any form to any other university. It therefore represents my original work, except where due acknowledgments are made.

Abel Chemura Signed: \_\_\_\_\_ Date: \_\_\_\_\_

As the candidate's supervisor, I certify the above statement and have approved this thesis for submission.

Prof. Onesimo Mutanga Signed: \_\_\_\_\_ Date: \_\_\_\_\_



## DECLARATION 1 - PLAGIARISM

I, **Abel Chemura**, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research,
2. This thesis has not been submitted for any degree or examination at any other university,
3. This thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons,
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - a. their words have been re-written, but the general information attributed to them has been referenced;
  - b. where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced;
5. This thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged and the source being detailed in the thesis and in the References section.

Signed: \_\_\_\_\_

Date: \_\_\_\_\_

## DECLARATION 2 - PUBLICATIONS AND MANUSCRIPTS

1. **Chemura, A.** and Mutanga, O. (2016) Developing detailed age-specific thematic maps for coffee (*Coffea arabica* L) for heterogeneous agricultural landscapes using random forests applied on Landsat 8 multispectral sensor data, *Geocarto International*, 32(7), 759-776.
2. **Chemura, A.**, Mutanga, O and 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.
3. **Chemura, A.**, Mutanga, O and Dube, T (In Press) Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests, *Physics and Chemistry of the Earth ABC*.
4. **Chemura, A.**, Mutanga, O and Dube, T (2016) Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions, *Precision Agriculture*, 17, 1-23
5. **Chemura, A.**, Mutanga, O., Sibanda, M., and Chidoko, P. (In Press) Machine learning prediction of coffee rust severity on leaves using spectroradiometer data, *Tropical Plant Pathology*.
6. **Chemura, A.**, Mutanga, O., and Odindi, J. (In Press) Empirical modelling of leaf chlorophyll content in coffee (*Coffea arabica*) plantations with Sentinel-2 MSI data: Effects of spectral settings, spatial resolution and crop canopy cover. Accepted in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
7. **Chemura, A.**, Mutanga, O., Kutwayo, D., and Odindi, J. (Accepted) Mapping spatial variability in foliar nitrogen in coffee (*Coffea arabica*) plantations with multispectral level Sentinel-2 MSI data, Submitted to *ISPRS Journal of Photogrammetry and Remote Sensing* (June 2017).

Signed: \_\_\_\_\_

## **DEDICATION**

This thesis is dedicated to my mother Mrs Bongai Chemura and my love Sitshengy.

## **ACKNOWLEDGEMENTS**

This study could not have been successful if it were not for the assistance, contributions and cooperation I got from organisations and individuals for which I am very grateful. I give sincere gratitude to my supervisor Prof Onesimo Mutanga whose guidance, patience and wit helped me to properly shape this research work. In addition, I am grateful to Dr John Odindi and Dr Timothy Dube for their support throughout my research project. The input and support of Dr D. Kutwayo, Mr P. Chidoko and Mr C. Mahoya is also greatly appreciated in facilitating field work and providing technical support. I also wish to thank my colleagues Dr. Terence Mushore, Dr. Mbulisi Sibanda, Victor M Bangamwabo, Dr Khoboso Seutloadi, Cletah Shoko, Thembeke Mhlongo and Pedzisai Kowe for their collaboration and moral support during the journey of putting together this study. I also want to thank my wife Sitshengisiwe, for supporting me, purifying my research ideas and proof-reading all my writings. The study also ran concurrently with International Foundation of Science (IFS) grant (D/5441) that provided financial support for carrying out some aspects of this study. I am also thankful to Mrs Shanita Ramroop, Sibongile Ntuli, Dr Mercy Ojoyi and Dr Concilia Danha for their support and encouragement. This research project and the entire study also drew a lot of inspiration from Anninlla-Nyasha for whom I will always work hard for.

## TABLE OF CONTENTS

ABSTRACT.....	i
PREFACE.....	iii
DECLARATION 1 - PLAGIARISM .....	iv
DECLARATION 2 - PUBLICATIONS AND MANUSCRIPTS .....	v
DEDICATION .....	vi
ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS.....	viii
LIST OF FIGURES .....	xv
LIST OF TABLES.....	xviii
SECTION 1: INTRODUCTION AND OVERVIEW.....	1
CHAPTER 1: GENERAL INTRODUCTION .....	2
1.1 Coffee and coffee production.....	3
1.2 Remote sensing in agronomic decision making.....	4
1.3 Challenges of remote sensing application in coffee production .....	5
1.4 Importance of multispectral remote sensing in coffee production.....	6
1.5 Aim and objectives .....	7
1.5.1 Aim .....	7
1.5.2 Objectives .....	7
1.6 Description of the study area .....	8
1.6.1 Coffee Research Institute.....	8
1.6.2: Ward 19 Chipinge District.....	8
1.7 Thesis structure .....	9
1.7.1 Section 1: Introduction and Overview .....	10
1.7.1.1 Chapter One: Introduction .....	10
1.7.1.2 Chapter Two: Literature review .....	10
1.7.2 Section 2: Setting the stage for coffee condition assessment.....	10
1.7.2.1 Chapter Three: Development of an age mask for coffee condition assessment.....	10
1.7.2.2 Chapter Four: Identification of anomalous patches in coffee plantations.....	11
1.7.3 Section 3: Remote sensing individual coffee stressors .....	11
1.7.3.1 Chapter Five: Predicting plant water content in coffee plants using multispectral remote sensing.....	11
1.7.3. 2 Chapter Six: Modelling coffee leaf rust using multispectral remote sensing.....	11
1.7.4 Section 4: Empirical modelling of coffee condition using multispectral data .....	11

1.7.4.1 Chapter Seven: Multispectral remote sensing of coffee chlorophyll content .....	11
1.7.4.2 Chapter Eight: Multispectral remote sensing of spatial variability of coffee foliar nitrogen..	12
1.7.4.5 Chapter Nine: Synthesis.....	12
<b>CHAPTER 2: ADVANCES IN SENSOR APPLICATIONS FOR COFFEE (COFFEA ARABICA L.)</b>	
<b>CROP CONDITION ASSESSMENTS AND MONITORING .....</b>	<b>13</b>
Abstract.....	14
2.1 Introduction.....	15
2.2 Physiological characteristics of coffee and their potential influence on remote sensing .....	17
2.3 Major biotic and abiotic stressors of coffee arabica plants that require monitoring .....	19
2.3.1 Nutrient deficiencies .....	19
2.3.2 Plant water stress and drought .....	20
2.3.3 Coffee Leaf Rust (CLR).....	21
2.3.4 Coffee Berry Disease (CBD) .....	21
2.3.5 Coffee Wilt Disease (CWD) .....	22
2.3.6 Coffee White Stem Borer (CWB).....	23
2.3.7 Other coffee diseases and pests.....	23
2.4. Current condition assessment methods in coffee crop management.....	24
2.5. Applications of remote sensing in coffee crop management .....	25
2.5.1 Mapping cropped areas .....	25
2.5.2 Detection of coffee physiological properties .....	26
2.5.3 Coffee condition assessment.....	28
2.5.4 Yield estimation and modelling .....	29
2.6 Challenges for remote sensing in coffee production and future perspectives.....	32
2.7 Conclusion .....	34
<b>SECTION 2: SETTING THE STAGE FOR COFFEE CONDITION ASSESSMENT .....</b>	<b>35</b>
<b>CHAPTER 3: DEVELOPMENT OF AN AGE MASK FOR COFFEE CONDITION ASSESSMENT ..</b>	<b>36</b>
Abstract.....	37
3.1 Introduction.....	38
3.2 Material and methods.....	40
3.2.1 Study area.....	40
3.2.2 Field data collection.....	41
3.2.3 Image data and pre-processing.....	42
3.2.4 Spectral analysis and characterization .....	43
3.2.5 Classification schemes .....	44

3.2.6 Random forest classification .....	44
3.2.7 Accuracy assessment.....	45
3.2.8 Comparing performance of classifiers .....	47
3.4 Results.....	48
3.4.1 Spectral characteristics of coffee age groups .....	48
3.4.2 Assessment of spectral separability of coffee classes .....	48
3.4.3 Assessing importance of individual bands in RF classification.....	49
3.4.4 Classification accuracy assessment.....	50
3.4.5 Comparing class area between the classifiers and schemes .....	55
3.5 Discussion.....	57
3.5.1 Comparison of performance of Landsat 8 OLI and ETM+.....	57
3.5.2 Effect of age on spectral reflectance of coffee.....	58
3.5.3 Comparison of accuracy performance .....	59
3.6 Conclusions.....	60
3.7: Link to next chapter .....	60
<b>CHAPTER 4: IDENTIFICATION AND MAPPING OF ANOMALOUS PATCHES IN COFFEE PLANTATIONS WITH MULTISPECTRAL DATA .....</b>	<b>61</b>
Abstract.....	62
4.1 Introduction.....	63
4.2 Materials and methods .....	67
4.2.1 Study area.....	67
4.2.2 Satellite imagery .....	68
4.2.2.1 Data acquisition.....	68
4.2.2.2 Data pre-processing.....	69
4.2.2.3 Derivation of vegetation indices .....	70
4.2.3 Field data and statistical analysis .....	70
4.2.4 Anomaly detection and mapping .....	71
4.2.5 Model evaluation.....	72
4.3 Results.....	73
4.3.1 Evaluating influence of age classes on vegetation indices.....	73
4.3.2 Determining anomalies .....	74
4.3.3 Accuracy assessment.....	78
4.3.4 Quantification of incongruous areas .....	79
4.4 Discussion.....	80

4.4.1 Effect of age on NDVI and LSWI values .....	80
4.4.2 Remote sensing-based identification of incongruent patches .....	81
4.4.3 Potential limitations and future improvements .....	83
4.5 Conclusions.....	84
4.6 Link to next chapter .....	84
<b>SECTION 3: REMOTE SENSING INDIVIDUAL COFFEE PLANT STRESSORS .....</b>	<b>85</b>
<b>CHAPTER 5: MULTISPECTRAL LEVEL REMOTE SENSING OF PLANT WATER CONTENT IN COFFEE.....</b>	<b>86</b>
Abstract.....	87
5.1 Introduction.....	88
5.2 Materials and methods .....	90
5.2.1 Study area.....	90
5.2.2 Experimental materials and design .....	90
5.2.3 Moisture stress treatments.....	91
5.2.4 Spectral reflectance measurements .....	91
5.2.5 Leaf water content measurements.....	91
5.2.6 Identification of water stress related wavebands .....	92
5.2.6.1 Cross-correlation threshold .....	92
5.2.6.2 Reflectance difference.....	93
5.2.6.3 Reflectance sensitivity .....	93
5.2.7 Modelling approach .....	93
5.2.8 Model evaluation.....	94
5.3. Results.....	95
5.3.1 Coffee plant water content and reflectance.....	95
5.3.2 Variable selection.....	95
5.3.2.1 Cross-correlation threshold.....	95
5.3.2.2 Reflectance difference and reflectance sensitivity .....	96
5.3.3 Variable importance.....	97
5.3.4 Model performance evaluation .....	97
5.3.4.1 Cross-correlation threshold.....	97
5.3.4.2 Reflectance difference.....	98
5.3.4.3 Reflectance sensitivity .....	98
5.4 Discussion.....	99
5.4.1 Comparison of variable selection methods .....	99



5.5 Conclusions.....	101
5.6 Link to next chapter .....	101
<b>CHAPTER 6: MULTISPECTRAL LEVEL REMOTE SENSING-BASED DISCRIMINATION AND SEVERITY MODELLING OF COFFEE LEAF RUST .....</b>	<b>102</b>
Abstract.....	103
6.1 Introduction.....	104
6.2 Materials and methods .....	107
6.2.1 Study area.....	107
6.2.2 Coffee leaf rust inoculation.....	107
6.2.3 Reflectance measurements and resampling .....	108
6.2.4 Vegetation Indices.....	109
6.2.5 Statistical approaches for CLR discrimination .....	111
6.2.5.1 Random forest algorithm. ....	111
6.2.5.2 Partial least squares discriminant analysis .....	111
6.2.6 Statistical approaches for severity modelling .....	111
6.2.7 Accuracy assessment.....	112
6.2.7.1 CLR discrimination.....	112
6.2.7.2 Severity modelling .....	112
6.3 Results.....	112
6.3.1 Spectral Resampling .....	112
6.3.2 CLR discrimination.....	113
6.3.2.1 Discrimination with all variables .....	113
6.3.2.2 Variable optimization.....	115
6.3.2.3 CLR discrimination with optimized variables .....	116
6.3.3 CLR severity modelling.....	118
6.3.3.1 Relationship between Sentinel-2 MSI spectral bands and CLR severity.....	118
6.3.3.2 Determining gaussian widths .....	119
6.3.3.3 Modelling CLR severity.....	120
6.4 Discussion .....	121
6.4.1 Spectral resampling.....	122
6.4.2 Discrimination.....	122
6.4.2.1 Performance of Sentinel-2 spectral bands in CLR discrimination.....	122
6.4.2.2 Performance of Sentinel-2 vegetation indices in CLR discrimination.....	123
6.4.2.3 Effect of variable optimization and algorithm on CLR discrimination .....	124

6.4.3 Severity modelling .....	124
6.4.3.1 Relationship between Sentinel-2 MSI spectral bands and CLR severity .....	124
6.4.3.2 Modelling CLR severity with spectral bands .....	125
6.4.4 Multispectral level remote sensing of biotic stress in coffee .....	126
6.5 Conclusion .....	126
6.6 Link to next chapter .....	127
<b>SECTION 4: EMPIRICAL MODELLING OF COFFEE CONDITION USING MULTISPECTRAL REMOTE SENSING DATA .....</b>	<b>128</b>
<b>CHAPTER 7: LANDSCAPE SCALE MULTISPECTRAL REMOTE SENSING OF COFFEE TOTAL CHLOROPHYLL CONTENT .....</b>	<b>129</b>
Abstract .....	130
7.1 Introduction .....	131
7.2 Materials and methods .....	133
7.2.1 Study area .....	133
7.2.2 Field data .....	134
7.2.3 Image acquisition and pre-processing .....	135
7.2.4 Machine learning modelling approach .....	137
7.2.5 Accuracy assessment and performance comparison .....	137
7.3 Results .....	138
7.3.1 Effect of biophysical and age characteristics .....	138
7.3.2 Total coffee leaf chlorophyll estimation with all coffee stands .....	139
7.3.3 Total coffee leaf chlorophyll estimation for mature coffee stands only .....	142
7.4 Discussion .....	145
7.4.1 Effect of biophysical and age characteristics .....	145
7.4.2 Effect of spectral settings and spatial resolution on coffee leaf Chl estimation .....	146
7.4.3 Influence of coffee stand age on coffee leaf Chl estimation .....	147
7.5 Conclusions .....	148
7.6 Link to next chapter .....	149
<b>CHAPTER 8: LANDSCAPE SCALE MULTISPECTRAL REMOTE SENSING OF COFFEE FOLIAR NITROGEN CONTENT .....</b>	<b>150</b>
Abstract .....	151
8.1 Introduction .....	152
8.2 Materials and methods .....	154
8.2.1 Study area .....	154
8.2.2 Field data .....	155

8.2.3 Foliar nitrogen determination.....	156
8.2.4 Image acquisition and preprocessing .....	156
8.2.5 Vegetation indices.....	158
8.2.6 Foliar N Prediction algorithm and its implementation.....	159
8.2.7 Accuracy assessment.....	160
8.3 Results.....	161
8.3.1 Correlation between bands and vegetation indices with foliar N.....	161
8.3.2 Foliar N prediction with spectral bands .....	162
8.3.3 Foliar N prediction with Sentinel-2 vegetation Indices .....	164
8.3.4 Foliar N prediction with combination of bands and vegetation indices.....	166
8.3.5 Mapping of low foliar nitrogen areas in coffee plantations .....	168
8.4 Discussion.....	169
8.4.1 Identifying Sentinel-2 bands and vegetation indices related to coffee foliar N .....	169
8.4.2 Quantification and mapping coffee foliar N distribution .....	171
8.5 Conclusions.....	172
<b>CHAPTER 9: MODELLING SPATIAL VARIABILITY OF COFFEE CROP CONDITION WITH MULTISPECTRAL REMOTE SENSING DATA: A SYNTHESIS .....</b>	<b>173</b>
9.1 Introduction.....	174
9.2 Mapping coffee plantations and development of age-masks .....	175
9.3 Identification of anomalous patches in coffee plantations .....	176
9.4 Predicting plant water content in coffee plants using multispectral remote sensing .....	178
9.5 Modelling coffee leaf rust using multispectral remote sensing .....	178
9.6 Multispectral remote sensing of coffee chlorophyll content.....	179
9.7 Multispectral remote sensing of spatial variability of coffee foliar nitrogen.....	180
9.8 Implications for the coffee sector.....	181
9.9 Conclusions.....	182
9.10 Outlook for future research.....	183
References.....	184

## LIST OF FIGURES

Figure 1. 1: Map of the study area showing (a) the location of the study area in Zimbabwe (b) the sites used to identify coffee areas (c) the sites used for anomalies and (d) the site used for field mapping of coffee condition.....	9
Figure 2. 1: Distribution of number of published papers across different subjects in coffee (2007-2017) across the first 100 search results. Search term was coffee for Web of Science, EBSCOHost and Scopus and Coffee arabica for Google Scholar as there was a user named Coffee whose name affected results. Other refers to coffee waste, packaging materials and socio-economic studies. ....	17
Figure 2.2: Physiological characteristics of coffee showing (a) planting arrangement in hedgerows, (b) flowering and (c) coffee fruits. Photo credits: A. Chemura (2017).....	18
Figure 2.3: Map showing the coffee producing countries, types of coffee produced and the number of papers published on use of remote sensing in coffee production.....	32
Figure 3.1: Map of the study area showing general landscape and roads. The insert shows the location of the study area in eastern Zimbabwe and in Southern Africa. ....	41
Figure 3.2: Mean reflectance of the coffee classes for the Landsat 8 OLI bands. ....	48
Figure 3.3: Importance of the spectral bands in class separation shown by decrease in training accuracy when the band is excluded for (3.3a) Scheme A samples on Landsat 8 OLI, (3.3b) Scheme B samples on Landsat 8 OLI, (3.3c) Scheme A samples on Landsat ETM+ and (3.3d) Scheme B samples on Landsat 7 ETM+. The highest percentage increase in mean standard error (%IncMSE) corresponds to the decrease in accuracy when that band is excluded and therefore indicates the most important bands. ....	50
Figure 3.4: Thematic maps produced from random forest classification of the agricultural landscape using Landsat 8 OLI data using (a) Scheme A samples (b) Scheme B samples. ....	51
Figure 3.5: Thematic maps produced from random forest classification of the agricultural landscape using Landsat 7 ETM+ data using (a) Scheme A samples and (b) Scheme B samples. ....	52
Figure 3.6: Age specific thematic maps for coffee produced from (a) Landsat 8 OLI and (b) Landsat ETM+. ....	56
Figure 3.7: Comparison of classifier estimated area with reference data from three farms showing relationships between (a) Landsat 8 OLI data and farm record of area of the three age groups (b) Landsat 8 ETM+ data and farm record of area of the three age groups and (c) Landsat 8 OLI areas and Landsat ETM+ area of the three age groups. The dashed line is the 1:1 line.....	57
Figure 4.1: Map of the study area showing the distribution of coffee ages used in the study adapted from Chemura and Mutanga (2016). The insert shows the location of the study area in Eastern Zimbabwe. ....	68
Figure 4.2: Flowchart of the methodology to detect and map incongruous patches in coffee plantations using time series global and age-adjusted Landsat 8 NDVI and LSWI anomalies. ....	73
Figure 4.3: Effect of age on mean (a) Landsat 8 Scene NDVI and (b) Landsat 8 Scene LSWI extracted from sample points for all the nine scenes. Means followed by the same letter are not significantly different after the Tukey HSD test ( $\alpha \leq 0.05$ ). ....	74
Figure 4.4: Histogram plots of the percentage differences of coffee NDVI (a-c) and LSWI (d-f) values from their age-expected mean values for the 4 <sup>th</sup> of December image scene. ....	75

Figure 4.5: Spatial distribution of percentage deviation from age-expected NDVI mean over the study area across the nine Landsat 8 OLI image dates. ....	76
Figure 4.6: Spatial distribution of percentage deviation from age-expected LSWI mean over the study area across the nine Landsat 8 OLI image dates .....	77
Figure 4.7: Map showing the distribution of incongruent coffee patches because of excessive age-specific NDVI deviation from the mean for the nine Landsat 8 OLI Scenes. ....	78
Figure 4.8: Producer’s and user’s accuracy for normal and incongruous patches for all different types of anomalies.....	79
Figure 4.9: Percentage area determined as incongruent from age-adjusted NDVI anomalies for all the nine Landsat 8 OLI image scenes.....	80
Figure 5.1: Distribution of plant water content of coffee leaves used in the study.....	95
Figure 5.2: Coefficient of determination for (5.2a) all wavebands between 400 and 900 and (5.2b) selected wavebands using the threshold $R^2$ .....	96
Figure 5.3: The application of the reflectance difference and reflectance sensitivity for variable selection. Figure 5.3a shows the variables selected using reflectance difference and Figure 5.3b shows the variables selected using reflectance sensitivity. ....	97
Figure 5.4: Variable importance of wavebands used in modelling plant PWC as selected by (5.4a) Cross-correlation (5.4b) Reflectance Difference and (5.4c) Reflectance Sensitivity. ....	97
Figure 5.5: One-to-one plots showing performance of the RF model for predicting PWC from wavebands for selected through (5.5a) Cross-Correlation, (5.5b) RD and (5.5c) RS.....	98
Figure 6.1: (a) The histogram showing distribution of leaf area of samples and (b) box plots of percent diseased area of leaves as measured on the day of reflectance measurements (N=63). ....	108
Figure 6.2: Mean spectral reflectance of CLR infection levels across the Sentinel-2 bands obtained from resampling hyperspectral imagery.....	113
Figure 6.3: Optimization of Sentinel-2 variables for CLR discrimination through (a) RF-OOB error for spectral bands (b) PLS-DA for spectral bands (c) RF-OOB error for vegetation indices and (d) PLS-DA VIP for vegetation indices. The dotted line shows the cut-off point for variables. .	115
Figure 6.4: Effect of RF and PLS-DA model variable optimization on CLR discrimination accuracies using Sentinel-2 bands and vegetation indices.....	116
Figure 6.5: Correlation between CLR severity and Sentinel-2 MSI (a) Band 2 reflectance, (b) Band 4 reflectance and (c) Band 6 Reflectance. ....	119
Figure 6.6: Determination of $\sigma$ through cross-validated RMSE for use as Gaussian widths in the non-linear RBF-PLS models for (a) All Sentinel-2 MSI bands and (b) Selected Sentinel-2 MSI variables. ....	120
Figure 6.7: Relationship between measured and modelled CLR severity with RBF-PLS with (a) All Sentinel-2 MSI bands and (b) Selected Sentinel-2 MSI variables. ....	121
Figure 7.1: The study area showing the distribution of the coffee fields and stand ages.....	134
Figure 7.2 Influence of stand age and variety on (a) coffee height (b) tree canopy area (c) total Chl obtained from field work (n=72). The numbers (3,4,6,7 and 8) are the ages while letters (A and B) are coffee varieties.....	139
Figure 7.3: Relationship between measured and predicted coffee leaf chlorophyll using (a) all 9 bands at 20m resolution, (b) 4 bands originally at 10m resolutions (c) 5 bands originally at 20m resolutions and (d) all 9 bands at 10m resolutions. ....	140

Figure 7.4: Total coffee leaf Chl distribution maps obtained using the RF on (a) all 9 bands at 20m spatial resolutions and (b) all 9 bands at 10m spatial resolutions for all coffee stands. ....	141
Figure 7.5: Relationship between measured and predicted mature coffee leaf chlorophyll using (a) all 9 bands at 20m resolution, (b) 4 bands originally at 10m resolutions (c) 5 bands originally at 20m resolutions and (d) all 9 bands at 10m resolutions. ....	142
Figure 7.6: Total coffee leaf Chl distribution maps obtained from the RF using (a) all 9 bands at 20m spatial resolution and (b) all 9 bands at 10m spatial resolutions for mature fields. ....	143
Figure 7.7: Effect of coffee canopy area on total coffee Chl estimation accuracy as shown by (a) prediction difference and (b) absolute difference with all coffee data (n=29). ....	145
Figure 8.1: The study area showing the distribution of the coffee fields and stand ages.....	155
Figure 8.2: 1:1 plot showing the relationship between measured and predicted coffee foliar N using (a) all spectral bands and (b) optimized spectral bands. ....	163
Figure 8.3: Predicted distribution of foliar nitrogen levels in coffee leaves obtained from (a) modelling with all spectral bands and (b) modelling with optimized spectral bands. ....	163
Figure 8.4: Variable optimization for prediction of coffee foliar N using (a) spectral bands and (b) vegetation indices. ....	164
Figure 8.5: 1:1 plot showing the relationship between the measured and the predicted foliar N levels with (a) all nine vegetation indices and (b) five optimized vegetation indices. ....	165
Figure 8.6: Predicted distribution of foliar nitrogen levels in coffee leaves obtained from (a) modelling with all vegetation indices and (b) modelling with optimized vegetation indices.....	166
Figure 8.7: 1:1 plot showing the relationship between the measured and the predicted foliar N levels using a combination of optimal bands and vegetation indices. ....	167
Figure 8.8: Predicted distribution of foliar nitrogen levels from combination of optimal bands and vegetation indices showing (a) variation in foliar N and (b) distribution of N levels. ....	167
Figure 8.9: Comparison of coffee area with positive (high foliar N) and negative (low N) anomalous foliar N levels from different modelling approaches. ....	168

## LIST OF TABLES

Table 2.1: Summary of major diseases of coffee, parts affected and author perceived potential for remote sensing based on extend and coffee plant parts affected. ....	24
Table 2.2: Summary of applications of remote sensing technology in coffee crop condition assessments showing the sensor types used, methods, maximum reported accuracy and the references. ....	30
Table 3.1: Description of the land cover classes used in the classification. ....	42
Table 3.2: Spectral and spatial characteristics of the Landsat 8 OLI and Landsat ETM+ data ....	42
Table 3.3: The number of training regions of interest (ROIs) and validation ROIs used for the classification by the two schemes with Landsat 8 OLI and Landsat 7 ETM+. Each ROI represents a field collected sample point. ....	47
Table 3.4: Transformed divergences indices showing pairwise interclass separability of coffee classes and other classes in the training samples. ....	49
Table 3.5: Confusion matrix and associated classification accuracies based on independent test data set for Scheme A produced from Landsat 8 OLI and Landsat 7 ETM+. The accuracies include overall accuracy (OA), kappa (Kc), user's accuracy (UA) and producer's accuracy (PA). ....	53
Table 3.6: Confusion matrix and associated classification accuracies based on independent test data set for Scheme B produced from Landsat 8 OLI and Landsat 7 ETM+. The accuracies include overall accuracy (OA), kappa (Kc), user's accuracy (UA) and producer's accuracy (PA). ....	54
Table 3.7: McNemar's test results for comparison between Landsat ETM+ and Landsat 8 OLI for Scheme A and Scheme B classifications*. ....	55
Table 3.8: Area (ha <sup>-1</sup> ) for each land cover class as obtained from the random forest classification of Landsat 8 OLI and Landsat 7 ETM+ with Scheme A and Scheme B samples. ....	57
Table 4.1: Detail on Landsat 8 Scenes used to derive NDVI. ....	69
Table 4.2: Accuracy assessment of model performance. ....	79
Table 5.1: Descriptive statistics of the PWC for the three stress levels. ....	95
Table 5.2: Summary model performance evaluation using test dataset (n=24). ....	98
Table 6.1: Description of levels of CLR infection levels, sample images and number of samples used in the study. ....	108
Table 6.2: Specifications of the Sentinel-2 Multispectral Instrument (MSI). ....	109
Table 6.3: Selected vegetation indices (VIs) evaluated in the study. ....	110
Table 6.4: CLR discrimination accuracies obtained using Sentinel-2 MSI derived spectral bands and vegetation indices. ....	114
Table 6.5: CLR discrimination accuracies derived using the most important selected model variables (i.e. spectral bands and optimized vegetation indices). ....	117
Table 6.6: Comparison of the performance of CLR discrimination accuracy results derived using RF and PLS-DA models with and without variable optimization. ....	118
Table 6.7: Correlation coefficients (r) and significance of correlation ( $\alpha < 0.05$ ) between Sentinel-2 MSI bands and CLR severity. ....	119
Table 6.8: Error metrics for all models in predicting CLR severity from Sentinel-2 MSI variables with RBF-PLS. ....	121
Table 7.1: Specifications of the Sentinel-2 Multispectral Instrument (MSI). ....	135
Table 7.2: Specifications of the Sentinel-2 Multispectral Instrument (MSI) band settings showing spatial/spectral combinations, number of bands and spatial resolution. ....	136

Table 7.3: Performance evaluation of different band settings in predicting coffee leaf chlorophyll using all coffee stands. ....	140
Table 7. 4: Performance evaluation of different band settings in predicting coffee leaf chlorophyll using mature coffee stands only. ....	143
Table 7.5: Correlation coefficients (r) between predicted coffee leaf chlorophyll using different band settings*.....	144
Table 8.1: Descriptive statistics of field data.....	156
Table 8.2: Specific details about the Sentinel-2 MSI image used.....	157
Table 8.3: Specifications of the Sentinel-2 Multispectral Instrument (MSI) band settings showing centre wavelengths, band width and spatial resolution. ....	158
Table 8.4: Name, formula and Sentinel-2 bands utilized in selected vegetation indices (VIs) evaluated in the study. ....	159
Table 8.5: Relationship between Senitnel-2 MSI variables and coffee foliar N.....	161
Table 8.6: Accuracy of coffee foliar N prediction with all and optimized Sentinel-2 bands.....	162
Table 8.7: Performance of predicting coffee foliar N with all and optimized Sentinel-2 vegetation indices. ....	164
Table 8.8: Performance of predicting coffee foliar N with a combination of optimal bands and vegetation indices.....	166
Table 8.9: Sizes (in ha) and % of low, sufficient and high coffee foliar N levels. ....	169



## **SECTION 1: INTRODUCTION AND OVERVIEW**

# CHAPTER 1: GENERAL INTRODUCTION



Photo credits: A. Chemura

## **1.1 Coffee and coffee production**

The production of coffee (*Coffea arabica* L.), a perennial tree crop whose fruits are used for making coffee, a very popular non-alcoholic beverage, is an important agricultural activity in over sixty countries across four continents. The majority of producer countries are developing countries in sub-Saharan Africa, South East Asia and Central America. In these countries, over twenty million coffee farmers tend about three billion coffee trees to supply the global coffee demand of over two billion coffee cups consumed daily (Waston & Achinelli, 2008; ICO, 2015). Accordingly, coffee production contributes over US\$20 billion annually to producer countries, with incomes cascading down to millions of smallholder farmers and farm workers for which coffee is a unique legal source of income and livelihoods. In addition to being a significant source of livelihoods, coffee plantations also contribute to landscape scale ecosystem processes such as carbon sequestration, erosion control and provision of other ecosystem services (Brauman et al., 2007). Unlike annual crops, perennial crops such as coffee represent a long-term investment because they are in the field for a long period of time.

Evidence shows that there are increases in frequency, magnitude and impacts of production challenges in the coffee sector threatening livelihoods, economies and ecosystems in coffee producing areas (Jayathilaka et al., 2012; Rahn et al., 2013). Water stress, increasing pest incidences, more severe and frequent disease pressure, and limited soil nutrient supply, coupled with slow developments in technologies to deal with these challenges, are affecting productivity of coffee. For example, Hillocks et al. (1999) pointed that some pests such as coffee white stem borer (CWB) and coffee leaf miner (CLM) previously considered minor are having significant impacts on coffee production. Similarly, African Development Bank (2010) observed that the coffee sector is facing many production challenges that are limiting yields and quality, and consequently reducing farmers' incomes.

Many studies have demonstrated that production challenges are increasingly becoming more pronounced, extended and severe because of climate change and variability (Ghini et al., 2011; Jayathilaka et al., 2012; Kutuwayo et al., 2013). Due to these challenges, productivity and area under coffee production of coffee are falling. In order to meet growing demand and export commitments, farmers have to expand their coffee fields, in some areas through conversion of environmentally sensitive land to crop land. This poses several environmental challenges

especially now that many coffee consumers are increasingly demanding sustainably produced coffee. Remote sensing can be used in enhancing coffee production to reduce production costs, increase productivity and safeguard ecosystem services in coffee producing areas.

## **1.2 Remote sensing in agronomic decision making**

There is an urgent need to increase coffee production volume and quality to support livelihoods without adversely affecting the environment, which is the basis of productivity. To achieve this, there is need to increase productivity of coffee. One way of increasing productivity per unit area is through crop condition assessments and season-long monitoring to support agronomic decision making (Baret et al., 2007). Current agronomic decision making in coffee is based on calendar based management programs, spontaneous field inspections and sampling by trained and experienced personnel who scout the often large coffee plantations looking for signs of production challenges. These methods are not only labour intensive, but also conclusive once economic damage has been inflicted on the crop. In addition, these methods are not spatially explicit and since they are based on sampling, assume that the crop condition is uniform across coffee fields. For example, if an identified pest is found to exceed the threshold on one tree, the whole field is sprayed. This is not only expensive to farmers who bear the cost of excessive agrochemical applications, but also to the environment that absorbs the excess chemicals.

There is overwhelming evidence that remote sensing utility in agriculture is significant especially under the broader goal of precision agriculture. For instance, it was demonstrated that remote sensing can provide automated routines for accurate and early detection and differentiation of plant diseases for farm-level decision making (Moshou et al., 2004; Rumpf et al., 2010; Mahlein et al., 2012a; Barbedo, 2013). Similarly, crop nutrient status can also be determined and mapped using remote sensing methods to identify anomalous areas and therefore ensuring uniform yields across fields (Scharf et al., 2002; Zhao et al., 2005). Furthermore, signs of plant water stress can also be detected using spectral bands and vegetation indices (Peñuelas et al., 1994; Barnes et al., 2000; Eitel et al., 2006; Brillante et al., 2016; Dangwal et al., 2016). More generally, cropped area and plant populations can be accurately estimated together with indications of crop vigour that are directly related to yield (Maxwell et al., 2004; Gitelson et al., 2005; Atzberger, 2013). This indicates that the ability of remote sensing to enhance agricultural productivity has since moved from potential to reality with evidence of practical applications across the world.

### **1.3 Challenges of remote sensing application in coffee production**

The majority of the studies on agricultural applications of remote sensing have been on annual crops, with many of the developed approaches difficult or impossible to transfer to perennial crops such as coffee and cocoa. This is attributed to the differences in physiology, functional photosynthetic pathways, phenological cycles, and type and extent of challenges between annual and perennial crops. For instance, many of the remote sensing applications in agriculture are reported for crop species in the grass family such as corn (*Zea mays* L) (Maxwell et al., 2004; Freeman et al., 2007; Kuri et al., 2014), wheat (*Triticum aestivum* L)(Sembiring et al., 2000; Liu et al., 2003; Riedell et al., 2003; Atzberger et al., 2010; Dangwal et al., 2016) and rice (*Oryza sativa* L.) (Lee et al., 2008; Cao et al., 2015; Kanke et al., 2016; Zhou et al., 2016). Soybean (*Glycine max* L. Merr.) (Nutter Jr. et al., 2002; Venteris et al., 2015), and leafy vegetables (Mahlein et al., 2012b), annual field crops such as cotton (*Gossypium hirsutum* L.) (Zarco-Tejada et al., 2005; Ko et al., 2006) and sunflower (*Helianthus annuus* L.) (Gutierrez et al., 2008) also feature prominently in remote sensing applications in agriculture.

As a result, in published reviews of applications of remote sensing for different aspects of precision agriculture, reports on perennial tree crops are scarce, far in-between and in a few cases, they appear to be overshadowed by annual crops. For instance, very few perennial woody species-related applications appear in reviews of broad remote sensing in agriculture (Hatfield et al., 2008; Ge et al., 2011; Mulla, 2013), unmanned aerial systems and precision agriculture (Zhang & Kovacs, 2012; Sankaran et al., 2015) and remote sensing for disease detection (Mahlein et al., 2012a). There is thus a huge gap between the application of remote sensing in annual and perennial agricultural systems, particularly in coffee production. This gap has to be closed to advance the benefits of coffee production in ensuring local and national economic development and provision of ecosystem services.

There are a number of challenges that make remote sensing applications in perennial tree crops in general and for coffee in particular more complicated than in a natural grassland, forest and in annual crops. Firstly, coffee is in the field all year round for up to between 15 and 25 years of production. Because of this, there are no distinct phenological cycles that can be used for its characterisation as it does not shed its leaves completely. Secondly, coffee is planted in hedgerows or covas that produce a systematic spatial planting arrangement. In addition, coffee is managed to

ensure that there are gaps between rows for efficient scouting, fertilisation and harvesting operations which results in a permanent background spectral effect from soil or weeds (Brunsell et al., 2009). This challenge is exacerbated in areas where coffee is being produced under shade. To ensure production continuity, farmers always grow coffee at different age groups and yet age has significant effects on the fractional components that influence the spectral signature of coffee plants (Netto et al., 2005). Furthermore, coffee has been shown to follow a specific biennial bearing effect that is related to its fruiting sequence that also has a bearing on remote sensing applications (Bernardes et al., 2012). There are also a number of coffee varieties produced and these have different growth and structural characteristics that may influence the remote sensibility of coffee plantations for crop biophysical and biochemical parameter assessment, and their relationship to yield.

In addition to the unique planting system used in coffee production system, and most importantly, production limiting factors in coffee plants are interrelated and thus may occur concurrently as confounding factors. For example, the coffee white stem borer (*Monochamus leuconotus L.*) is prevalent in areas of low soil fertility while *Cercospora* leaf spot is common in coffee with frequent water stress and nutrient imbalances (Logan & Biscoe, 1987; Kutuywayo et al., 2013). Thus, it may be difficult to know exactly which factor is contributing to low agricultural productivity and what to deal with first in such instances. These broad challenges explain the limited application of remote sensing in perennial tree crops such as coffee and yet they are hugely important in that they are part of a few systems that simultaneously support economic development and ecosystem services.

#### **1.4 Importance of multispectral remote sensing in coffee production**

Remote sensing, particularly the use of hyperspectral sensors provides very promising options for early, objective and spatially resolved crop condition assessments in agricultural crops. Hyperspectral remote sensing data consists of many, very narrow contiguous spectral wavebands located from the visible, near infrared, mid-infrared and thermal infrared portions of the electromagnetic spectrum (Blackburn, 2007; Knyazikhin et al., 2013). However, relying on the hyperspectral data for detecting and quantifying coffee condition assessments for field crop management presents challenges in data handling, dimensionality, high costs, noise and unavailability, all of which dissuade potential users from relying on this range of data for

determining crop condition assessments (Dube & Mutanga, 2015). New generation multispectral space-borne earth observation instruments, such as WorldView-2, RapidEye and Sentinel-2 multispectral imager have incorporated narrow wavebands including those in the red-edge position that were not available in predecessor sensors (e.g. Landsat series, MODIS, SPOT, ASTER etc). These technological advancements therefore provide an opportunity for timely landscape or farm-based assessment of crop condition (that is health status and yield estimation). Unlike the hyperspectral sensors, multispectral sensors have a huge swath-width and are currently available at low or no costs for many developing countries where coffee is produced. They are also available sequentially, providing potential for temporal analysis of coffee condition which is important for long-term monitoring of perennial crops such as coffee.

## **1.5 Aim and objectives**

### **1.5.1 Aim**

The overall aim was to develop a landscape scale approach to model spatial variability in coffee (*Coffea arabica* L.) condition using multispectral remote sensing data.

### **1.5.2 Objectives**

The specific objectives were to:

1. Develop detailed age-specific thematic coffee plantation maps in heterogeneous agricultural landscapes.
2. Integrate coffee plantation age in identification and mapping of incongruous patches using multi-temporal remotely sensed data.
3. Evaluate the potential of multispectral remote sensing data for quantifying leaf water stress in coffee.
4. Assess the potential of multispectral remote sensing bands and vegetation indices for discriminating and predicting coffee leaf rust infection.
5. Determine the effects of spectral band settings, spatial resolution and crop canopy cover in predicting leaf chlorophyll content in coffee plantations.
6. Model the spatial variability in foliar nitrogen in coffee plantations with multispectral level data for agronomic decision making.

## **1.6 Description of the study area**

Two study areas were used; one for experimental studies and the other for field studies.

### **1.6.1 Coffee Research Institute**

For laboratory spectroscopy studies, research was carried out at Coffee Research Institute (CoRI) in Chipinge, Zimbabwe. CoRI is located at coordinates 20°14'4.11"S and 32°38'49.98"E at an altitude of 1100 meters above sea level (m.a.s.l.). The average annual rainfall at CoRI is 1180mm of which 80% falls in five months from November to March. The mean maximum daily temperature is 20°C and minimum is 14°C. Most of the soils in this area are leached and strongly weathered and in the Orthoferrallitic group derived from Umkondo quartzite and sandstone (Chemura, 2014). Chipinge is in the main coffee production zone in Zimbabwe.

### **1.6.2: Ward 19 Chipinge District**

Field studies were carried out in Ward 19 of Chipinge district, which consists of coffee estates and smallholder communal coffee farmers. The site is located in Mossurize sub catchment, South-east of Zimbabwe between latitude 32°36'1.00"E and 32°48'1.00"E, and longitude 20°20'1.00"S and 20°33'1.00"S in Chipinge district (Figure 1.1). The study site represents the current largest coffee producers in Zimbabwe in terms of volume and area. The climate of the area is subtropical with two distinct seasons, divided almost equally between months of the year (October to March is the growing season while April to September is the dry season). Compared to other parts of Zimbabwe, the area receives relatively high mean annual rainfall totals (1200-1300 mm/year) with mean annual temperatures around 22.5°C. Together with deep red clayey soils formed from mafic rocks, the climatic conditions make the area suitable for quality coffee production. As the area is dominated by large scale coffee farms, the coffee production system is sun-coffee, which means plantations are exclusively grown with coffee and not mixed with shade trees. In Chipinge district, the mean size of coffee farms is 25 ha and is dominated by Catimor varieties due to their resistance to coffee leaf rust (Chemura et al., 2015a).



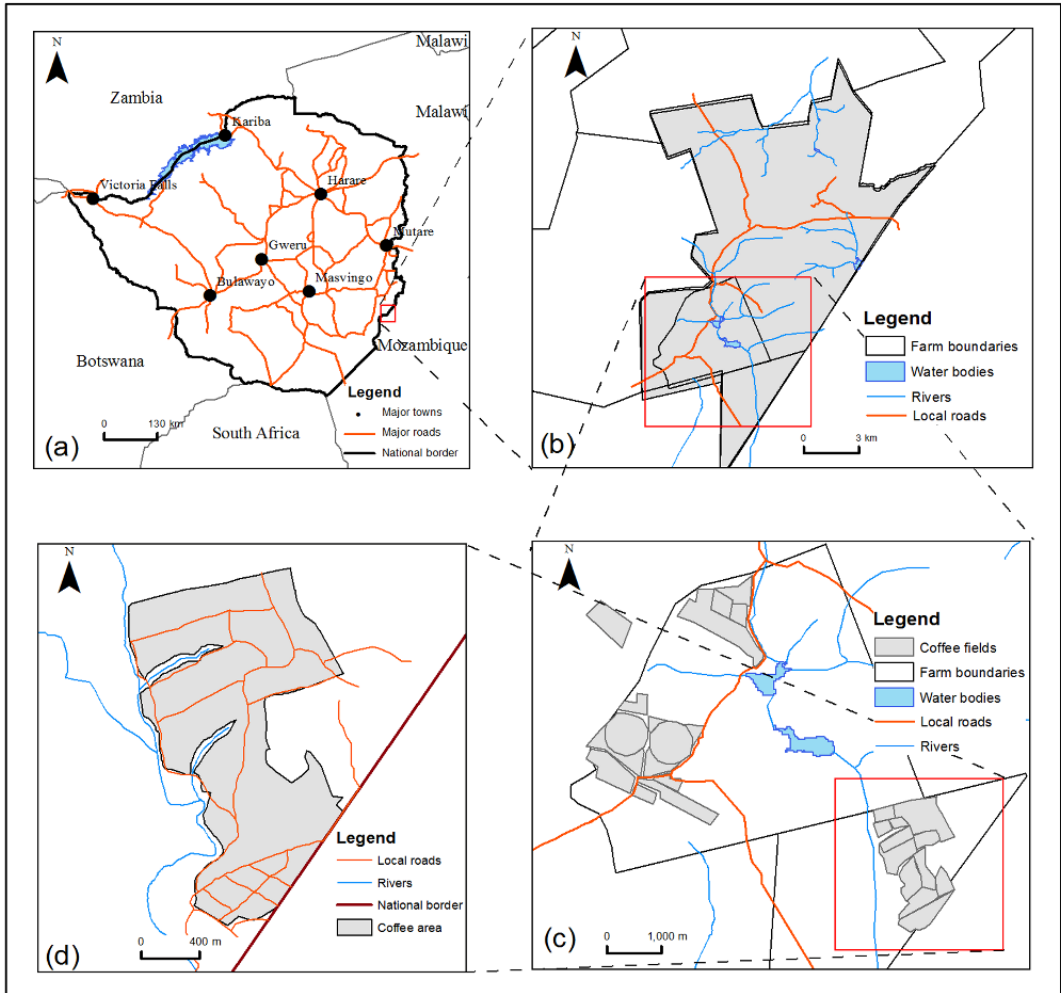


Figure 1. 1: Map of the study area showing (a) the location of the study area in Zimbabwe (b) the sites used to identify coffee areas (c) the sites used for anomalies and (d) the site used for field mapping of coffee condition.

**1.7 Thesis structure**

This thesis is presented in four sections, each with standalone chapters that can be independently read as research articles published in international journals but all aimed towards developing spatial models for coffee condition assessment. Section 1 provides an introduction and overview. Section 2 sets the stage for coffee condition assessment where initial processes required for successful coffee characterisation are presented. Section 3 presents the experimental studies on dealing with potential biotic and abiotic confounding factors to coffee condition assessment. Section 4 presents application of multispectral data in estimating coffee chlorophyll and nitrogen content, two important indicators of coffee condition together with the synthesis. Most of the

content and structure of the manuscripts submitted to peer-reviewed journals has been retained and thus each chapter has an abstract, introduction, materials and methods, results, discussion and conclusion sections. However, because of this, there are some duplications and overlaps, particularly in the introduction and methods sections of these chapters. This is assumed to be of little consequence since these chapters are independent but related peer-reviewed journal articles, that can be read separately, without losing the overall context of the study. This thesis is made up of nine chapters as summarized in the following section.

## **1.7.1 Section 1: Introduction and Overview**

### **1.7.1.1 Chapter One: Introduction**

This chapter serves as an introduction that gives the background and broad statement of the problem that underpins this study. It describes the importance of coffee production, its physiological and plant characteristics and how this affects remote sensing potential of the crop. This chapter also explains the importance of multispectral level remote sensing in tropical field crops such as coffee. The aim and objectives of the study and a detailed description of the study area are also provided.

### **1.7.1.2 Chapter Two: Literature review**

A state of the art literature review on crop condition assessments with remote sensing in coffee is presented in this chapter. A synopsis of the global coffee industry is provided together with the physiology of the coffee plant. An analysis of major coffee biotic and abiotic stressors that require monitoring is provided with an assessment of the application of remote sensing in their monitoring. A review of reported applications of remote sensing in coffee crop management is presented together with sensors, algorithms and future perspectives.

## **1.7.2 Section 2: Setting the stage for coffee condition assessment**

### **1.7.2.1 Chapter Three: Development of an age mask for coffee condition assessment**

This chapter presents an approach for development of age-specific maps of coffee from multispectral remote sensing data in heterogeneous agricultural landscapes. These age-specific maps are important for use as age-masks for coffee condition assessment. The cost of producing

these age specific masks in terms of accuracy in comparison to general landcover maps is evaluated in relation to the utility of such a map for crop condition assessment.

### **1.7.2.2 Chapter Four: Identification of anomalous patches in coffee plantations**

In this chapter, the age-masks produced from the preceding chapter are used to identify areas in coffee fields that are anomalous and therefore require attention. An approach using deviation from age-specific mean Normalised Difference Vegetation Index (NDVI) and Land Surface Water Index (LSWI) is developed using Landsat data and applied to monitor coffee condition over time. The value of such an approach is analysed for field application in coffee condition assessment at various scales.

### **1.7.3 Section 3: Remote sensing individual coffee stressors**

#### **1.7.3.1 Chapter Five: Predicting plant water content in coffee plants using multispectral remote sensing**

This chapter presents an experimental evaluation of the potential for remote sensing abiotic stress that contribute to anomalous patches in coffee plantation. A machine learning approach to predict plant water stress in coffee using multispectral level data is presented. Modelling plant water content using secondary water absorption features is presented as a promising approach for coffee condition assessment to explain anomalies.

#### **1.7.3. 2 Chapter Six: Modelling coffee leaf rust using multispectral remote sensing**

In this chapter, an experimental evaluation of the possibility to discriminate and quantify biotic factors that are responsible for causing anomalous conditions in coffee is presented. Coffee leaf rust, one of the major diseases of coffee, is discriminated into three levels (healthy, moderate and severe) using multispectral level band settings and derived vegetation indices. The severity of the infection is also quantitatively modelled. For both discrimination and severity modelling, machine learning algorithms are applied.

### **1.7.4 Section 4: Empirical modelling of coffee condition using multispectral data**

#### **1.7.4.1 Chapter Seven: Multispectral remote sensing of coffee chlorophyll content**

Having successfully demonstrated the possibility to identify and quantify biotic and abiotic factors that can confound coffee condition assessment at field level in experimental studies, this chapter

presents field level application of multispectral data in estimating coffee chlorophyll content. The Sentinel-2 MSI data is used to map and quantify the spatial variability of chlorophyll in coffee fields that can be related to the biotic or abiotic factors discussed in Chapter 5 and 6.

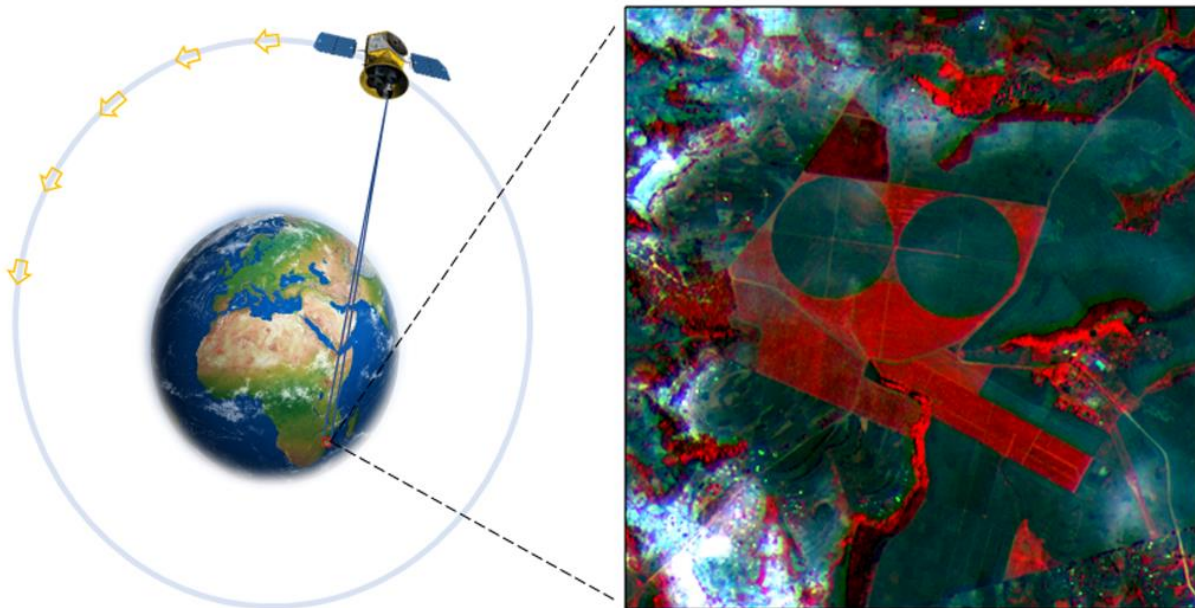
#### **1.7.4.2 Chapter Eight: Multispectral remote sensing of spatial variability of coffee foliar nitrogen**

This chapter developed an approach for monitoring coffee foliar nitrogen, one of the most important determinants of crop productivity, with multispectral level remote sensing data. The role of spectral bands and vegetation indices and their optimisation is presented to identify important parameters for the task.

#### **1.7.4.5 Chapter Nine: Synthesis**

This chapter provides a synthesis of this research work. It presents a summary of the major findings of the study and their meaning for the coffee sector. Deductions are also made about the achievements and limitations of the presented project that feed into recommendations for future research in coffee crop condition assessment. A single reference list is provided at the end of the thesis.

## CHAPTER 2: ADVANCES IN SENSOR APPLICATIONS FOR COFFEE (*COFFEA ARABICA* L.) CROP CONDITION ASSESSMENTS AND MONITORING



**This chapter is based on:**

Chemura A and Mutanga O. (Under Review) Advances in sensor applications for coffee (*Coffea arabica* L.) crop condition assessments and monitoring, **Agronomy for Sustainable Development**.

## **Abstract**

Coffee is an important crop for producers in developing countries and their economies. These producers are increasingly facing production constraints due to climate change and are required to sustainably produce their coffee to meet consumer demands. They therefore require intra and inter-seasonal monitoring of coffee plants to achieve sustainable production. There are many developments in technology, data storage, processing algorithms and platforms for applications of remote sensing that provide opportunities for application of remote sensing in coffee production. This paper provides a detailed survey on constraints to coffee production and the progress of remote sensing applications in coffee production across the world. The major limiting factors to coffee productivity and the potential of remote sensing to detect, quantify and model these were assessed. The availability and applicability of different remote sensing datasets, in terms of spatial radiometric and temporal resolutions are discussed in relation to physiology, climate and physical conditions in coffee producing countries. It was observed that the majority of research on coffee has focused on brewed coffee and related products with less research papers focusing on production. Of the papers on remote sensing of coffee production, many have focused on the application of sensors in mapping cropping areas with very little research on coffee condition assessment. It was concluded from the analysis that although there is potential for application of remote sensing in coffee condition assessment, there is need for more research focusing on high fidelity sensors and high spatial and spectral resolutions that utilizes readily available data and advanced algorithms targeted for remote producing areas.

## 2.1 Introduction

Perennial tree crops such as coffee are important agricultural crops for many developing countries. The coffee plant belongs to the family *Rubiaceae* and genus *Coffea*, which is subdivided into *Mascarocoffea*, *Eucoffea*, and *Paracoffea* sections. The cultivated species of coffee (*Coffea canephora* and *Coffea arabica*) belong to the subsection *Paracoffea* having originated from sub-tropical montane areas in Africa (Nair, 2010). Coffee does not shade its leaves although fruiting load alters leaf area index (LAI). *Coffea arabica* is an allotetraploid with a chromosome constitution of  $2n = 4x = 44$ , and is known for its high quality, intense aroma, lower caffeine content, and a less bitter taste, all of which give it higher aggregate prices (Lashermes & Anthony, 2007). Under commercial production, coffee has a gestation period of 2-3 years after planting before economic yields can be obtained, and can be productive for an average of 15 years, although productive plants over 50 years have been reported (Nair, 2010).

Coffee is sensitive to extremes of temperature, preferring temperatures between 15 and 25°C, and well distributed rainfalls are a key requirement (>1800 mm/annum) for healthy growth and productivity (Wrigley, 1988). It is known that the rate of net photosynthesis of coffee plants slows down progressively above a temperature threshold of 25°C to the detriment of the plant. On the other hand, the plant responds in several ways to moisture deficit stress such as leaf folding, leaf and branch dieback, leaf shape change and altered biomass allocation between the roots, the stems, and leaves (DaMatta, 2004; Dias et al., 2007; Chemura et al., 2014). The crop is produced across a range of altitudes although, generally, it does very well between 1200–1800 m above sea level in the equatorial and tropical zones (Coste, 1992). The ideal soils are deep, clayey, slightly acidic, and rich in humus and exchangeable bases (Logan & Biscoe, 1987; Nair, 2010).

In addition to being a significant agricultural crop, coffee represents a long-term capital investment. This is because it is in the field all-year round and in production for longer periods and therefore requires accurate, reliable and cost-effective crop condition monitoring strategies for diseases, pests, water stress, soil fertility and other crop stressors. However, the current monitoring methods largely rely on spontaneous field inspections and sampling. There is thus a general lack of fine-scale long-term datasets to monitor plantation tree crop stressors (Jeger & Pautasso, 2008). Long term monitoring is required in identifying and analysing inter and intra-annual variations in crop conditions useful for farm managers, investors, insurers and other stakeholders interested in

monitoring coffee condition and productivity.

There is an increasing body of literature that shows that producing the required amounts of good quality coffee is now more difficult than before (ICO, 2009; Hagggar & Schepp, 2011). There are a number of reasons for this. In traditional production areas rainfall patterns have become unpredictable and unreliable exposing the coffee plants to frequent and often severe water deficit stress. Water deficit stress has permanent and short-term effects on coffee physiology and productivity potential (DaMatta, 2004). These changing weather patterns due to climate change and variability are even projected to reduce the suitability of many coffee producing areas. This is through increasing production costs, reducing coffee quality and increasing the frequency and severity of coffee pests and diseases, all of which ultimately erodes the profitability of the coffee enterprise (Laderach et al., 2006; ICO, 2009; Hagar & Schepp, 2011).

The growing coffee production challenges increase the demand for crop assessments, particularly given the limitations in current coffee monitoring methods. These therefore call for increased efforts in development, testing and implementation of more efficient production technologies and systems in coffee producing countries such as Zimbabwe, Zambia, Malawi, Kenya, Tanzania, Ethiopia, Burundi, Uganda in Africa and Brazil, Guatemala, Costa Rica and Uruguay in South America. These are important to sustain the economies dependent on coffee in these countries and maintain the culture in which coffee is associated with in developed countries. Despite this urgent requirement, much of the research around coffee from major research databases has focused on coffee consumption rather than on production with over 50% of results from Web of Science, EBSCOHost, Scopus and Google Scholar research databases being coffee consumption studies. The effects of coffee on health and coffee beverage chemistry dominate the coffee research space in published papers (Figure 2.2). On the other hand, there have been very few studies on climate change and use of remote sensing in coffee production despite the pressing climate change related challenges the sector is facing and the potential of the later to help with handling these challenges.



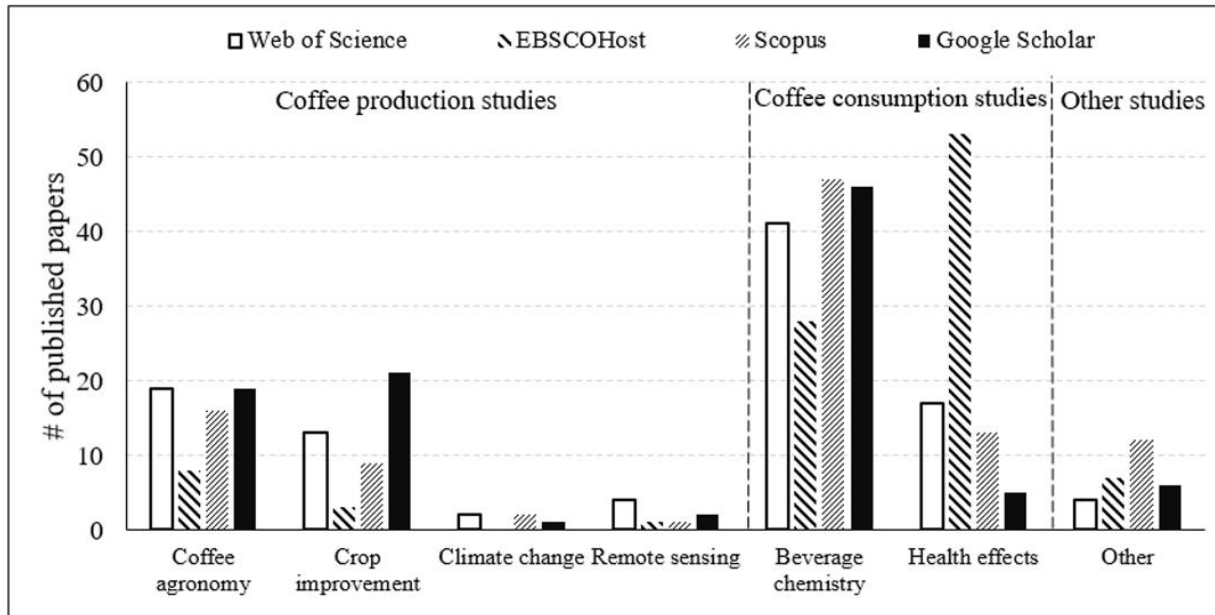


Figure 2. 1: Distribution of number of published papers across different subjects in coffee (2007-2017) across the first 100 search results. Search term was coffee for Web of Science, EBSCOHost and Scopus and Coffee arabica for Google Scholar as there was a user named Coffee whose name affected results. Other refers to coffee waste, packaging materials and socio-economic studies.

The aim of this paper is to review the physiological characteristics of coffee in relation to remote sensing applications and demonstrate the progress made in remote sensing applications in coffee production in order to determine future research directions. Firstly, the review highlights specific challenges faced in perennial tree crops in comparison to annual crops. Secondly, the effect of physiology, phenology and limiting factors to operational remote sensing are highlighted. Furthermore, the applicability of available remote sensing sensors, platforms and algorithms is then discussed with regards to the identified challenges. The future prospects of condition assessments are given with specific focus on monitoring coffee leaf rust as a potential example application.

## 2.2 Physiological characteristics of coffee and their potential influence on remote sensing

Coffee is planted in hedgerows and can reach up to 3m depending on varieties, environmental conditions and crop husbandry, and has an average leaf area index (LAI) of 6 (Coste, 1992). Plant populations are based on spacing and planting arrangements and vary between 2000 to 6000 plants/ha as influenced by variety, region, soil types, shading design or just farmer experimentation (Logan & Biscoe, 1987). The coffee plant has an orthotropic stem from which series of branches (primary, secondary, tertiary and quaternary) develop. Leaves develop from each of these

depending on the age of the crop as opposite pairs at right angles to the previously formed pair. Older leaves are found at the base of the tree while young leaves are on the top parts of the plant (Coste, 1992; Lashermes & Anthony, 2007). Leaf life is variable and new leaf production compensates for old leaf fall which occurs throughout the year. In new leaves, light green and bronze are predominant colours depending on age and variety (Logan & Biscoe, 1987). Mature leaves are lanceolate, ribbed and waxy, and vary in size due to variety, environmental conditions, crop management and season (Coste, 1992).

The development of a coffee serial bud into an inflorescence is catalysed by rainfall or irrigation after a required short period of drought. Full floral blossoming then occurs 10 days later and then the flowers begin to wither with development of pinheads (Coste, 1992). The vegetative cover in coffee field progressively increases with age and can reach up to 90% for old coffee (Vieira et al., 2006). Thus, there is always a background effect in coffee plantations unlike in natural forests and annual crops such as maize and wheat for which much of the agricultural remote sensing has been done.



Figure 2.2: Physiological characteristics of coffee showing (a) planting arrangement in hedgerows, (b) flowering and (c) coffee fruits. Photo credits: A. Chemura (2017)

As a perennial tree, the coffee plant is in the field throughout the seasons and the year, and as such, seasonal and year-to-year fluctuations in climatic conditions are considered as part of the transient environmental fluctuations the plant should adapt to. There is overwhelming evidence that one key physiological adaptation method of coffee is biennial bearing cycle where one year of good harvest is followed by a subsequent year of lower yields (Nair, 2010; Bernardes et al., 2012). One of the

reasons for this phenomenon is that up to 95% of total plant nitrogen can be taken up by fruits in heavily bearing coffee trees (Wrigley, 1988), which results in remarkable decreases in the leaf nitrogen concentrations after harvests, contributing to subsequent lower yields the following year. Another explanation is that higher yields contribute to lower new primary development which provided little basis for next year fruit development (Logan & Biscoe, 1987). The following year will be a year of recovery that will culminate in higher yields the following year and the cycle continues. From a field level remote sensing perspective, the condition assessment may need to be sensitive to capture these ‘natural’ changes or at least be adjustable for them. Thus, regardless of other biotic and abiotic stressors, there is an inherent fluctuation in coffee plant physiology that can influence observed year-to-year changes in spectral signatures of coffee.

### **2.3 Major biotic and abiotic stressors of coffee arabica plants that require monitoring**

Although the genetic diversity of coffee is much higher in tropical Africa where coffee originated, the majority of the coffee under current production are from a few genetically similar trees (Nair, 2010). Due to this, cultivated varieties have a very narrow genetic base, making them susceptible to pest and disease epidemics and limiting selection and improvement efforts (Van der Vossen & Walyaro, 2009). Major plant stressors for coffee are nutrient deficiencies, plant water stress and pests. Major diseases of economic importance for coffee are Coffee Leaf Rust (*Hemileia vastatrix* (B & Br)), Coffee Berry Disease (*Colletorichum kahawae* (J.M.Waller & Bridge)) and Fusarium Bark Disease (*Fusarium lateritium*) while major pests are the Coffee White Stem Borer (*Monochamus leuconotus*) (Brown, 2008; Kutuwayo et al., 2013).

#### **2.3.1 Nutrient deficiencies**

Coffee plants require a high level of fertility and an intensive fertiliser programme is a prerequisite for successful coffee production. Without proper soil fertility management, nutrient deficiencies will affect productivity and survival of the coffee plants. Nitrogen (N) is known as the single most limiting factor to coffee productivity (Coste, 1992). Mature coffee requires a total of about 105 kg ha<sup>-1</sup> of N to achieve yield levels of 1 tone ha<sup>-1</sup> per year and yet some coffee farms report up to 6 tonnes clean coffee per ha<sup>-1</sup> per year (Chemura, 2014). N plays several interconnected roles in coffee plant development and productivity and its deficiency results in stunted growth, the characteristic yellowing of leaves and floral abortion and fruit drop that is usually accompanied with die back. Phosphorus (P) is another important nutrient for coffee as the element has a

stimulating effect on root growth and assists in hastening crop maturity (Logan & Biscoe, 1987). Furthermore, in coffee areas where Potassium (K) is inherently low, the appearance of K deficiency can be spectacular. This normally happens in areas where irrigation water has high level of Mg and Ca as these are antagonistic to the uptake of K by coffee plants. The appearance of K deficiency symptoms often coincides with berry development and are shown by small necrotic areas on the leaf margins that rapidly extend towards the centre of the leaf and towards each other, eventually coalescing (Wrigley, 1988). K deficiency results in serious defoliation, often along the complete lengths of the coffee branches, and this significantly alters the coffee plant LAI (Wrigley, 1988; Nair, 2010).

### **2.3.2 Plant water stress and drought**

Irrigation water management is another critical aspect of coffee management in coffee production as plant water stress and drought have significant impacts on the crop. Excessive irrigation, in addition to being costly and environmentally unfriendly, may result in root rot caused by *Armillaria mellea*. On the other hand, moisture deficit stress results in dieback, wilting and opportunistic disease and pest attacks, resulting in reduced production in both cases (DaMatta et al., 1993; Kutuwayo et al., 2010). Depending on the timing in coffee plant cycle, moisture stress can also result in floral abortion and premature berry drop, which results in reduced economic yields of coffee (Chemura et al., 2014). Coffee plants affected by plant water stress go through a variety of physiological processes, such as damage or removal of the waxy cuticle, destruction of cell walls, reduced stomatal conductance and retarded rates of net carbon assimilation (DaMatta, 2004).

Coffee plants respond in different ways to drought and plant water stress depending on the stage in the growth cycle. For example, during the pinhead stage, the effects of moisture stress are exhibited by scotched berries but during berry expansion, moisture stress results in blue-green coloured berries that are later dropped. Towards berry maturity, plant water stress results in reduced size beans which are ragged with reduced quality and mass (Logan & Biscoe, 1987). It is therefore imperative that water requirements are regularly monitored for irrigation scheduling with current methods relying on soil water holding capacity, measurements of soil tension, evaporation, rooting depth and the crop factor. These, however, are not direct estimates of plant water content of the coffee plants, but are only used as proxies on which decisions are based.

### **2.3.3 Coffee Leaf Rust (CLR)**

There is consensus that CLR caused by the fungus *Hemileia vastatrix* is by far the most severe of all coffee diseases, especially for *Coffea arabica* (Dinesh et al., 2011; Cressey, 2013). The basidiomycete *H. vastatrix* is an obligate biotrophic fungus that is found in almost all coffee producing countries and capable of long distance dispersal (Brown & Høvmøller, 2002). Unlike other fungal plant diseases, CLR is not necrotic and its symptoms appear only on the underside of the leaves (Belan et al., 2015). The optimum environmental conditions for CLR spore germination are temperatures between 20-25°C, presence of free water through rain, high humidity or dew and low light intensity (Deepak et al., 2012; Vandermeer et al., 2014). It is also known that CLR severity is positively correlated with coffee tree fruit load of the previous season because of the effect the load has on the status of mineral nutrition (Muller et al., 2004; Alves et al., 2009). The disease is spread passively by wind, rain splash, farm workers or vectors mainly from diseased leaves of the preceding campaign that remain on the coffee trees but also from far afield (Avelino et al., 2012; Cristancho et al., 2012).

Current management approaches are preventive and curative fungicides, host plant resistance and biological control (Haddad et al., 2009; Jackson et al., 2012; Kilambo et al., 2013). In the absence of early detection and proper management, CLR results in up to 50% loss of leaves and 70% yield reduction in coffee through premature leaf drop, dieback and debilitation of trees, which will eventually lead to death of coffee plants (Avelino et al., 2004). It is also well documented that CLR does not only affect yield quantity, but also quality, with devastating impacts on farmers' incomes (Ribeyre & Avelina, 2011). One of the key challenges in coffee leaf rust identification is that the characteristic signs of infections only occur on the abaxial leaf side with little to no signs of infection on the upper side. Secondly, CLR infection affects the distribution of nutrients in the coffee leaf (Belan et al., 2015) and thus characteristic reflectance associated with nutrient deficiency such as N can be depicted in CLR infected leaves. In addition, leaf fall appears to be a common problem caused by many of the coffee plant stressors including for CLR.

### **2.3.4 Coffee Berry Disease (CBD)**

The second most important disease of *C. arabica* after CLR is CBD, which is caused by the fungus *Colletorichum kahawae* (J.M.Waller & Bridge). CBD is endemic in Africa (Rutherford & Phiri, 2006). Field studies have shown that CBD is an anthracnose of the green and ripening berries that

result in mummified berries with dark, brown sunken lesions that make berries to be dry, wrinkled, decayed and with a hard skin (Bedimo et al., 2007; Mtenga et al., 2008). It has been established that the pathogen infects berries at all stages of bean development (flowers, pinhead, expanding, mature and ripe stages) although the greatest losses are due to infection of the green expanding berries, between four and six weeks after flowering (Waller et al., 2007; Zeru et al., 2008). Like CLR, the severity and resultant impact of CBD varies between seasons depending on weather conditions as it favours wet conditions and warmer temperatures (Bedimo et al., 2008). It is known that CBD results in considerable yield losses of up to 90% and significantly degrades coffee quality when not adequately controlled (Coste, 1992; Mtenga et al., 2008; Ribeyre & Avelina, 2011). Current management methods are routine calendar preventive fungicidal sprays and curative sprays in affected areas as well as host plant resistance (Mtenga et al., 2008; Van der Vossen & Walyaro, 2009).

### **2.3.5 Coffee Wilt Disease (CWD)**

Another serious threat to *C. arabica* production is CWD caused by *Fusarium xylaroides* Steyaert, the conidial stage of *Gibberella xylarioides* Hem. & Saccas, although so far it has been more significant on Robusta coffee. CWD has devastated millions of coffee trees in central, eastern and southern Africa, costing millions in control and lost earnings in affected countries (Waller et al., 2007; Musoli et al., 2008; Flood, 2009). Unlike other coffee diseases that affect plant parts, CWD kills the whole tree making it more devastating as it takes years for coffee to fully establish and is therefore more difficult to control once it has established (Musoli et al., 2008; Phiri & Baker, 2009). CWD is a vascular disease causing yellowing and wilting of the trees leading to eventual death. It has been observed that the first signs of CWD are yellowing of the leaves on one side of the plant which then wilt, curl, dry up and fall off the branches. Infection and symptoms start on one side of the coffee stem where the vascular bundles become blocked by the interaction of fungal colonization and host responses (Phiri & Baker, 2009; Nair, 2010). Depending on the age of the plant and other factors such as water stress and soil fertility, the period between infection and death of the plant ranges from one month to eighteen months. Current management methods are cultural control, fungicides and host plant resistance but once a tree is infected, there is no remedy other than to uproot the tree and burn it in situ to reduce the chances of spreading the infection (Muller et al., 2004; Waller et al., 2007; Musoli et al., 2008; Phiri & Baker, 2009). Since infection is progressive, there is potential for remote sensing CWD. In addition, since the CWD kills the whole

plant it is relatively easy to apply remote sensing to identify affected areas and fields as this alters the spectral signature of coffee plants and fields. Patterns of disease spread across a field or a region can be monitored from spectral reflectance and vegetation indices.

### **2.3.6 Coffee White Stem Borer (CWB)**

The CWB (*Monochamus leuconotus*, Pascoe) is the most serious pest of coffee in many of the coffee producing countries in Africa with yield losses of up to 25% and its prevalence and severity is increasing from the recent past (Murphy et al., 2008). The presence of CWB is detected by yellowing of leaves, signs of ring barking, exit holes and frass. CWB attack leads to stunted growth, wilting, dieback and reduced yields on affected plants (Logan & Biscoe, 1987; Murphy et al., 2008). The complete life cycle of the CWB can be up to two years with the larval stage taking about 20 months (Logan & Biscoe, 1987; Shoeman et al., 1998). The common management practices for the CWB in Zimbabwe is picking and killing adults as well as uprooting and burning infested plants. According to Kutuwayo (2002a), there is evidence of a kairomonal relationship between nitrogen deficiency and CWB infestation in coffee fields. Since there is selective infestation that depend on other individual tree factors such as N, there is possibility of remote sensing CWB infestation at field levels. Factors such as the spatial patterns in potentially confusing signs such as yellowing of leaves that also occur in nitrogen and plant water stress can be used to separate these. For example, N and water stress affect whole fields or whole blocks while CWB affects specific individual coffee plants in a hedgerow.

### **2.3.7 Other coffee diseases and pests**

Other diseases of Arabica coffee are fusarium bark disease (FBD) caused by *Fusarium lateritium* (Wollen) (telemorph: *Gibberella stilboides*). The characteristic symptoms of this disease are scaling of the bark leading to stem cankers and a progressive dying back of the whole tree (Logan & Biscoe, 1987; Nair, 2010). In addition, the brown eye spot or berry blotch caused by *Cercospora coffeicola* (Berk and Cooke) is also another disease affecting nurseries and plantations. This disease infects the coffee leaves and the fruits, although impact is usually less significant (Waller et al., 2007). Although these diseases are considered minor, there is potential that they can develop to become significant diseases with climate change and changes in cultural management practices (Baker & Haggar, 2007). In addition to diseases, there are also threats from pests of economic significance such as coffee leaf minor (*Leucoptera meyricki* (Ghesquiére)) and antestia bug

(*Antestiopsis orbitalis* (Kirk)) and parasites such as nematodes (Coste, 1992; Kutuywayo, 2002b; Ghini et al., 2011). The plant parts affected, extend of damage on the plant and the remote sensing potential for each of the identified major coffee stressors are described in Table 2.1.

Table 2.1: Summary of major diseases of coffee, parts affected and author perceived potential for remote sensing based on extend and coffee plant parts affected.

Disease/Pest/Condition	Parts affected	Extend of effect	Potential for remote sensing
Water stress	Leaves and stem	Whole plant	Very High
Nutrient deficiency	Leaves	Whole plant	Very High
CLR	Leaves	Few leaves to whole plant	High
CBD	Berries	Few berries to whole plant	Low
CWD	Main stem	Whole plant	Moderate
FBD	Stem	Whole plant	Moderate
<i>Cercospora</i>	Leaves	Few leaves	Low
CWB	Stem and branches	Whole plant	Moderate
Leaf minor	Leaves	Few leaves to many	Low
Antestia bug	Berries	Variable	Very Low
Nematodes	Roots	Whole plant	Moderate

#### 2.4. Current condition assessment methods in coffee crop management

Current coffee condition assessment methods rely on occasional field surveys by teams of specially trained and experienced personnel who use eyeballing and pacing in fields looking for disease signs. However, besides being the largely adopted approach, particularly in resource limited areas, the technique is strenuous especially for large coffee plantations and subjective, with the results being confirmatory, mostly once the disease has fully established. Remote sensing therefore, offers timely and spatially explicit objective assessment of plant condition throughout the coffee growing season (Sankaran et al., 2010; Junior et al., 2015).



## **2.5. Applications of remote sensing in coffee crop management**

There has since been a realisation of the need for remote sensing applications in coffee crop management, and many studies have been done to apply remote sensing in coffee production. These applications can be loosely grouped into mapping cropped areas, determination of coffee physiological properties, crop condition assessments and yield modelling.

### **2.5.1 Mapping cropped areas**

Much of the remote-sensing based research in coffee production has been to map coffee areas to develop spatially explicit inventories of coffee farms and fields for local, regional and national planning. An array of spatial resolutions and sensors ranging from high, medium and low-resolution imagery for this task. For instance, Mukashema et al. (2014) developed an automated approach that utilises an object-based Bayesian classifier on high resolution aerial orthophotos and QuickBird imagery to develop maps of coffee production areas for Rwanda. When this was coupled with a DEM and a forest map, an overall accuracy of 87% was achieved. In a separate comparative study, Ramirez et al. (2006) showed that the IKONOS imagery performed better than Landsat ETM+ in land cover classification in a coffee-dominated landscape owing to its high resolution, with correlations low between reflectance values of the two sensors for most bands. Also using RapidEye and GeoEye high resolution imagery, Johl et al. (2014) obtained thematic maps of coffee areas with an overall accuracy of 80.7% in Tanzania. High classification accuracies are also obtained when high resolution imagery such as aerial photos are coupled with advanced classification methods such as hierarchical boost-classifiers adapted to multi-scale segmentation as done by Dos Santos et al. (2012). It was concluded from this study that the classification approach does not only produce a better classifier which result in higher accuracy (OA=76.7%) but also reduces the training time, making the classification process more efficient.

Most of the studies though, utilised the medium resolution Landsat imagery for coffee cropped area mapping. N'Doume et al. (2000) did a detailed supervised classification of Landsat imagery to map coffee fields in Cote d'Ivoire by categorising fields according to shade and field maintenance, and obtained an overall accuracy of 72% with producer's accuracy of 69% for coffee fields. A study by Moreira et al. (2010) demonstrated that unsupervised classification and visual analysis of Landsat ETM+ image composites can be used to develop accurate thematic maps of coffee farms at municipal level. Alternatively, Bispo et al. (2014) used a computationally intensive

SVM to map crop areas in Minas Gerais in Brazil and reported a low overall classification accuracy of 67% mainly because of the high omission error in the coffee class of 39%. Also, Southworth et al. (2002) observed expansion of coffee production areas in Honduras from parametric classification of Landsat TM imagery. Bolanos (2007), in a study in Colombia, concluded that object based classification resulted in higher accuracies (74.9%) compared to pixel-based classification (71.9%), with producer and user accuracies much higher for sun coffee compared to shaded coffee when using Landsat ETM+ imagery.

In a study in El Salvador, Ortega-Huerta et al. (2012) used Landsat TM and obtained an overall accuracy of 76.7% when classifying coffee according to levels of shades, a difficult task given the complexities and heterogeneity introduced by other trees in influencing the spectral signature of coffee. In a recent study, Chemura and Mutanga (2016) pointed that although these previous studies were important in developing thematic maps for coffee areas as a shift from non-spatial data in planning, it would have been better to develop age-specific thematic maps as age influences management, yield and crop stand and therefore influences the spectral signatures of coffee. They observed that the spectral signatures of coffee are indeed age variant but disaggregating the coffee class into young mature and old coffee classes significantly reduces the accuracy of the thematic maps produced by the random forest (RF) algorithm applied on Landsat 8 OLI imagery (from 90.3% to 86.2%). They however concluded that, the benefit of splitting the age-classes could be more than the accuracy reduction cost in terms of utility of the thematic maps for other purposes such as anomaly detection, disease modelling and yield prediction. Coarse resolution imagery has also been used in coffee area mapping. For instance, Stibig et al. (2007) and Vancutsem et al. (2009) reported mapping coffee areas using multi-temporal SPOT Vegetation composite data as part of mapping perennial croplands in South-East Asia and tropical Africa, respectively. In both cases, results on coffee area mapping were poor mainly due to the coarse resolution of the imagery, indicating that this resolution is not suited for the task.

### **2.5.2 Detection of coffee physiological properties**

Apart from mapping coffee areas, some studies have used remote sensing to characterize the field, tree and fruit physiological properties of coffee. These properties are mainly related to yield quantity and quality and therefore can be used as proxies to plan for field labour for harvesting, expected harvests and their quality attributes. Moreira et al. (2004) used time series spectral

signatures obtained from Landsat TM and ETM+ of coffee to identify changes in coffee phenology at landscape scale. They concluded that the spectral behaviour of coffee was variable between the seasons with better differences with other land cover types more pronounced during the dry season. In another study, Campos et al. (2005) was able to detect and map scene fractional components in a coffee field such as plants, shadow and soil through some segmentation and correlated the spectral characteristics of each component to Landsat TM bands. They concluded that the fractional components in a coffee field were related to most Landsat TM bands (the highest correlation being between Band 5 and shadows which had a correlation coefficient of -0.97). Furthermore, they also demonstrated that it was possible to classify and map the spatial patterns of scene fractions at field level using fused high resolution QuickBird imagery and therefore providing detailed information about crops in the field while excluding confounding soil background and shadows. A study by Gomez et al. (2010) coupled canopy heterogeneity and tree crown size characteristics obtained from Quickbird with a DEM to predict shaded coffee areas in New Caledonia and reported overall accuracy of 83.2% at regional scale.

Physical characteristics of fields and plants have also been related to remotely sensed spectral values and derived vegetation indices. For example, Vieira et al. (2006) correlated coffee parameters such as height of coffee plants, plant density, vegetative vigour, plant diameter, productivity, proportion ground cover and slope gradient to Landsat TM spectral band 4 (NIR). They observed that a complex spectral response relationship exists between coffee plants and reflectance as only proportion ground cover was significantly related ( $\alpha > 0.05$ ) to reflectance of Landsat band 4 ( $r = 0.61$ ). They therefore concluded that proportion ground cover can be estimated using this band in coffee fields at Landsat spatial resolution of 30m, especially in older coffee with ground cover over 50%. After considering challenges using medium resolution Landsat imagery, Coltri et al. (2013) further developed the idea by relating coffee plant characteristics (fresh weight, dry weight, height, basal circumference, plant area index, and lower canopy circumference) with reflectance and vegetation indices derived from high resolution GeoEye imagery. It was concluded from this study that vegetation indices, as expected, had highest correlations with coffee plant parameters than spectral bands, with the highest being for generalised normalised difference vegetation index (GNDVI) and fresh and dry biomass ( $r = 0.97$  for both). Therefore, high resolution imagery can be used to model coffee plant biophysical properties such as height, fresh biomass and dry biomass.

LAI is one of the plant physiological characteristics that can be estimated using remote sensing approaches. Marcon et al. (2011) successfully developed an approach based on a digital camera to estimate the LAI of coffee using the dimensions of the plant as captured in the width and height model (WHM), producing results that are similar to those from time consuming and fastidious field measurements after adjusting for outliers ( $R^2=0.91$ ). Studies have shown that it is possible to characterize and map ripening of coffee in a field for harvest planning and yield estimation with UAV-mounted digital cameras (Herwitz et al., 2004; Furfaro et al., 2005). They were both able to reliably estimate percentages of coffee crop that were at different ripening stages (under-ripe, ripe and over-ripe), demonstrating the ability of remote sensing to provide information on very specific coffee plant characteristics. In a related study, it has also been shown that coffee quality can be reliably estimated by remotely sensed colour of coffee beans at field levels, providing indications of an important attribute that determines income from coffee (Silva et al., 2014).

### **2.5.3 Coffee condition assessment**

Although relatively fewer, some studies have attempted to assess crop condition of coffee using remote sensing-based approaches. Using in-vitro studies, it was determined that the percentage diseased area for CLR can be accurately estimated using digital image processing techniques of remotely sensing imagery but results are dependent on both sensor quality and algorithms for data processing (Price et al., 1993). It was therefore suggested from this study that sensor systems that are able to provide information at leaf level on the shape, area, and perimeter of individual rust lesions are better suited at that scale to provide better epidemiological models for predicting CLR epidemics and related yield losses. In another study that used a high resolution unmanned aerial vehicles (UAV) system, Herwitz et al. (2002) was able to identify and map areas with invasive weeds in coffee areas, providing an improved way of targeted spraying for affected plants or explaining field level yield variability. Literature has also shown that the selection of data processing algorithms matter in coffee crop condition assessment as random forest (RF) was able to more accurately predict plant water content in coffee leaves ( $R^2= 0.76$ ) than simple multiple linear regression (SMLR) ( $R^2=0.59$ ) and partial least squares regression (PLSR) ( $R^2 =0.56$ ) from field spectra (Chemura & Mutanga, 2015). Given the array of crop conditions affecting coffee and other crops in general, these studies are just too few and show underdevelopment in remote sensing applications not just in coffee but in perennial field crops.

#### **2.5.4 Yield estimation and modelling**

Crop yield estimation is important in agricultural management including in coffee production. In coffee, this is even more important because the local and global market prices of coffee are dependent on the exchange market, which are mainly influenced by projected production of coffee. In this regard, attempts to predict the yield of coffee at farm or regional levels using remote sensing have been reported. For example, Brunsell et al. (2009) converted coarse resolution MODIS NDVI into fractional vegetation cover, derived its lagged cross-correlation to surface precipitation and air temperature data, and successfully correlated the obtained vegetation fractions to coffee yields and their annual trends in Brazil. The reported approach is very attractive in that it is based on high temporal resolution data that is able to show, quantify and explain inter-annual variability in yields of coffee as influenced by climate and coffee physiology. Using similar MODIS data, Bernardes et al. (2012) were able to capture biennial bearing trends in coffee yields using lagged cross-correlations of yields and metrics of NDVI and EVI. They concluded that the best correlations were between variation on yield and variation on vegetation indices during the previous year, indicating that the conditions of coffee as measured by MODIS data this year is related to the yield of the following year ( $r = 0.74$  for minEVI and  $r = 0.68$  for minNDVI).

#### **2.5.5 Platforms and algorithms used in remote sensing applications in coffee**

Most of the reported remote sensing applications in coffee management used satellite data. Landsat data series dominate the list of reported sensors because the majority of the studies have been in cropped area mapping (Table 2.2). It is interesting to note that unmanned aerial vehicles and aerial imageries follow satellites in prevalence of platforms, which indicates that there is great interest in application of technology in the coffee industry. Many of the algorithms that were applied on remote sensing data were parametric statistical methods but recent studies are showing application of non-parametric and machine learning algorithms in remote sensing applications in coffee. The spatial distribution of coffee studies using remote sensing that are publicly available is shown in Figure 2.3.

Table 2.2: Summary of applications of remote sensing technology in coffee crop condition assessments showing the sensor types used, methods, maximum reported accuracy and the references.

Area	Sensors	Methods/Algorithms	Accuracy	References
Mapping cropped areas	Landsat TM	Visual, PCA, statistical		Croome (1989)
	SPOT	Maximum Likelihood	OA=80.0%	N'Doume et al. (2000)
	Landsat TM	Maximum Likelihood	OA=85%	Southworth et al. (2002)
	Ikonos-II, Landsat/ETM+	Visual & statistical differences	r=0.81	Ramirez et al. (2006)
	Landsat ETM+	Object-based classification	OA=74.8%	Bolanos (2007)
	Landsat TM	Unsupervised classification	OA=92.8%	Moreira et al. (2010)
	Landsat TM	ISODATA & hierarchical data clustering	OA=74.7%	Ortega-Huerta et al. (2012)
	Aerial images	Object-based classification	OA= 82.7%	Dos Santos et al. (2012)
	Aerial orthophotos & QuickBird	Bayesian Network	R <sup>2</sup> =0.92	Mukashema et al. (2014)
	GeoEye, RapidEye	Maximum Likelihood	OA=80.7%	Johl et al. (2014)
	SPOT-Vegetation	Unsupervised maximum likelihood	OA=72.0%	Stibig et al. (2007)
	SPOT-Vegetation	ISODATA	OA=76.9%	Vancutsem et al. (2009)
	QuickBird	Neural Network and textual analysis	OA=83.2%	Gomez et al. (2010)
	MODIS	Support Vector Machine	OA=67.0%	Bispo et al. (2014)
	SPOT, Aerial photos	Predictive probability models	OA=88.9%	Hailu et al. (2015)
Landsat 8 OLI	Random Forest	R <sup>2</sup> =0.88	Chemura and Mutanga (2016)	
Physiological properties	Landsat TM and ETM+	Time series analysis	NA	Moreira et al. (2004)
	Landsat ETM+, QuickBird	Histogram thresholds, regression	R <sup>2</sup> =0.95	Campos et al. (2005)
	Digital camera	Otsu thresh-holding, Regression	R <sup>2</sup> = 0.91	Marcon et al. (2011)

	UAV with Hasselblad 555ELD camera	Thresh-holding vegetation indices	NA	Herwitz et al. (2004)
	Landsat TM	Correlation	$r=0.61$	Vieira et al. (2006)
	GeoEye	Correlation and regression	$R^2=0.92$	Coltri et al. (2013)
	Field Spectra	Ward's Method & K-means	$R^2=0.88$	Silva et al. (2014)
Condition assessment	Digital Camera imagery, Planimeter	Correlations	$R^2=0.84$	Price et al. (1993)
	UAV with Hasselblad 555ELD camera	Visual, Correlations	$R^2=0.81$	Herwitz et al. (2002)
	UAV	Leaf-Canopy radiative transport Model	NA	Furfaro et al. (2005)
	UAV with DuncanTech MS3100 camera	Ripeness Index, correlations	$R=0.94$	Johnson et al. (2004)
	Field spectra	SMLR, PLSR, RF	$R^2=0.79$	Chemura and Mutanga (2015)
Yield modelling	MODIS and SPOT 5	Fractional NDVI & lagged correlations	NA	Brunsell et al. (2009)
	MODIS	Wavelet-based filtering & correlations	$R^2=0.74$	Bernardes et al. (2012)

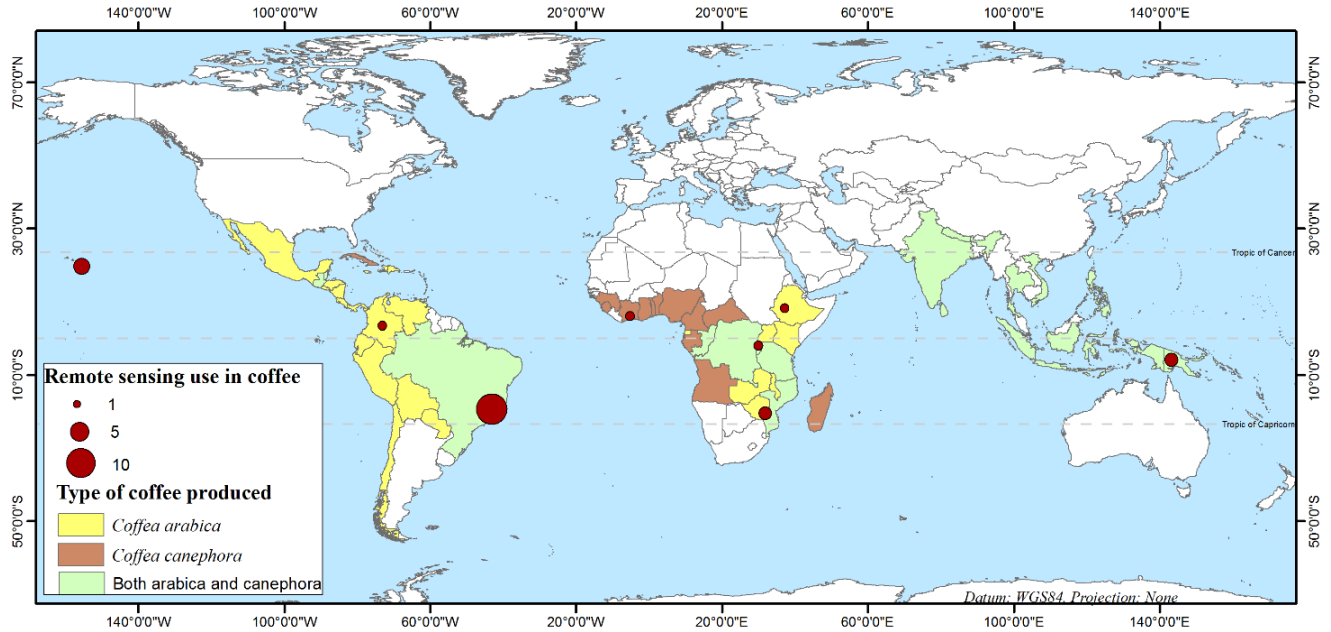


Figure 2.3: Map showing the coffee producing countries, types of coffee produced and the number of papers published on use of remote sensing in coffee production.

## 2.6 Challenges for remote sensing in coffee production and future perspectives

There are a number of physiological, operational and technological challenges that make coffee sensing a challenging task (Section 1.3). The observation that coffee always has a background effect means that spectral scanners will not experience the problem of saturation of the red band that limits application of generic vegetation indices such as NDVI in closed canopy vegetation. However, always having the background effect presents its own challenges for operational remote sensing. There are other challenges reported in coffee area mapping and condition assessment such as influence of age, plants spacing and varieties that limit remote sensing applications (Moreira et al., 2004; Marcon et al., 2011). Dealing with these will require higher spectral resolution datasets such as hyperspectral imagery that are not just expensive, but come with their own new challenges such as dimensionality or the Hugh' curse. Higher spatial resolution sensors can also come handy in dealing with these challenges but these are usually inaccessible in coffee producing areas, are expensive to obtain or maybe limited spectrally.

This review has shown that there has been great interest in mapping coffee areas using remote sensing methods from as early as the 80s (Croome, 1989) and that possibility of doing so has been increasing with improvements in sensor technology as well as with development of robust algorithms to handle the remote sensing data. However, coffee area mapping studies have covered a small fraction of coffee producing countries (Figure 2.3) and given the heterogeneity in coffee systems, terrain and other factors, it is not certain if the algorithms and sensors remain effective across borders. Major issues limiting accuracy reported in these studies are the influence of age of the coffee crop, where young coffee is confused with



bare soil because of the significant soil background influence while old coffee is confused with forest because of the high LAI (Ramirez et al., 2006; Mukashema et al., 2014). Thus, factors such as the health of the crop, shade and the planting arrangement of coffee often in straight rows present further classification challenges (Ramirez et al., 2006; Bolanos, 2007; Bispo et al., 2014). In dealing with variability in coffee ages at landscape scale and associated influence of canopy cover on reflectance, a pre-data handling or masking approach is required. If this pre-masking data can be developed from remote sensing the better, but if not available, it can be obtained from extensive field based methods and farm maps. These pre-masks are vital in dealing with age and other potentially confounding factors in remote sensing-based characterisation of coffee biophysical and biochemical properties related to coffee condition.

Current studies on yield assessment are based on MODIS data. These studies are very important in demonstrating application of remote sensing for direct assessment of coffee yield but the purity of the large 250m x 250m MODIS pixels is a serious issue of concern because of the influence of edges of fields and for intensive land use systems where coffee is just but one of the many options for the farmer. Alternatively, a generic coffee condition map based on field based anomalies can be developed for directing high resolution sensors and new technologies such as UAVs in condition monitoring. This is important as a hierarchical approach to coffee condition assessment where different sensor systems are applied for different but interrelated objectives. Medium resolution remote sensing data such as Landsat and MODIS can be applied for coffee field mapping and anomaly detection while higher resolution datasets such as WorldView, GeoEye and UAVs can be targeted to identify anomalous patches to characterise and quantify the specific cause of the anomaly. In this way, an integrated multi-sensor approach will benefit the coffee condition assessment as each of the sensors have their strengths and limitations.

It was also observed that the spectral fidelity of the imagery is important in mapping coffee areas. Johl et al. (2014) showed that RapidEye imagery (10m), which has the red-edge bands, had the highest overall accuracy (80.7%) compared to GeoEye (1.65m) in mapping coffee areas in Tanzania. The accuracy obtained in mapping coffee areas from high resolution imagery (e.g. IKONOS, aerial photos, QuickBird) is comparable and in many cases less than that obtained from medium resolution sensors such as Landsat (Table 2.2). This may also indicate that local factors such as terrain and heterogeneity influence the accuracy. In addition, classification and modelling algorithms also appear to play a significant role in performance of coffee area and condition assessment. Advanced machine learning algorithms such as Bayesian networks, Random Forests and Neural networks consistently showed high accuracies than maximum likelihood for classifications and linear regression for modelling in coffee. Therefore, the choice of an algorithm is important and future research in coffee condition assessment could benefit from the use

of kernel learning nonparametric machine learning algorithms. These appear more suited to remote sensing data in addition to having high transferability potential (Verrelst et al., 2012).

## **2.7 Conclusion**

Coffee is an important export commodity in international trade and it can benefit from application of remote sensing in condition assessment for decision support. The limiting factors to coffee production such as nutrient deficiency, moisture stress deficit, diseases and pests were discussed in relation to the application of remote sensing. The reported interests and efforts in the application of remote sensing in coffee production were reviewed to analyse achievements and limitations so far. The major challenges to the application of remote sensing in coffee condition assessments that are related to coffee physiological characteristics and production were identified and discussed. It was concluded from this review that there are a number of significant coffee plant conditions that limit yield and quality of coffee and associated economic and environmental benefits. The existence of these stressors concurrently may pose challenges in their remote sensing which makes general assessments such as anomaly detection or estimation of biochemical properties affected by the majority of these stressors a more useful approach. There have been recent developments in multispectral scanners such as Landsat 8 OLI and Sentinel-2 which build on the limitations of predecessor sensors to provide an opportunity to possibly overcome some of the challenges faced in the application of remote sensing in coffee condition assessment. Also related to that are the developments in nonparametric and robust machine learning algorithms that have the potential to further increase the identification and estimation of coffee condition parameters. Future studies should therefore build on the development of sensors and algorithms to necessitate coffee condition assessment with multispectral remote sensing data.

**SECTION 2: SETTING THE STAGE FOR COFFEE CONDITION  
ASSESSMENT**

## CHAPTER 3: DEVELOPMENT OF AN AGE MASK FOR COFFEE CONDITION ASSESSMENT



Photo credits: A. Chemura (2017)

**This chapter is based on:**

**Chemura, A.** and Mutanga, O. (2016) Developing detailed age-specific thematic maps for coffee (*Coffea arabica* L) for heterogeneous agricultural landscapes using random forests applied on Landsat 8 multispectral sensor data, *Geocarto International*, **32**(7), 759-776.

## **Abstract**

Coffee is a commodity of international significance and its production can benefit from spatially explicit age-specific thematic maps. The aim of this study was to evaluate the potential to develop age-specific maps of coffee from Landsat 8 OLI coupled with an intelligent classifier, the random forest. Results showed that the classifier achieved higher overall classification accuracy when coffee was classified as a single class (Scheme B; 90.3% for OLI and 86.8% for ETM+) than the three age-based coffee classes (Scheme A; 86.2% for OLI and 81.0% for ETM+). Comparing the RF mapped area with farm records indicated that the Landsat 8 OLI mapped area closely matched farm records ( $R^2=0.88$ ) compared to that of Landsat 7 ETM+ ( $R^2=0.78$ ). It was concluded that disaggregating coffee classes to produce age-specific maps reduce overall accuracy, but the usefulness of a thematic map with age-specific classes is more than the value of the marginal decrease in accuracy.

**Keywords:** intra-class discrimination, coffee, age mapping, Landsat 8 OLI, random forest

### **3.1 Introduction**

Despite the importance and wide distribution of coffee production, reliable spatially explicit information on where, how much and by whom coffee is produced is lacking in many producer countries. Such datasets are imperative for management decisions for various stakeholders operating at different spatial scales. Most of the current data on coffee production in many countries is on coffee marketing inventories where producers are logged at processing mill level or at washing stations. In some countries, agricultural extension and central statistical agencies collect information on agricultural commodity production, usually at administrative unit level (Mukashema et al., 2014). In addition to this information being scarce, unreliable and expensive to collect, it lacks spatial attributes and therefore maybe limited in its application.

Realising the need for spatial information for the coffee sector, there have been various attempts on mapping coffee production zones in some producer countries. Some studies have characterized and mapped potential coffee production areas using physical and environmental factors (Alves et al., 2006; Devi & Kumar, 2008; Trabaquini et al., 2010; de Carvalho et al., 2013). Campos et al. (2005) concluded that coffee fractional components were related to most Landsat TM bands and demonstrated that it was possible to classify coffee at landscape scale using scene fractional components on QuickBird images. An object-based Bayesian classifier was used to develop a coffee production map for Rwanda that can be useful in agricultural planning and yield prediction (Mukashema et al., 2014). It was also demonstrated that it is possible to estimate coffee yield from phenological behaviour of coffee extracted from remotely sensed data (Moreira et al., 2004; Brunsell et al., 2009; Bernardes et al., 2012). In another study, high resolution IKONOS imagery performed better than Landsat ETM+ in land cover classification in a coffee-dominated landscape (Ramirez et al., 2006). Although these studies are positive steps towards development of reliable and spatially explicit coffee inventories, one important missing component is age of the coffee plantations.

The mapping of coffee plantations, disaggregated according to age is an important agricultural management component. Coffee has a gestation period of four years (young coffee) in which production is not profitable. Mature coffee (4-8 years) is the most profitable stage beyond which productivity and profitability decreases. Age-specific thematic maps are required for precision

agricultural planning that ensures optimal resource utilization and management of farm operations such as fertilization and pest control. Since productivity of coffee is age-dependent, age-specific coffee thematic maps are also useful in yield forecasting required for pre-harvesting and post harvesting operational planning. Reliable yield forecasts can then be used for coffee market intelligence necessary for economic planning at local, regional and national levels (Mukashema et al., 2014). This is especially important for coffee, as local prices are determined by activities on the international commodity markets. All this information leads to more efficient farm management. Efficient coffee management results in higher productivity per unit area, while reducing economic and environmental costs of production (Balasundram et al., 2013). As a result, producer countries will be able to meet their production targets, which is important in securing and meeting contractual agreements without expanding area under production, thereby preserving the environment.

There has been some interest in developing remote sensing-based approaches to produce age-specific thematic maps for agricultural and forestry applications, with different levels of success. Some of the methods included directly relating age to reflectance characteristics. For example, McMorro (2001) found strong relationships between oil palm age and reflectance in Landsat NIR bands. Some studies evaluated the potential of remotely sensed plant biophysical parameters such as leaf area index, canopy cover, crown size and shadow lengths that are directly related to age of plantations (Suratman et al., 2004; Ozdemir, 2008; Tan, 2013). These studies have focused on discriminating age in single-crop dominated landscapes such as for oil palm where the degrees of heterogeneity were not much a factor. In addition, many of these studies relied on the use of high resolution imagery such as airborne hyperspectral data, Worldview-2 and IKONOS which have been reported as the most successful in discriminating different aged stands when coupled with advanced methods such as wavelet transform (Ghiyamat et al., 2015), object based image classification (Chemura et al., 2015b) and support vector machines (Kamiran & Sarker, 2014). The high-resolution imagery are however expensive and unavailable for many parts of the developing countries where coffee is produced. These limitations have necessitated reliance on more readily available satellite datasets such as the Landsat series. However, to achieve age-based intra-species discrimination for developing age-specific thematic maps for coffee from medium resolution imagery such as Landsat 8 OLI, a more robust discriminating algorithm is required.

Machine learning algorithms are among the highly performing classifiers in terms of accuracy assessment. These methods are typically nonparametric and thus do not require any assumptions regarding the distribution of input data which are difficult to meet in remotely sensed data (Srivastava et al., 2012; Adelabu et al., 2014). In addition, they are flexible and robust with respect to nonlinear and noisy relations among input features and class labels (Li et al., 2013). The random forest (RF) is among such machine learning methods. For example, Adam et al. (2014) reported high accuracies in classifying age-specific categories for sugarcane and conditions of grassland using RF classification method and the RapidEye imagery.

The aim of this study was therefore to evaluate the potential of the RF classifier and Landsat 8 OLI imagery in discriminating and mapping coffee plantations of different ages in a heterogeneous agricultural landscape in comparison to Landsat 7 ETM+. Precisely, the objectives of this work were to (i) test for spectral separability of three age-based coffee classes using Landsat 8 OLI data, and (ii) identify the Landsat 8 OLI bands that are specific for coffee detection and age separation and (iii) evaluate performance of the RF classifier applied on Landsat 8 OLI data in comparison to Landsat 7 ETM+ for intra-class and inter-class discrimination and mapping of heterogeneous agricultural landscapes to produce age-specific thematic maps for coffee areas. These maps are meant to be detailed so that they can be useful for many applications at landscape scale.

## **3.2 Material and methods**

### **3.2.1 Study area**

The study was conducted in Ward 19 of Chipinge district, which consists of four large coffee estates and smallholder communal coffee farmers. The site is located in the Mossurize sub-catchment, South-east of Zimbabwe between latitude 32° 36'00E and 32° 48'00E, and longitude 20° 20'00S and 20° 33'00S in Chipinge district (Figure 1). The study site represents the current largest coffee producers in Zimbabwe in terms of volume and area. The climate of the area is subtropical with two distinct seasons, divided almost equally between months of the year (October–March is the growing season while April to September is the dry season). Compared to other parts of Zimbabwe, the area receives relatively high mean annual rainfall totals (1200-1300 mm/year) with mean annual temperatures around 22.5°C (Lagerblad, 2010; Nicolin, 2011). Together with deep red clayey soils formed from mafic rocks, the climatic conditions make the area suitable for quality coffee production. As the area is dominated by large scale coffee farms, the coffee production



system is sun-coffee, which means plantations are exclusively grown with coffee and not mixed with shade trees. In Chipinge district, the mean size of coffee farms is 25 ha and is dominated by Catimor varieties due to their resistance to coffee leaf rust (Chemura et al., 2015a).

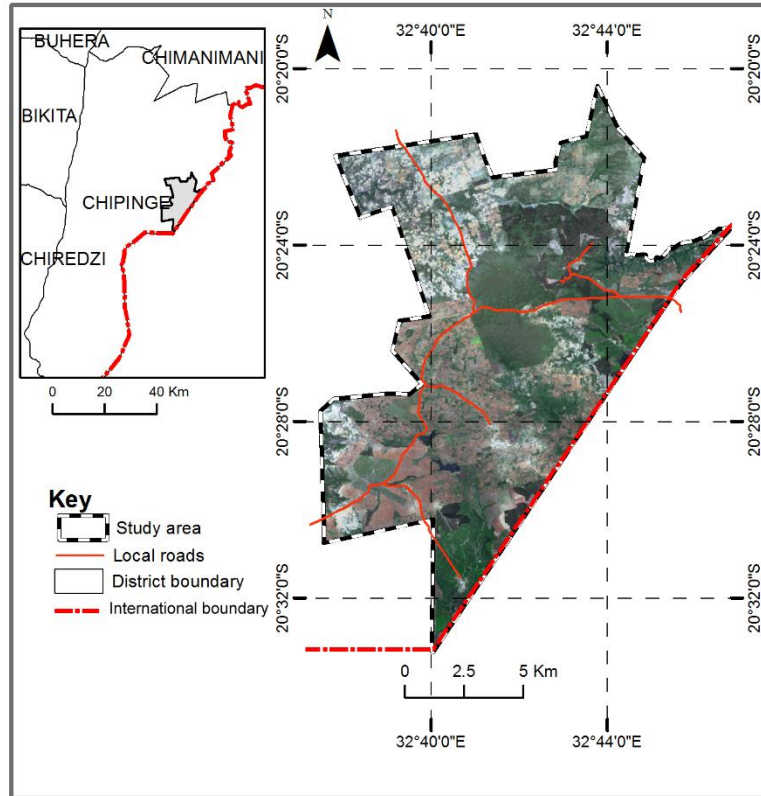


Figure 3.1: Map of the study area showing general landscape and roads. The insert shows the location of the study area in eastern Zimbabwe and in Southern Africa.

### 3.2.2 Field data collection

The training and validation data was collected during a field campaign conducted from the 15<sup>th</sup> to the 16<sup>th</sup> of December 2014. A general landscape map of the area was produced from Google Earth® to guide the fieldwork. A stratified sampling scheme was designed from the Google Earth® imagery to enable proportional sampling of reference points with respect to their sizes. Nine classes as described in Table 3.1 were sampled in the study area. The centre coordinates of the selected locations were recorded using a handheld GPS receiver with an accuracy of ~3m (Garmin® eTrex Vista). Only sample sites that were greater than 1000m<sup>2</sup> were selected.

Table 3.1: Description of the land cover classes used in the classification.

Class	Description
Bare	Areas with open soil mainly due to cultivation.
Forest	Area with tree crown cover of more than 10% covering at least 0.5 ha.
Built-up	Areas with human settlements in high concentrations.
Grassland	Open areas predominantly covered by grass species.
Tea	Plantations with tea plants ( <i>Camellia sinensis</i> ).
Water	Water bodies exceeding 5000m <sup>2</sup>
Young coffee	Areas with coffee plants aged between 1 and 4 years.
Mature coffee	Areas with coffee plants aged between 5 and 8 years.
Old coffee	Areas with coffee plants aged 9 years and above.
All coffee	Sample areas with sites selected from young, mature and old coffee.

### 3.2.3 Image data and pre-processing

One Landsat OLI scene (path 168 and row 74) was obtained for the study area from the USGS-EROS Centre archive ([www.earthexplorer.usgs.gov](http://www.earthexplorer.usgs.gov)). The image acquisition date is the 4<sup>th</sup> of December 2014 at a sun-azimuth angle of 99.7°, a sun elevation angle of 64.38° and was cloud-free. In addition, one Landsat ETM+ image for the same area was acquired and used for comparative analysis. The Landsat ETM+ data was taken on the 11 of November 2014 at sun azimuth angle of 87.39° and sun elevation angle of 64.85° and was cloud free. These images were the available images taken on a date closest to the dates of field data collection. For the purposes of this study, only band 2-7 of Landsat 8 OLI were considered necessary and used for image classification. This is because they have a uniform spatial resolution of 30 m and matched the bands in the Landsat 7 ETM+ data (Table 3.2).

Table 3.2: Spectral and spatial characteristics of the Landsat 8 OLI and Landsat ETM+ data

Landsat 8 OLI				Landsat ETM+			
Band	Name	Bandwidth(μm)	GSD (m)	Band	Name	Bandwidth(μm)	GSD (m)
2	Blue	0.452–0.512	30	1	Blue	0.441–0.514	30
3	Green	0.533–0.590	30	2	Green	0.519–0.601	30
4	Red	0.636–0.673	30	3	Red	0.631–0.692	30
5	NIR	0.851–0.879	30	4	NIR	0.772–0.898	30
6	SW1	1.566–1.651	30	5	SW1	1.547 – 1.749	30
7	SW2	2.107–2.294	30	7	SW2	2.064 – 2.345	30

The two Landsat (OLI and ETM+) scenes were obtained in digital number (DN), and were converted to reflectance values. All Landsat ETM+ images acquired from 2003 have scan-line errors (striping problem) and were corrected for scanline errors using the Landsat toolbox in ArcGIS 10.2 before being converted to reflectance (Walawender et al., 2012). The Landsat ETM+ bands were then converted to Top-Of-Atmosphere (TOA) spectral radiances and then to at-sensor reflectance using the reflectance rescaling coefficients provided in the image's metadata files. Atmospheric correction of Landsat ETM+ and Landsat 8 OLI images to surface reflectance was performed using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) radiative transfer model in ENVI Environment. Ten ground control points collected using a GPS with ~3m accuracy at noticeable places such as road intersections and sharp turns were used for geometric correction. A first order polynomial (Affine) transformation was used for geometric correction of the Landsat imagery and the root mean square error (RMSE) obtained was considered acceptable for the classification process, as it was less than half the pixel dimensions for both Landsat OLI and ETM+.

### **3.2.4 Spectral analysis and characterization**

The spectral profiles of the three coffee age groups were determined by extracting reflectance values from Landsat 8 OLI bands using sample points. To test for the significant differences in spectral reflectance of the three coffee age groups in each band, the non-parametric Kruskal-Wallis one-way analysis of variance was used at  $\alpha < 0.05$  significance level (McKight & Najab, 2010). The Kruskal-Wallis test was chosen because it is known that spectral data seldom follows the normal distribution (Srivastava et al., 2012) and also to deal with the un-balanced number of the samples between classes.

Digitized regions of interest (roi) were used for both classification and accuracy assessment. This is because the sample size used in classification training ( $n=79$ ) and accuracy assessment ( $n=55$ ) was small. Evaluation of the suitability of the training samples to perform classification was done by statistically testing for separability using the transformed divergence separability index. The transformed divergence separability index (TDSI) is calculated as the Jeffries-Matusita (JM) distance squared (Equation 3.1).

$$TDSI = \left( \sqrt{2(1 - e^{-\alpha})} \right)^2 \quad (3.1)$$

where  $\alpha$  is the Bhattacharya distance (Equation 3.2).

$$\alpha = \frac{1}{8} (\mu_i - \mu_j)^T \left( \frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left( \frac{|(C_i + C_j)/2|}{\sqrt{|C_i| |C_j|}} \right) \quad (3.2)$$

where  $i$  and  $j$  are the two signature classes being compared,  $C_i$  is the covariance matrix of signature  $i$ ,  $\mu_i$  is the mean vector of vector  $i$  and  $|C_i|$  is the determinant of  $C_i$  (Asmala, 2012; Dian et al., 2014). In the JM distance, values greater than 1.9 show that spectral classes are highly separable while values less than 1.0 are not statistically different to enable a proper classification (Castillejo-González et al., 2014).

### 3.2.5 Classification schemes

The performance of the RF classifier on Landsat 8 OLI and Landsat 7 ETM+ for intra-species discrimination of coffee was evaluated by running classifications with three age-based classes for coffee and other classes in Scheme A. This classification procedure had nine classes. In Scheme B, the three age-based classes were combined into one class and all other classes remained constant as defined in Table 3.1. In this way, Scheme B classification procedure had seven classes. Scheme B was used as the standard control as it represented the generic approach to running classifications. Both Scheme A and Scheme B were used for classification with the RF classifier on Landsat OLI and Landsat ETM+.

### 3.2.6 Random forest classification

Random forest (RF) is an ensemble machine learning algorithm developed by Breiman (2001) to solve classification and regression problems through a multitude of decision trees. RF employs an iterative bagging (bootstrap aggregation) operation where a number of trees ( $n_{tree}$ ) are independently built using a random subset of samples from the training samples. Each tree is then independently grown to a maximum size based on a bootstrap sample of about two-thirds the training dataset. Each node is then split using the best, among a subset of input variables ( $m_{try}$ ). The ensemble then classifies the data that are not in the trees as out-of-bag (OOB) data, and by averaging the OOB error rates from all trees, the RF algorithm gives an error rate called the OOB

classification error for each input variable. This way, the RF algorithm assesses the importance of each input variable to the outcome by comparing how much the OOB error increases when a variable is removed, while all others are left unchanged (Gislason et al., 2004; Breiman & Cutler, 2007; Adelabu et al., 2013).

In many applications, this algorithm produces one of the best accuracies to date and has important advantages over other techniques in terms of ability to handle highly non-linear data, robustness to noise and tuning simplicity (Caruana & Niculescu-Mizil, 2006; Rodriguez-Galiano et al., 2012; Lebedev et al., 2014; Lu & Weng, 2007). When running the RF, the parameters, *mtry* and *ntree* have to be optimized for improved accuracy (Breiman, 2001; Adelabu et al., 2014). The default number of trees (*ntree*) in EnMAP box is 100 while *mtry* is determined as the square root of the total number of variables used which in this study was seven spectral bands (Breiman, 2001; Liaw & Wiener, 2002). To determine the optimal *ntree*, the random forest model was iteratively ran between 500 and 2500 while assessing the accuracy on training data. An optimal default *ntree* of 500 was used for subsequent analysis as confirmed in literature (Breiman & Cutler, 2007). In the present study, the RF classification algorithm was implemented in EnMAP Box 2.2 (Rabe et al., 2014). The algorithm was applied at the original spatial resolution of 30m after pre-processing.

### **3.2.7 Accuracy assessment**

A confusion matrix was used for accuracy assessment of the thematic maps. A confusion matrix is an empirical estimate of the probabilistic association between remotely sensed and ancillary data versus reference data. To assess the accuracy, the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA) and Kappa coefficient ( $K_c$ ) were used (Congalton, 1991; Liu et al., 2007). OA expresses as percentage, the probability that a pixel is classified correctly by the thematic map and is a measure of the overall classification accuracy (Equation 3.3). PA for a certain class expresses the percentage of a category on the ground that is correctly classified by the classifier, and measures proportion of pixels omitted from a reference class (omission error) (Equation 3.4). UA expresses the proportion of a category on the ground that is included erroneously in another category (commission error) (Equation 3.5) (Congalton, 1991; Foody, 2002).  $K_c$  is a statistical measure of the actual agreement between reference data and classified data versus the chance of agreement between the reference data and a random classifier (Equation 3.6) (Congalton & Green, 1999).  $K_c$  measures the amount of agreement between attributes and

corrects for the expected amount of agreement. If  $K_c$  is one or close to one then there is perfect agreement between the classified map and the reference data. However, limitations of  $K_c$  have been reported especially in terms of giving misleading information and difficulties in interpretation (Pontius et al., 2004; Pontius & Millones, 2011). The quantity disagreement and allocation disagreement measures in addition to the above measures were also used. Quantity disagreement (QD) is the amount of pixels of a class in the training data that is different from the quantity in the test data. Allocation disagreement (AD) is the location of a class of pixels in the training data that is different from the location of the same class in the test data (Pontius & Millones, 2011; Safaa & Pontius, 2012).

$$OA = \frac{1}{N} \sum_{i=1}^r n_{ii} \quad (3.3)$$

$$PA = \frac{n_{ii}}{n_{icol}} \quad (3.4)$$

$$UA = \frac{n_{ii}}{n_{irow}} \quad (3.5)$$

$$Kc = N \sum_{i=1}^r n_{ii} - \sum_{i=1}^r \frac{n_{icol} n_{irow}}{N^2} - \sum_{i=1}^r n_{icol} n_{irow} \quad (3.6)$$

where  $n_{ii}$  is the number of pixels correctly classified in a category,  $N$  is the total number of pixels in a confusion matrix,  $r$  is the number of rows, and  $n_{icol}$  are the column totals representing reference data and  $n_{irow}$  is row total representing the predicted classes (Congalton & Green, 1999; Liu et al., 2007; Petropoulos et al., 2012).

In order to determine the above accuracy matrices, at least 100 pixels from each class (a total of 1532 pixels for Scheme A and 1283 pixels for Scheme B) that were set aside as validation regions of interest developed from fieldwork were used. This represented less than 0.5% of the study area. These validation points were selected from equally representative areas as the training data and away from the training pixels to avoid overlaps between training and validation data. To ensure a fair comparative evaluation of performance of the two schemes, the same set of validation points

were used for classification with Scheme A and B and for Landsat 8 OLI and Landsat ETM+ (Table 3.3). The only differences were in the class ‘coffee’.

Table 3.3: The number of training regions of interest (ROIs) and validation ROIs used for the classification by the two schemes with Landsat 8 OLI and Landsat 7 ETM+. Each ROI represents a field collected sample point.

Class	Training ROIs	Validation ROIs	Total ROIs
Bare	9	6	15
Forest	12	9	21
Built-up	10	7	17
Grassland	8	5	13
Tea	12	9	21
Water	8	5	13
Young coffee	7	5	12
Mature coffee	6	4	10
Old coffee	7	5	12
All coffee	15	12	27

### 3.2.8 Comparing performance of classifiers

The statistical significance of the difference in accuracy performance of the RF classifier on Landsat 8 OLI and Landsat ETM+ was evaluated using the McNemar’s test (Foody, 2004). The McNemar’s test is used in evaluating the superiority of one thematic map over another using the same validation data and when developed from the same training samples. This test is based on the chi-square ( $\chi^2$ ) test computed from a two by two matrix  $F$  (Equation 3.7) based on correctly and incorrectly classified pixels in both classifications (Equation 3.8).

$$F = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \quad [3.7]$$

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad [3.8]$$

where  $f_{11}$  is the number of cases correctly classified by both data one (Landsat 8 OLI) and data two (Landsat 7 ETM+),  $f_{12}$  is the number of cases that are correctly classified in data one (Landsat 8 OLI) but incorrectly classified in data two (Landsat 7 ETM+),  $f_{21}$  is the number of cases that are correctly classified in data two (Landsat 7 ETM+), but wrongly classified in data one (Landsat 8

OLI) while  $f_{22}$  is the number of cases wrongly classified in both datasets. In addition, farm records on area under each age category from three large scale coffee farms in the area were obtained. The area on the farm records was compared with the area estimated to be in each category for the farms to understand how well the RF classifier produce age-maps related to actual field data.

### 3.4 Results

#### 3.4.1 Spectral characteristics of coffee age groups

The differences between the age-based coffee classes are apparent in the spectral profiles of coffee shown in Figure 3.2. The young coffee class had the highest reflectance in all bands except the NIR band where mature coffee class had the highest reflectance. Significant differences ( $p < 0.01$ ) in reflectance between the three coffee classes were obtained from band 3 to band 7. This indicates these bands are important in discriminating the coffee age groups.

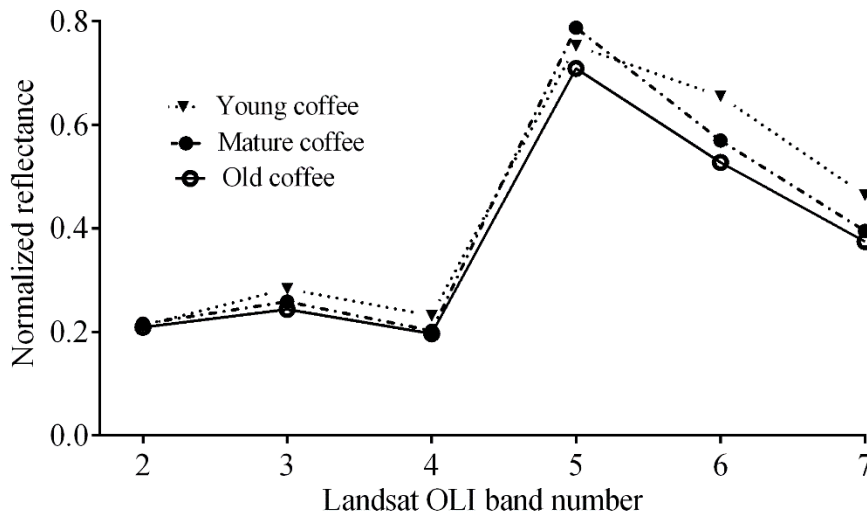


Figure 3.2: Mean reflectance of the coffee classes for the Landsat 8 OLI bands.

#### 3.4.2 Assessment of spectral separability of coffee classes

The transformed divergence spectral separability index showed that the classes were statistically different and could therefore be discriminated using the collected training samples for Landsat 8 OLI and Landsat ETM+. For Landsat 8 OLI, the lowest pairwise transformed divergence separability index was 1.70, which was between mature coffee and old coffee. Mature coffee and forest yielded a score of 1.91 while young coffee and mature coffee had a TDSI of 1.98, which,



when compared to other classes, showed lower separability using the Landsat 8 OLI data (Table 3.4). The highest separability on Landsat 8 OLI data for coffee classes was between young coffee and old coffee (TDSI=1.99). The results for Landsat ETM+ gave the lowest separability as being between mature coffee and old coffee (TDSI=1.61) followed by mature coffee and young coffee (TDSI=1.67). There was also lower separability between in Landsat ETM+ data between old coffee and grassland (TDSI=1.82) and between mature coffee and forest (TDSI=1.89). These results show that spectral separability for most classes was less in Landsat ETM+ than in Landsat 8 OLI data.

Table 3.4: Transformed divergences indices showing pairwise interclass separability of coffee classes and other classes in the training samples.

Class	Landsat 8 OLI			Landsat ETM+		
	Young coffee	Mature coffee	Old coffee	Young coffee	Mature coffee	Old coffee
Water	1.99	2.00	2.00	1.99	2.00	2.00
Forest	1.99	1.91	1.98	1.99	1.89	1.98
Built-up	1.98	1.96	2.00	1.99	1.95	2.00
Grassland	2.00	1.92	1.88	1.99	1.90	1.82
Bare	1.99	2.00	2.00	1.96	1.96	2.00
Tea	2.00	2.00	2.00	2.00	1.99	2.00
Young coffee	-	1.98	1.99	-	1.67	1.97
Mature coffee	1.98	-	1.70	1.67	-	1.61
Old coffee	1.99	1.70	-	1.95	1.61	-

### 3.4.3 Assessing importance of individual bands in RF classification

An evaluation of the importance of image bands in the RF classification showed that the NIR band is the most important parameter for both Scheme A and Scheme B for Landsat 8 OLI while results for Landsat 7 ETM+ were less consistent. Figure 3.3a shows that the accuracy of the RF model drastically drops by 70% on Landsat 8 OLI when the NIR band is not included in running the RF model when coffee is split into three age-based classes. The importance of the Landsat 8 OLI NIR band in the classification however drops by 46.1% when the classification is run with coffee as one compound class (Figure 3.3b). For Landsat ETM+, the SWIR1 band was the most important when classes were split to have age-specific coffee classes (Figure 3.3c) while the NIR band was the most important when a compound coffee class was used (Figure 3.3d).

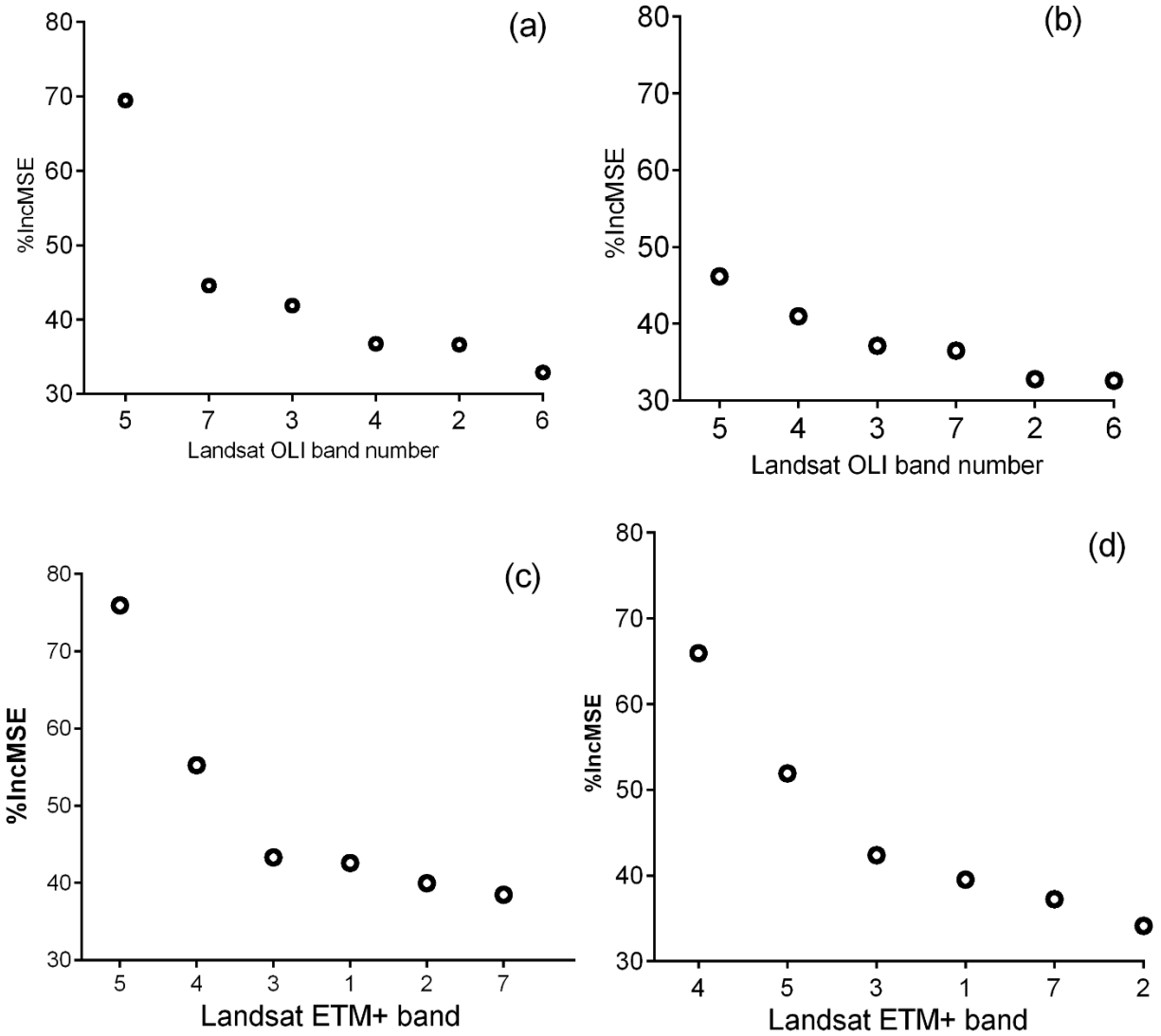


Figure 3.3: Importance of the spectral bands in class separation shown by decrease in training accuracy when the band is excluded for (3.3a) Scheme A samples on Landsat 8 OLI, (3.3b) Scheme B samples on Landsat 8 OLI, (3.3c) Scheme A samples on Landsat ETM+ and (3.3d) Scheme B samples on Landsat 7 ETM+. The highest percentage increase in mean standard error (%IncMSE) corresponds to the decrease in accuracy when that band is excluded and therefore indicates the most important bands.

### 3.4.4 Classification accuracy assessment

Figure 3.4 shows thematic map outputs from RF classifier for scheme A and Scheme B on Landsat OLI and Figure 3.5 shows the thematic maps produced from Scheme A and Scheme B samples on Landsat ETM+. The accuracy metrics showed that the accuracy of the thematic maps was better in Scheme B than in Scheme A. The RF classification achieved an overall accuracy of 90.3 percent in Scheme B when coffee was one class compared to 86.2 percent in Scheme A when coffee was

split into three age-based classes in Landsat 8 OLI. This represents a 4.1% decrease in overall classification accuracy due to splitting the classes in Landsat OLI data. Scheme B also yielded a higher Kc value (0.9) than Scheme A which had 0.8. The allocation disagreement and quantity disagreement increased by 2% when coffee classes were split into three age-based classes, compared to classification of coffee as a single class on Landsat 8 OLI data (Table 3.5).

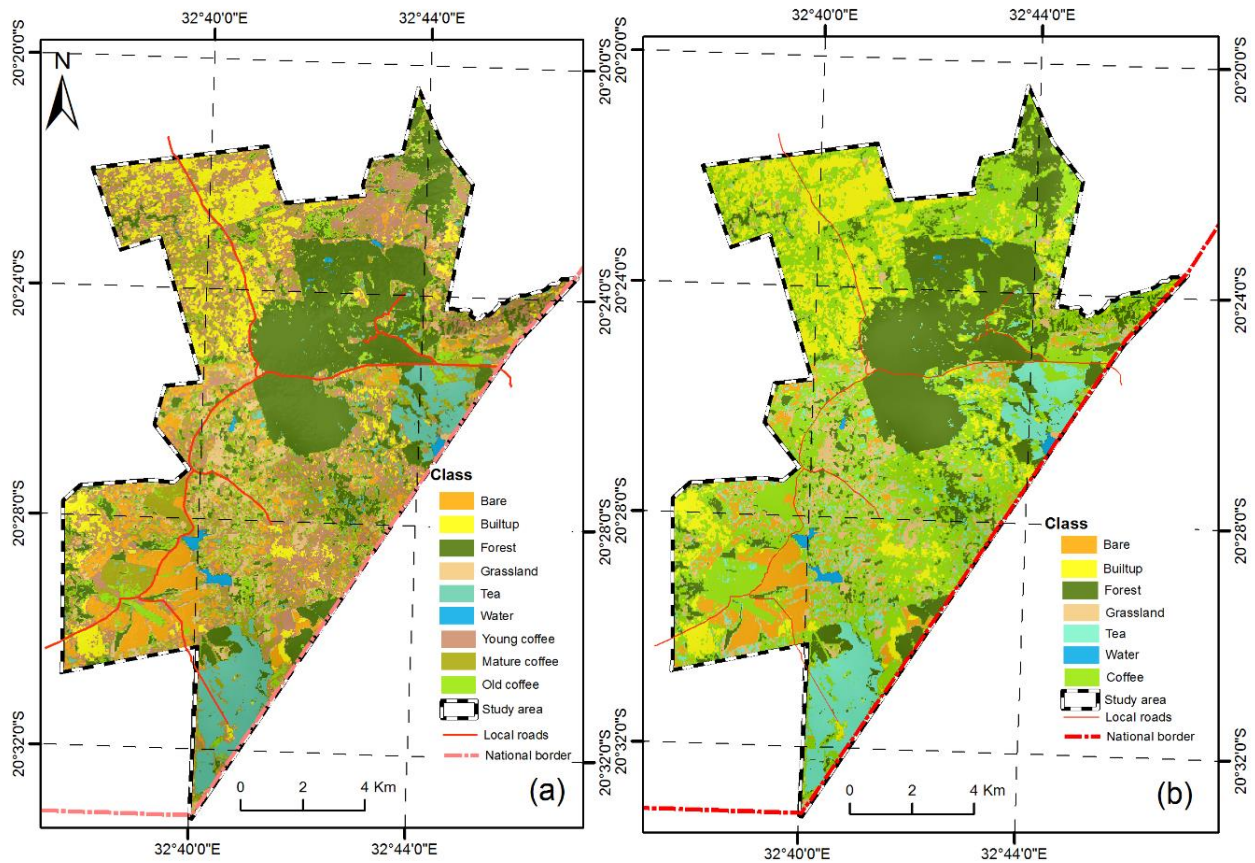


Figure 3.4: Thematic maps produced from random forest classification of the agricultural landscape using Landsat 8 OLI data using (a) Scheme A samples (b) Scheme B samples.

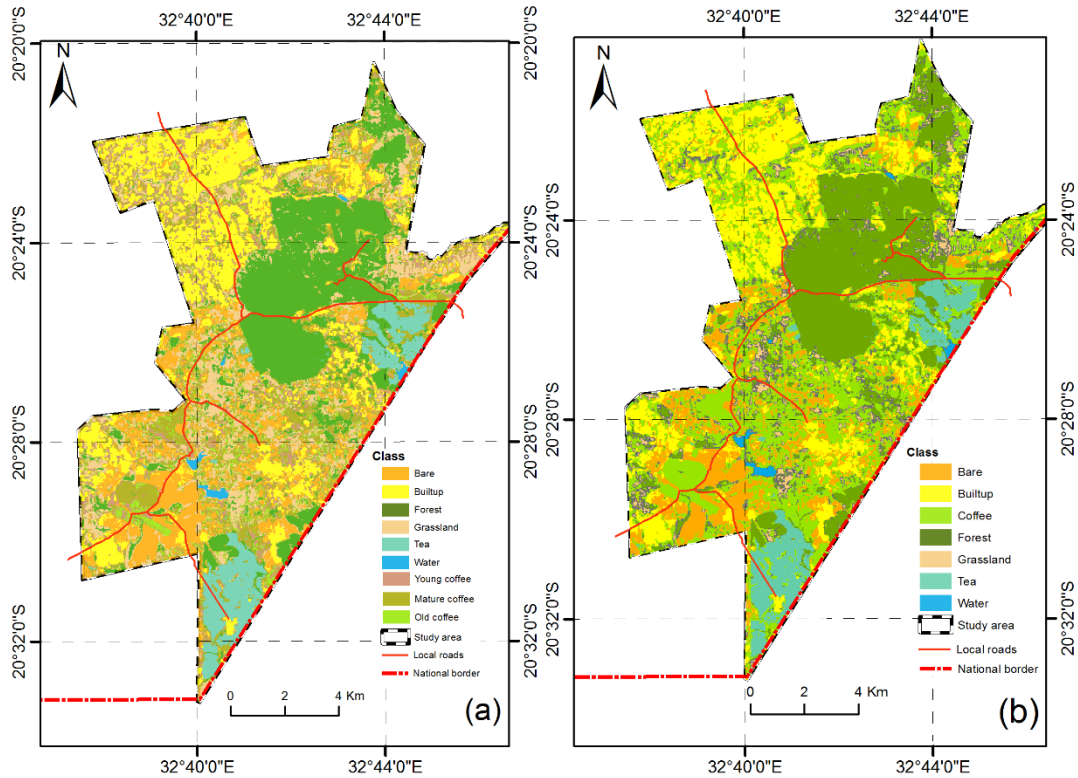


Figure 3.5: Thematic maps produced from random forest classification of the agricultural landscape using Landsat 7 ETM+ data using (a) Scheme A samples and (b) Scheme B samples.

When the classification was run on Landsat ETM+, lower accuracies were realised for both Scheme A and Scheme B, with Scheme B being better than Scheme A (Table 3.5). The accuracy metrics were consistently better for Landsat 8 OLI compared to Landsat ETM+ (higher OA and Kc). The individual class producer's accuracy for Landsat 8 OLI changed only for the forest class (91.9% to 85.2%) due to splitting of the coffee into three age groups, indicating possible spectral confusion with one of the three coffee classes (Table 3.6). The producer's accuracy of water, bare, built-up, tea and grassland remained relatively unchanged after the split of the coffee class in Landsat 8 OLI. Unlike in Landsat 8 OLI, there were remarkable changes in producer accuracies due to splitting the coffee class into three age-based classes. For example, the producer accuracy of tea reduced from 96.5% with classification of coffee as a compound class to 69.4%. On the other hand, the producer accuracy of grassland increased from 67.4% when the compound class was used to 90.2% when split coffee classes were used.

**Table 3.5:** Confusion matrix and associated classification accuracies based on independent test data set for Scheme A produced from Landsat 8 OLI and Landsat 7 ETM+. The accuracies include overall accuracy (OA), kappa (Kc), user’s accuracy (UA) and producer’s accuracy (PA).

Scheme A: Landsat OLI		Reference class										
Classified	Class	Water	Forest	Bare	Built-up	Tea	Grassland	Young	Mature	Old	Total	PA
		Water	94	0	0	0	0	0	0	0	0	94
	Forest	7	190	0	0	1	0	0	0	10	208	85.2
	Bare	0	1	181	0	0	8	4	0	0	194	94.3
	Built-up	0	1	0	213	0	0	9	0	0	223	87.3
	Tea	0	0	1	6	166	0	30	13	0	216	96.5
	Grassland	4	8	1	0	0	158	0	0	0	171	92.4
	Young	0	0	7	20	0	2	126	0	0	155	70
	Mature	0	0	2	4	5	0	8	90	19	128	80.4
	Old	1	23	0	1	0	3	3	9	103	143	78
	Total	106	223	192	244	172	171	180	112	132	1532	
	UA	100	91.4	93.3	95.5	76.9	92.4	81.3	70.3	72.0		
	OA	86.2				Kc	0.8					

Scheme A: Landsat ETM+		Reference class										
Classified	Class	Water	Forest	Bare	Built-up	Tea	Grassland	Young	Mature	Old	Total	PA
		Water	105	0	0	0	0	0	0	0	0	105
	Forest	0	220	0	0	0	1	1	0	10	232	98.7
	Bare	0	0	183	0	0	16	44	0	0	243	95.3
	Built-up	0	0	2	241	2	1	36	0	0	282	98.8
	Tea	0	0	0	0	164	0	0	0	0	164	95.4
	Grassland	1	3	4	0	0	119	2	1	1	131	69.6
	Young	0	0	0	3	0	3	53	6	34	99	29.4
	Mature	0	0	2	0	6	19	39	101	28	195	90.2
	Old	0	0	1	0	0	12	5	4	59	81	44.7
	Total	106	223	192	244	172	153	180	112	132	1532	
	UA	100.0	94.8	75.3	85.5	100.0	90.8	53.5	51.8	72.8		

Table 3.6: Confusion matrix and associated classification accuracies based on independent test data set for Scheme B produced from Landsat 8 OLI and Landsat 7 ETM+. The accuracies include overall accuracy (OA), kappa (Kc), user's accuracy (UA) and producer's accuracy (PA).

Scheme B Landsat 8 OLI		Reference class								
Classified	Class	Water	Forest	Bare	Built-up	Tea	Grassland	Coffee	Total	PA
	Water	94	0	0	0	0	0	0	94	88.7
	Forest	7	205	0	0	1	0	10	223	91.9
	Bare	0	2	181	0	0	10	0	193	94.3
	Built-up	0	1	3	214	0	0	2	220	87.7
	Tea	0	2	1	6	166	0	19	194	96.5
	Grassland	4	7	1	0	0	158	3	173	92.4
	Coffee	1	6	6	24	5	3	141	186	80.6
	Total	106	223	192	244	172	171	175	1283	
	UA	100	91.9	93.8	97.3	85.6	91.3	75.8		
	OA	90.3				Kc	0.89			

Scheme B: Landsat 7 ETM+		Reference class								
Classified	Class	Water	Forest	Bare	Built-up	Tea	Grassland	Coffee	Total	PA
	Water	103	0	0	0	0	1	0	104	97.2
	Forest	2	220	0	0	0	0	0	222	98.7
	Bare	0	0	184	0	0	5	20	209	95.8
	Built-up	0	0	3	242	2	1	0	248	99.2
	Tea	0	0	0	0	166	0	37	203	96.5
	Grassland	0	3	0	0	0	81	0	84	47.4
	Coffee	1	0	5	2	4	83	118	213	67.4
	Total	106	223	192	244	172	171	175	1283	
	UA	100	98.7	88.0	84.9	100	96.4	55.4		
	OA	86.8				Kc	0.84			

In terms of the user’s accuracy, the water, forest, bare, built-up and grassland, classes remained relatively unchanged by splitting the coffee class into three in Landsat 8 OLI data. Notable changes were in the tea class whose accuracy decreased by 8.7% from 85.6% in Scheme B to 76.9% in Scheme A. For Landsat 7 ETM+, user accuracies were better for tea and grassland when the classification was done with the compound while they were better for split coffee classes for built-up and bare, which are spectrally very close classes. Although, there was a decrease in overall and individual class accuracies, the degree of change was not remarkable in Landsat 8 OLI data, which show that RF classifier was able to accurately discriminate, inter and intra-species variation of coffee in a heterogeneous agricultural landscape. This was further supported by the McNemar’s test that showed that the accuracy of Landsat OLI significantly outperformed that of Landsat ETM+ for both Scheme A and Scheme B (Table 3.7).

Table 3.7: McNemar’s test results for comparison between Landsat ETM+ and Landsat 8 OLI for Scheme A and Scheme B classifications\*.

	$f_{11}$	$f_{12}$	$f_{21}$	$f_{22}$	Total	$\chi^2$	$p$ -value
Scheme A	1169	152	76	135	1549	24.67	0.0001338
Scheme B	1069	90	45	79	1283	25.04	5.39E-07

\*Scheme A classes: Water, Forest, Bare, Built-up, Tea, Grassland, Young Coffee, Mature Coffee, and Old Coffee. Scheme B: Same classes as Scheme A except all coffee classes were combined into one class, Coffee.

### 3.4.5 Comparing class area between the classifiers and schemes

In concurrence to the changes in producer and user accuracies, the areas determined for each class by Scheme A closely matched the area determined by Scheme B for bare, water, built-up, and tea when classified using Landsat 8 OLI. Only the forest class (10.3%) and grassland class (15.3%) changed in area by more than 10% in area between Scheme A and Scheme B for Landsat 8 OLI data. The total area predicted as the coffee class increased by 8% when the coffee class was split from one class to three age-based classes by the Landsat 8 OLI data. For Landsat 7 ETM+, splitting coffee into three age-based classes increased area of forest by 3%, of tea by 8 %, bare reduced by 3% while that of coffee remained relatively unchanged. The age-specific thematic maps for coffee are shown in Figure 3.6.



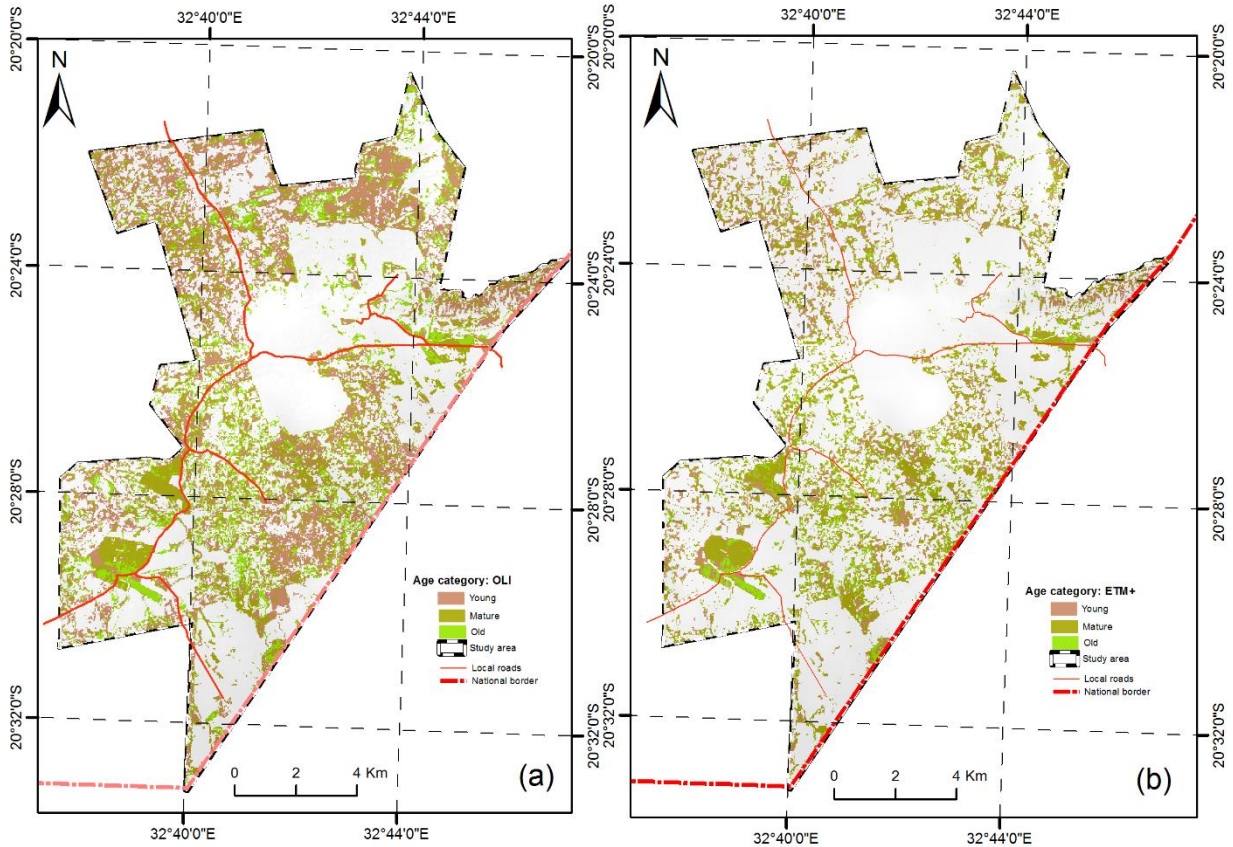


Figure 3.6: Age specific thematic maps for coffee produced from (a) Landsat 8 OLI and (b) Landsat ETM+.

When areas for each class were compared between Landsat 8 OLI and Landsat ETM+, thematic maps from Landsat 8 OLI had more area for young coffee (+19.6%) and mature (+5.9%) and while it was less for old coffee (-10.6%, Table 3.7). The area under coffee classified by Landsat OLI and Landsat 7 ETM+ matched, indicating that differences between the two schemes are apparent in intra-class discrimination.

A comparison of the farm records from three large-scale farms in the study area with the area predicted by the RF classifier using Landsat 8 OLI and Landsat 7 ETM+ was done. Results indicated that the area estimated by Landsat 8 OLI was closer to field farm record of area ( $R^2=0.88$ ) than that of Landsat 7 ETM+ ( $R^2=0.78$ ). However, there was apparent over-estimation of coffee area by the Landsat 8 OLI for all the age categories and farms (Figure 3.7a). This pattern was different from that of Landsat 7 ETM+, which had both over-estimation and under-estimation of areas (Figure 3.7b). The relationship between area estimated by the Landsat 8 OLI and Landsat ETM+ is shown in Figure 3.7c.



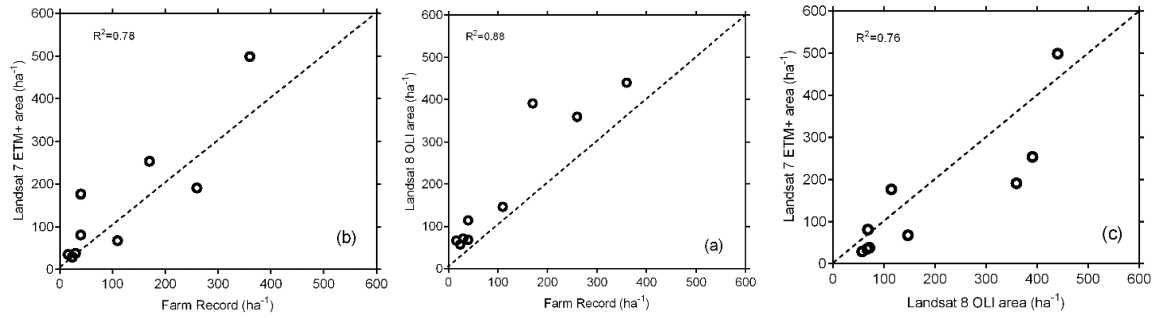


Figure 3.7: Comparison of classifier estimated area with reference data from three farms showing relationships between (a) Landsat 8 OLI data and farm record of area of the three age groups (b) Landsat 8 ETM+ data and farm record of area of the three age groups and (c) Landsat 8 OLI areas and Landsat ETM+ area of the three age groups. The dashed line is the 1:1 line.

Table 3.8: Area ( $\text{ha}^{-1}$ ) for each land cover class as obtained from the random forest classification of Landsat 8 Oli and Landsat 7 ETM+ with Scheme A and Scheme B samples.

	Scheme A( $\text{ha}^{-1}$ )		Scheme B( $\text{ha}^{-1}$ )	
	Landsat 7 ETM+	Landsat 8 OLI	Landsat 7 ETM+	Landsat 8 OLI
Bare	2205	2134	2265	1113
Forest	3866	3520	3755	3970
Built-up	1943	1736	1949	2337
Grassland	1102	1086	1068	1101
Tea	1179	1257	1275	1751
Water	89	91	89	89
Young	2433	3025	-	-
Mature	2178	2314	-	-
Old	1825	1650	-	-
Coffee*	6436	6989	6418	6453

\*The area for class Coffee under Scheme A is the total area of classes young coffee, mature coffee and old coffee.

### 3.5 Discussion

The overall outcome of this study is that it is possible to achieve acceptable accuracies for developing age-specific thematic maps for coffee in a heterogeneous agricultural landscape using Landsat 8 OLI and the RF algorithm. This result provides an opportunity for a cost-effective way of producing reliable age-specific thematic maps that are useful in coffee crop condition assessments and other applications.

#### 3.5.1 Comparison of performance of Landsat 8 OLI and ETM+

The ability to achieve reliable classification accuracy is premised on the significance of the differences in spectral reflectance of the coffee classes of different age categories as well as with other land cover classes. This difference is a factor of the sensor characteristics, particularly its spectral fidelity. The sensor technology in Landsat 8 OLI, particularly the use

of numerous elongated sets of detectors for each waveband, makes it capable of a detailed scan of the surface along track. Along track scanning is known to increase the sensitivity of the sensor to the important biophysical characteristics that determine vegetation reflectance (Dube & Mutanga, 2015). In addition, when compared to the Landsat TM/ETM+, Landsat 8 OLI performed better possibly as it has an enhanced image radiometric resolution of 12 bits which permits more accurate detection of variations in vegetation characteristics (Pahlevan & Schott, 2013; Jia et al., 2014). This, when coupled with Landsat 8 OLI prolonged sensor radiation sampling residence-period for each field-of-view, explain the ability of the RF classifier to produce accurate age-specific thematic maps from Landsat 8 OLI data than on Landsat ETM+.

### **3.5.2 Effect of age on spectral reflectance of coffee**

In this study, it was demonstrated that the spectral reflectance of coffee age categories (young, mature and old) classes are significantly different in the spectral range from the red to the SWIR2 bands of Landsat 8 OLI data. These differences are mostly apparent in the NIR and SWIR1 band, which also showed the greatest uniqueness in spectral characteristics for all age groups. This study therefore underscores the importance of analysing the spectral characteristics of classes in order to discriminate age groups of coffee accurately, and to distinguish these from other land cover classes in order to develop age-specific thematic maps at landscape scale. Furthermore, utilising the random forest inherent variable importance assessments, the results showed the contribution of each band in modelling the land cover classes. This is very important in understanding the usefulness of the Landsat 8 OLI bands in general land cover classification and in vegetation discrimination in particular.

The findings that mature coffee had the highest reflectance in the NIR band was somewhat expected. This is because at mature stage, the coffee is most productive in terms of photosynthetic potential and yield, which is then captured in reflectance in these bands where biochemical properties are least absorptive. The difference between mature coffee and old coffee was not discernible from the visible bands but only in the infrared bands. This confirms the importance of the NIR bands in vegetation characterization particularly in intra-species discrimination. The reduced reflectance of old coffee compared to young and mature coffee could be a result of the normal dieback associated with old coffee. Old coffee also has generally reduced nutrient utilisation and increased leaf fall from plant debilitation (Logan & Biscoe, 1987; Wrigley, 1988). The results indicated a higher degree of spectral overlap between mature coffee and old coffee, and this could be explained by the forward and backward transition of coffee into and out of the mature stage in some samples. The influence of general dieback that

occurs in old coffee was ostensive in the overlaps between the old coffee and young coffee as old coffee droops, thereby having soil background influence similar to that of young coffee. This influence of crop parameters was also reported for coffee by other studies (Vieira et al., 2012).

The spectral profiles showed that there could be significant variations in reflectance of an age group particularly in the infrared bands. This points to the fact that for large-scale seemingly uniform plantation crops such as coffee, there can be variation in reflectance even from the same age groups. These are because of local specific field factors which may confound the age-reflectance relationships utilised by the classifier (McMorrow, 2001). In this particular case, there is therefore need to consider a more robust classifier that is able to better deal with these subtle differences, which is the premise of the use of the kernel random forest classifier in this study. In other studies, the random forest classifier was reported to be able to discriminate between degraded and healthy grassland, between young and mature sugarcane and between inland and coastal sand (Adam et al., 2014). This ability to deal with marginal spectral class changes makes the RF classifier an ideal candidate for the development of age-specific thematic maps from Landsat 8 OLI data.

### **3.5.3 Comparison of accuracy performance**

When all accuracy evaluation metrics are considered (overall accuracy, Kappa, allocation disagreement and quantity disagreement), it is apparent that splitting coffee into three age classes reduces the classification accuracy. However, the reduction in classification accuracy that occur when coffee is split into three classes is marginal (only 4.1% decrease in overall classification accuracy and only 2% increase in allocation and quantity disagreement). This indicates that it could be worthwhile to trade-off some accuracy to obtain a more functional thematic map with age-specific classes in mapping complex agricultural landscapes.

Although Landsat data has been used for age discrimination in landscape vegetation analysis (McMorrow, 2001; Franklin et al., 2003), most successful approaches have involved mostly high spatial and spectral resolution imagery such as Worldview-2 (Chemura et al., 2015b), IKONOS (Thenkabail et al., 2004b) and RapidEye imagery (Adam et al., 2014). This study therefore points to the potential utilisation of enhanced Landsat 8 OLI data combined with clever machine learning classifiers to achieve what has traditionally been in the domain of high-resolution imagery. This high spatial and spectral resolution data is very expensive, unavailable for most areas, suffers from speckle effect and often have high dimensionality (Dalponte et al.,

2009; Mutanga et al., 2012). Some earlier studies had dismissed the potential of thematic maps produced from Landsat 8 OLI data for operational applications because of obvious errors that were associated with the common classification methods at the time (Foody, 2002). Even at the same spatial resolution of 30m, with the improvements in spectral characteristics, coupled by intelligent classifiers, the utility of Landsat data can be further extended to areas that previous studies considered not possible.

### **3.6 Conclusions**

This study tested the utility of Landsat 8 OLI data to discriminate coffee ages in a complex and heterogeneous agricultural landscape using the random forest classifier. The following conclusions were drawn from this study:

1. There are fundamental changes in the spectral profiles of coffee as it grows.
2. There is a marginal increase in class prediction error when coffee classes are separated into three age-based classes (young, mature, and old) as compared to a composite class used in classification.
3. The Landsat 8 OLI data, in combination with the RF classifier, was able to produce satisfactorily compared to Landsat 7 ETM+ accurate age-specific thematic maps for coffee in a heterogeneous agricultural landscape.

### **3.7: Link to next chapter**

This chapter developed age-specific thematic maps using multispectral level Landsat 8 OLI remote sensing data, thereby dealing with age of plants which was identified as one of the most important problems in condition assessment and coffee area mapping. The next chapter utilises the age-specific thematic maps developed from this chapter for identifying areas in coffee plantations that are not growing well.

## CHAPTER 4: IDENTIFICATION AND MAPPING OF ANOMALOUS PATCHES IN COFFEE PLANTATIONS WITH MULTISPECTRAL DATA



Photo credits: A. Chemura (2017)

**This chapter is based on**

**Chemura, A., Mutanga, O. and 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(4), 1–13.**

## **Abstract**

The development of cost-effective, reliable and easy to implement crop condition monitoring methods is urgently required for perennial tree crops such as coffee (*Coffea arabica*), as they are grown over large areas and represent long term and higher levels of investment. These monitoring methods are useful in identifying farm areas that experience poor crop growth, pest infestation, diseases outbreaks and or to monitor response to management interventions. This study compares field level coffee mean NDVI and LSWI anomalies and age-adjusted coffee mean NDVI and LSWI anomalies in identifying and mapping incongruous patches across perennial coffee plantations. To achieve this objective, deviation of coffee pixels from the global coffee mean NDVI and LSWI values of nine sequential Landsat 8 OLI image scenes was first derived. The influence of coffee age class (young, mature and old) on Landsat-scale NDVI and LSWI values was evaluated using a one-way ANOVA and since results showed significant differences, NDVI and LSWI anomalies were adjusted for age-class. The cumulative inverse distribution function ( $\alpha \leq 0.05$ ) was then used to identify fields and within field areas with excessive deviation of NDVI and LSWI from the global and the age-expected mean for each of the Landsat 8 OLI scene dates spanning three seasons. Results from accuracy assessment indicated that it was possible to separate incongruous and healthy patches using these anomalies and that using NDVI performed better than using LSWI for both global and age-adjusted mean anomalies. Using the age-adjusted anomalies performed better in separating incongruous and healthy patches than using the global mean for both NDVI (Overall accuracy=77.2% and 66.4% respectively) and for LSWI (Overall accuracy=66% and 49.2% respectively). When applied to other Landsat 8 OLI scenes, the results showed that the proportions of coffee fields that were modelled incongruent decreased with time for the young age category and while it increased for the mature and old age classes with time. It was concluded that the method could be useful for the identification of anomalous patches using Landsat scale time series data to monitor large coffee plantations and provide an indication of areas requiring particular field attention.

**Keywords:** Monitoring, multi-temporal Landsat NDVI, age, coffee, crop condition

## 4.1 Introduction

Perennial tree crops are important for economic development, food security and alternative energy development, particularly in the developing world, where they are mostly produced (Ghini et al., 2011; Lin, 2011). Coffee (*Coffea arabica* L), an evergreen, perennial tree crop that typically grows up to 4m tall under a variety of production systems in tropical montane areas, is one such important crop (Ortega-Huerta et al., 2012). As part of perennial agricultural systems, coffee plantations protect water resources, improve soil quality, buffer floods, sequester carbon dioxide and provide jobs for millions of people (Omont et al., 2006; Kahn et al., 2011; Dixon & Garrity, 2014). Coffee also represents a long-term capital investment as it is in production for longer periods and therefore requires a robust, reliable and cost-effective monitoring strategy for diseases, pests, water stress, soil fertility and other crop stressors, to safeguard investments and other related ecosystem services. So far, there is however, a general paucity of fine-scale long-term datasets to monitor plantation tree crop stressors (Jeger & Pautasso, 2008). These types of data are required in tracking fingerprints of inter and intra-annual variations in crop conditions useful for plantation managers, investors, insurers and other stakeholders interested in monitoring crop performance. The current monitoring methods largely rely on spontaneous field inspections and sampling, which are not only labour intensive, but also conclusive once economic damage has been inflicted on the crop.

Remotely sensed data provides valuable opportunities for detailed crop condition assessments using the spectral, spatial and temporal domains of raw image bands or their derived vegetation indices or a combination of both. The normalised difference vegetation index (NDVI) because of its robustness is, for instance, one of the most commonly applied remote sensing indices in assessing plant condition i.e. health, growth and vegetation productivity (Glenn et al., 2008; Ding et al., 2014). The NDVI takes advantage of contrasting effects of vegetation on the red and near infrared portion of the spectrum to provide information of vegetation condition. It is a linear estimator of the fraction of photosynthetically active radiation (fPAR) intercepted by vegetation and thus useful in analysing patterns of net primary productivity (NPP) at different spatial and temporal scales (Wang et al., 2004; Alcaraz-Segura et al., 2009). However, NDVI is known to saturate under closed canopy and to be sensitive to atmospheric conditions and soil background (Xiao et al., 2003). Other vegetation indices such as the Land Surface Water Index (LSWI) and the Enhanced Vegetation Index (EVI) are thus also commonly used in land cover assessments and crop monitoring. The LSWI is especially used because it is more sensitive to

equivalent water thickness as it is generated by ratios of spectral bands that are known to be sensitivity to vegetation, soil moisture and water properties (Xiao et al., 2005; Torbick & Salas, 2014).

Vegetation indices can be used to map vegetation and crop conditions over a large area and to detect these changes over time. This is because if the fPAR intercepted by vegetation for the 'healthy' vegetation condition is estimated, then the departure from this indicates an anomaly. For example, Rulinda et al. (2012), used MODIS NDVI anomalies to detect drought-hit areas in east Africa, by applying threshold values to assess the spatial and temporal transition into drought condition. Funk and Budde (2009), using NDVI time-series data for drought assessment, concluded that smoothing, masking and phenological adjustments provide scale-invariant quantitative and visual drought assessment results that are consistent across sub-national, national and regional spatial aggregations. However, the downside of the commonly used hyper-temporal NDVI data in assessing crop condition is that its spatial resolution is very large (>250m) for localized applications due to the mixed pixel effect (Atzberger, 2013; Rembold et al., 2013). Furthermore, most hyper-temporal sensors, such as MODIS, have wide viewing angles that result in significant Bi-Directional Reflectance Distribution Function (BRDF) (Hansen & Loveland, 2012). This confounds such time-series data making them less attractive for field scale applications.

With the high costs and limited coverage of alternative high-resolution imagery and the aforementioned issues with coarse resolution time series data, research efforts have started shifting towards the utilisation of medium resolution multispectral data such as Landsat data series in mapping crop area and condition assessments. These efforts are meant to enable updating of currently available coarser agricultural inventories (e.g. national data) to localized scales (e.g. fields) that help in better understanding of crop dynamics. For example, Dangwal et al. (2016), used Landsat TM data for monitoring water stress in wheat and the use of Landsat derived vegetation indices yielded promising results (RMSE = 0.12,  $R^2 = 0.65$ ). This concluding remark was also validated by a couple of studies across the globe which showed that Landsat NDVI anomalies can be successfully used for identifying land degradation hotspots in the Eastern Tibetan Plateau, China (Fassnacht et al., 2015) and in Basilicata, Italy (Lanfredi et al., 2015). In a comparative study, Ding et al. (2014) concluded that the use of NDVI values produces the greatest observable heterogeneity in the early stage of crop growth at Landsat spatial resolution of 30m. They also concluded that this significantly decays with increase in the spatial resolution of imagery. This observation is further confirmed by Venteris



et al. (2015), whose work successfully showcased the ability of applying the 26 year Landsat data in identifying and mapping anomalous patches (i.e. abnormal vegetation conditions) in soybean and corn at field level, based on their deviation from the expected NDVI. In addition, Tatsumi et al. (2015) demonstrated that Landsat time series can be used for agricultural crop mapping in homogeneous areas by harnessing spatial and temporal characteristics (Overall accuracy of 81% and Kappa of 0.71).

Many studies have developed remote sensing-based phenological metrics such as onset of greenness, time of peak NDVI, senescence period, among others, for monitoring growth and conditions of annual crops. However, the spectral characteristics and growth behaviour of annual crops are very different from those of perennial tree crops that have cumulative growth characteristics, making remotely-sensed distinctions more difficult (Atzberger, 2013). Therefore, much of the focus of remote sensing-based anomaly detection applications in agriculture have been on annual crops, such as wheat, maize and soybeans and yet these are in the field for a relatively shorter period to make same crop assessments (Wang et al., 2011). In addition, in annual crops, unlike in perennial tree crops, it is difficult to interpret season by season anomalies because of potential differences introduced by planting new cultivars, different planting dates and different management between seasons and years, which can contribute to observed NDVI anomalies.

Consequently, attempts to apply coarse-resolution time series NDVI anomalies to map perennial tree crops, such as coffee have not been as widespread and as successful as in annual crops. For example, Vancutsem et al. (2009) using monthly SPOT Vegetation composite data in mapping perennial croplands in tropical Africa, observed incongruent results with those from reference datasets. In a separate study, Stibig et al. (2007), used the same data to map expanding coffee plantations in southern Asia and concluded that results were not consistent as coffee areas were rather mapped as shrubland. This therefore shows that coarse-resolution hyper-temporal data, such as MODIS and SPOT vegetation are only limited to the discrimination and characterisation of land cover types that explicitly exhibit seasonality and that have very contrasting growth cycles, such as annual crops and deciduous species. Medium resolution Landsat data, given the spectral improvements on Landsat 8 may therefore, provide opportunities for the development of detailed spatial datasets on coffee areas and where sequential data is available, to monitor growth or lack thereof.

There is currently more potential in utilising the Landsat 8 imagery for crop condition assessment than what was possible with TM and ETM+. This is because the Landsat-8 satellite, launched on February 11th, 2013, installed into orbit the multispectral operational land imager (OLI) which has many significant technical improvements when compared to its predecessors. These improvements include a higher signal-to-noise ratio (increased by a factor of at least eight compared to ETM+), a higher radiometric resolution (12 bits compared to 8 bits on ETM+) and better radiometric sensitivity from tens of thousands of highly sensitive detectors (Roy et al., 2014; Dube & Mutanga, 2015). In addition, it has an improved geometric accuracy from an on-board global positioning system and an improved combination of pre-launch, on-board and vicarious calibration procedures for a better sensor fidelity (Roy et al., 2014). The Landsat 8 OLI imagery also has an added coastal band designed for atmospheric correction and this is particularly useful in tropical areas where many perennial tree crops are produced. The Landsat 8 OLI therefore provides further opportunities for detailed and accurate detection and mapping of crop conditions than its predecessors. As a result of these improved characteristics, comparative studies have shown that Landsat 8 OLI derived NDVI has better capability of capturing inherent subtle differences across different land surface characteristics (Roy et al., 2014; Ke et al., 2015), which among other things, is attributable to greater signal-to-noise ratio and the radiometric sensitivity.

Given the improvements on Landsat 8 OLI data, it should be possible to identify anomalies in the structural and functional characteristics of coffee by integrating phenology in order to detect incongruous areas where the crop stand is less productive or lagging behind. Perennial tree crops are in production for many years that can exceed 50 years and age is important in influencing fertility management, weed regimes and yield dynamics of these crops (Bhojaraja et al., 2015). Consequently, the age of the crop has significant effect on reflectance and subsequently on derived vegetation indices because of the decrease in the effect of soil background, changes in tree morphology, increase in productivity, and expansion of leaf area index (LAI) with age (McMorrow, 2001; Moreira et al., 2004). The yield of coffee, for example, is known to start after four years (young gestation period), peak between four and eight years (mature period) and then stagnates or decreases thereafter (Logan & Biscoe, 1987; Nair, 2010). This indicates that the age of the crop influences its spectral behaviour and characteristics and thus influences the potential for remote sensing. With these influences of age on spectral characteristics of plantation crops, age-based adjustment of vegetation indices

anomalies is mostly likely to produce better representation of crops than using the global means.

In this study, it was hypothesized that age-adjusted Landsat 8 NDVI and LSWI anomalies can be used to identify incongruous patches in coffee plantations as part of development and application of spatially explicit geospatial, agronomic and economic tools to advance productivity of perennial tree crops. In addition to most of the past studies on the use of NDVI and LSWI anomalies being on annual crops, they have not demonstrated the applicability of the determined anomalies in farm and field-level management decisions. The aim of this study was therefore to assess the utility of multi-temporal age group adjusted Landsat 8 derived NDVI and LSWI data in detecting and mapping anomalous growth patches in continuously growing agricultural crops of economic importance, using coffee as an example.

## **4.2 Materials and methods**

### **4.2.1 Study area**

This study was conducted in south-eastern Zimbabwe, the country's commercial coffee production hub. The study focussed on three of the country's top coffee producing estates and smallholder farmers in the area. The dominant coffee production system is sun-coffee where plantations are exclusively grown with coffee and not mixed with shade trees (Chemura et al., 2015a). The study area has field geometry of 23 km by 16 km, covering a total of 16,815 ha and of this, 860 ha is under coffee production at different ages (Figure 4.1). The site is located in Mossurize sub-catchment, between longitude 32° 36'00E and 32° 42'00E, and latitude 20° 26'00S and 20° 32'00S in Chipinge district. The region is characterized by a subtropical climate with two distinct seasons, divided almost equally between months of the year (October to March - growing season, while April to September - dry season).

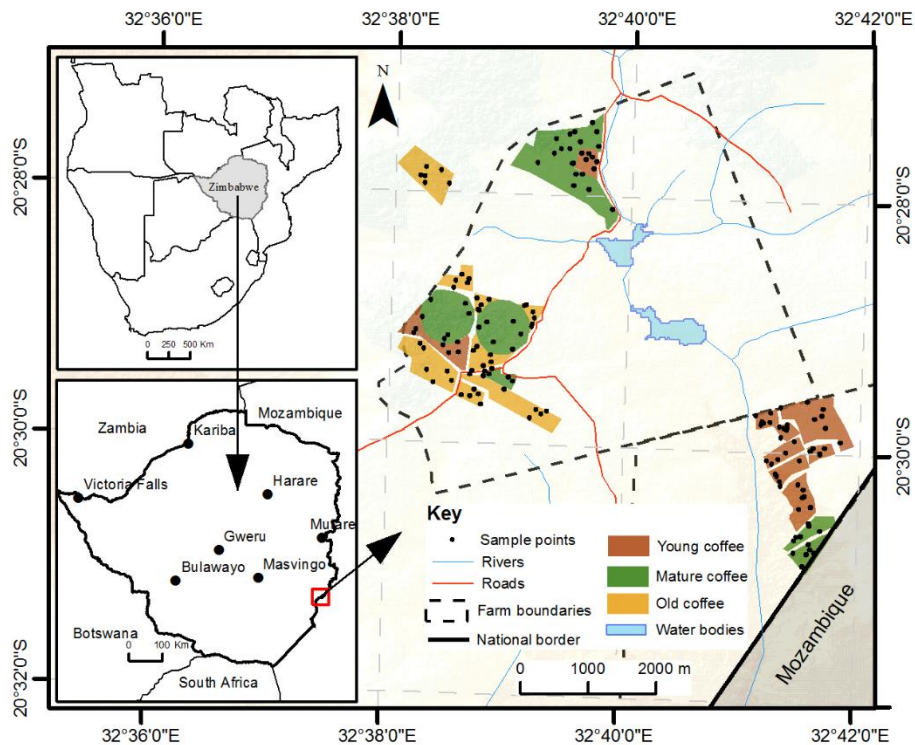


Figure 4.1: Map of the study area showing the distribution of coffee ages used in the study adapted from Chemura and Mutanga (2016). The insert shows the location of the study area in Eastern Zimbabwe.

## 4.2.2 Satellite imagery

### 4.2.2.1 Data acquisition

Nine Landsat OLI scenes (path 168 and row 74) were obtained for the study area from the USGS-EROS Centre archive ([www.earthexplorer.usgs.gov](http://www.earthexplorer.usgs.gov)) covering a period from September 2014 to January 2016. This period covers three contrasting seasons for sufficient monitoring of perennial tree growth. The images were acquired as Level 1 terrain corrected (L1T) products with systematic geometric correction, orthorectification with a digital elevation model (DEM) and precision correction assisted by ground control chips (Roy et al., 2014; Peña & Brenning, 2015). The reference period for data collection was December 2014 when field data was available. Three image scenes were selected for each season (October to March is the rainy season and May to September in the dry season) at a temporal spacing of about a month. A spacing of a month was considered sufficient for temporal monitoring of coffee crop health and to provide opportunities for obtaining images with less cloud cover for practical applications. The dates, day of year (DOY), cloud cover percentage and sun elevation angles of all the image scenes are shown in Table 4.1. Image scenes with the lowest percentage cloud cover over the study area were selected.

Table 4.1: Detail on Landsat 8 Scenes used to derive NDVI

Scene Date	Day of year	Cloud cover (%)	Sun Elevation	Local Season
15 Sept 2014	258	0.00	53.5	Wet
17 Oct 2014	290	7.00	62.5	Wet
04 Dec 2014	338	0.00	64.4	Wet
13 May 2015	133	0.05	41.9	Dry
14 June 15	165	0.28	37.5	Dry
16 July 2015	197	1.70	38.2	Dry
20 Oct 2015	293	7.24	62.9	Wet
23 Dec 2015	357	24.00	102.1	Wet
08 Jan 2016	008	18.00	60.6	Wet

#### 4.2.2.2 Data pre-processing

Since derivation of vegetation indices from satellite imagery is known to be sensitive to atmospheric conditions and also given the diversity of the dates and seasons of the Landsat 8 OLI scenes, atmospheric correction was therefore necessary. Landsat 8 OLI L1T Digital Number (DN) images were first converted to TOA reflectance based on the available Landsat 8 OLI calibration coefficients and standard correction formulas ([http://landsat.usgs.gov/Landsat8\\_Using\\_Product.php](http://landsat.usgs.gov/Landsat8_Using_Product.php)) as described by Chander et al. (2009). These were then used to perform atmospheric correction of the 9 Landsat 8 OLI image scenes using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) radiative transfer model in ENVI 5.1 Software (Exelis Visual Information Solutions, Inc.; Boulder, CO, USA). This method takes into consideration the most likely atmospheric conditions for each DOY. Cloud and cloud shadow contamination were then detected and removed using mask layers generated with the “Fmask” method and code (Zhu & Woodcock, 2012; Zhu et al., 2015). Since the period between the first and the last image is only 18 months, there was no need to account for satellite signal decay over time. In order to ensure that there was a spatial match of the image pixels that were acquired at different dates, further geometric co-registration of the stacked Landsat 8 OLI scenes was performed using ten-ground control points. These were collected using a GPS with ~3m accuracy at noticeable places using first order polynomial (Affine) transformation. Root mean square errors (RMSE) obtained from each scene were checked and were considered acceptable, as they were all less than half the pixel dimensions.

#### 4.2.2.3 Derivation of vegetation indices

NDVI was derived from Landsat 8 OLI band 5 (NIR, 0.851–0.879 $\mu$ m) and band 4 (Red, 0.636–0.673 $\mu$ m) as shown in Equation 4.1 (Rouse et al., 1973; Tucker, 1979). LSWI was derived from band 5 (NIR) and band 6 (SWIR1, 1.57–1.65 $\mu$ m) as shown in Equation 4.2 (Xiao et al., 2005; Torbick & Salas, 2014). Both bands had 30m ground sampling distance (GSD) and a radiometric resolution of 12 bits.

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}} \quad [4.1]$$

$$\text{LSWI} = \frac{\rho_{\text{nir}} - \rho_{\text{swir}}}{\rho_{\text{nir}} + \rho_{\text{swir}}} \quad [4.2]$$

#### 4.2.3 Field data and statistical analysis

Field data was collected during a field campaign conducted between the 15th and 16th of December 2014. During field campaigns, only healthy coffee sample locations were marked using a GPS and their age, variety and other general characteristics were also recorded (N=18). Three age groups which are young (1-4 years), mature (5-8 Years) and old (9 years +) were used for age-adjustment of the anomalies. To increase the number of samples for statistical analysis, the mean and the standard deviation of the 18 sample points were used to obtain pixels that represent healthy coffee as those whose NDVI values lie within two standard deviations of the mean of healthy coffee. Using this approach the number of samples increased to 71 for each age class (N=213). The one-way analysis of variance (ANOVA) was performed in R (R Core Team, 2013) by taking age class as independent variable and pixel values extracted from sample points as dependent (response) variables ( $\alpha \leq 0.05$ ). Coffee NDVI and LSWI values extracted from field sample points (N=213) were first tested for the assumptions of normality using the Shapiro-Wilk test before being subjected to ANOVA. Where means were significantly different, ANOVA was followed by the Tukey pairwise comparison test for mean separation. The three coffee classes were selected because of their significance in determining coffee productivity. There is little to no productivity in the first four years of coffee (young) followed by maximum productivity between four and eight years (mature) and dwindling yield thereafter in the old age group (Logan & Biscoe, 1987; Nair, 2010). These groups also correspond to management requirements for coffee as intensive monitoring is required in the young and the old age groups to ensure productivity.

#### 4.2.4 Anomaly detection and mapping

In order to identify anomalous coffee patches, a method that uses pixel-based difference for identifying incongruous crop areas was adopted (Lanfredi et al., 2015; Venteris et al., 2015). The procedure presented here involves a statistical analysis of anomalies to identify incongruent patches in perennial coffee plantations. The analysis was done by identifying and analysing departures from expected mean values, with and without putting into consideration their age classes. The age map was derived from the map generated by Chemura and Mutanga (2016), Chapter 3. The mean coffee NDVI and LSWI from the sample points were determined and used as the global coffee NDVI and LSWI means on which deviation of all coffee pixels was calculated. The global mean is the average coffee NDVI and LSWI for all the points and was computed for each image scene date. In addition, the average coffee NDVI and LSWI values for each age class were determined from the group means in the ANOVA procedure and used as the age- expected values for a particular age class. Using this age-class based mean, the deviation of each pixel from the age-expected mean for each image scene date was calculated using Equation 4.3.

$$Index_{Diff}(\%) = \frac{X_{iA} - \bar{X}_A}{\bar{X}_A} \cdot 100 \quad [4.3]$$

where  $X_{iA}$  is the pixel NDVI or LSWI value in the image and  $\bar{X}_A$  is the mean NDVI or LSWI for a particular age group where the pixel belongs. This produces the distribution of deviation from the mean of NDVI and LSWI values around 0 for each age class and for each Landsat scene. This transformation is important as it filters the amplitude NDVI and LSWI deviation of each age group into values between 0 and -100 (patches that are less healthy than the normal sampled plots), and +100 (higher than the mean representing those pixels that are healthier than the normal sampled plots) around their age-class mean.

To determine the pixels with excessive NDVI and LSWI deviation, focus was put on the left tail where values were less than the mean. We used the inverse cumulative distribution function (ICDF) to obtain areas that are excessively lower in NDVI and LSWI than the mean for all identified coffee pixels. This was done for both the deviation of the global mean and the age-based anomalies. The ICDF gives the value associated with a specific cumulative probability given the mean and the standard deviation of that data. The cumulative distribution function and the inverse ICDF are related by Equation 4.4 and Equation 4.5. The NDVI and LSWI anomalies were tested for normality using Shapiro-Wilk test before application of the ICDF to ensure that the assumptions of normality are met.

$$p = F(x) \quad [4.4]$$

$$x = F^{-1}(p) \quad [4.5]$$

The threshold probability at 0.05 (outside two standard deviations) was set to be incongruent patches that were not performing well, when compared to their age peers. The area for those pixels was calculated and mapped in ArcGIS 10.3 (ESRI, Redlands, CA, USA). These calculations were done for each age and each scene from September 2014 to January 2016.

#### **4.2.5 Model evaluation**

The approach applied in this study was evaluated using the final binary coffee condition map obtained from one independent image scene. Reference data on coffee condition were collected using visual interpretations of very high resolution (VHR) imagery within Google Earth® as described by (Clark & Aide, 2011). Google Earth VHR is increasingly being used in evaluation of remote sensing approaches particularly for those that involve historical archival data such as in this study (Dong et al., 2014; Landmann & Dubovyk, 2014; Reschke & Hüttig, 2014; Zhou et al., 2016). The sampled VHR imagery (taken on July 10, 2015) was close to the July 16, 2015 Landsat scene and this was therefore considered appropriate for evaluation. The image provider in the study area on which reference data was obtained is Digital Globe. The interpretation operator applied consistent rules including minimum area, observable crop density, crop stand and soil background cover for identification of a pixel as incongruous or normal. Only areas larger 30m x 30m were considered and interpreted in order to match Landsat 8 resolution. The kml reference data points were exported to ArcGIS 10.3, re-projected and overlaid on the final incongruous/normal map and used to extract the condition for statistical analysis. A total of 47 points were used for validation (34 healthy and 13 unhealthy) from the VHR imagery. Using this a confusion matrix from the reference validation data and classified map was developed.

To evaluate model performance, five overall performance (overall accuracy (OA), prevalence, area under the receiver operating curve (AUC), false positive rate (FPR), false negative rate (FNR) and two class specific (producer's and user's accuracy) assessment measures were used (Foody, 2002; Brenning, 2009). The process of identifying healthy and incongruous patches in coffee plantations used in this paper is summarized in Figure 4.2.



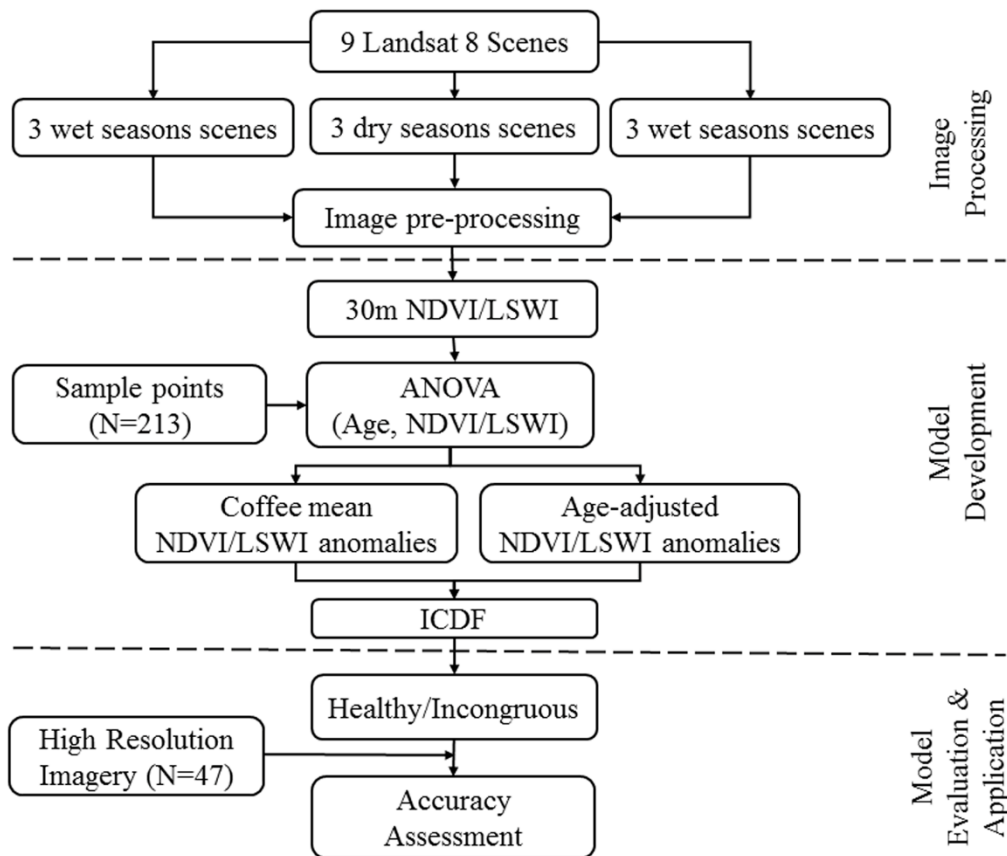


Figure 4.2: Flowchart of the methodology to detect and map incongruous patches in coffee plantations using time series global and age-adjusted Landsat 8 NDVI and LSWI anomalies.

## 4.3 Results

### 4.3.1 Evaluating influence of age classes on vegetation indices

Age class resulted in significant differences ( $p < 0.05$ ) in coffee NDVI values for all the applied nine Landsat OLI scenes ( $p < 0.05$ ), with four of the dates showing highly significant difference ( $p < 0.001$ ) between classes (Figure 4.3a). Results also showed that age resulted in significant differences in LSWI for all image scene dates except one, 15 July 2015. For all scene dates, younger coffee had the lowest NDVI and LSWI values, followed by old coffee, with mature coffee exhibiting the highest NDVI values (Figure 4.3 and Figure 4.4). Although a similar trend was followed, the magnitude of differences in NDVI and LSWI among the three coffee age classes differed with scene date. For example, the difference between mean NDVI of young and mature coffee was quite remarkable for some dates such as 13 May 2015 and 8 January 2016 when compared to other scene dates. Post-hoc pairwise comparisons showed that for some dates there were no significant differences in coffee NDVI means between mature and old (Figure 4.3a). This was more pronounced in LSWI as these two classes were significantly

different for only two dates out of the nine, indicating that LSWI was less able to differentiate between mature and old coffee when compared to NDVI (Figure 4.3b).

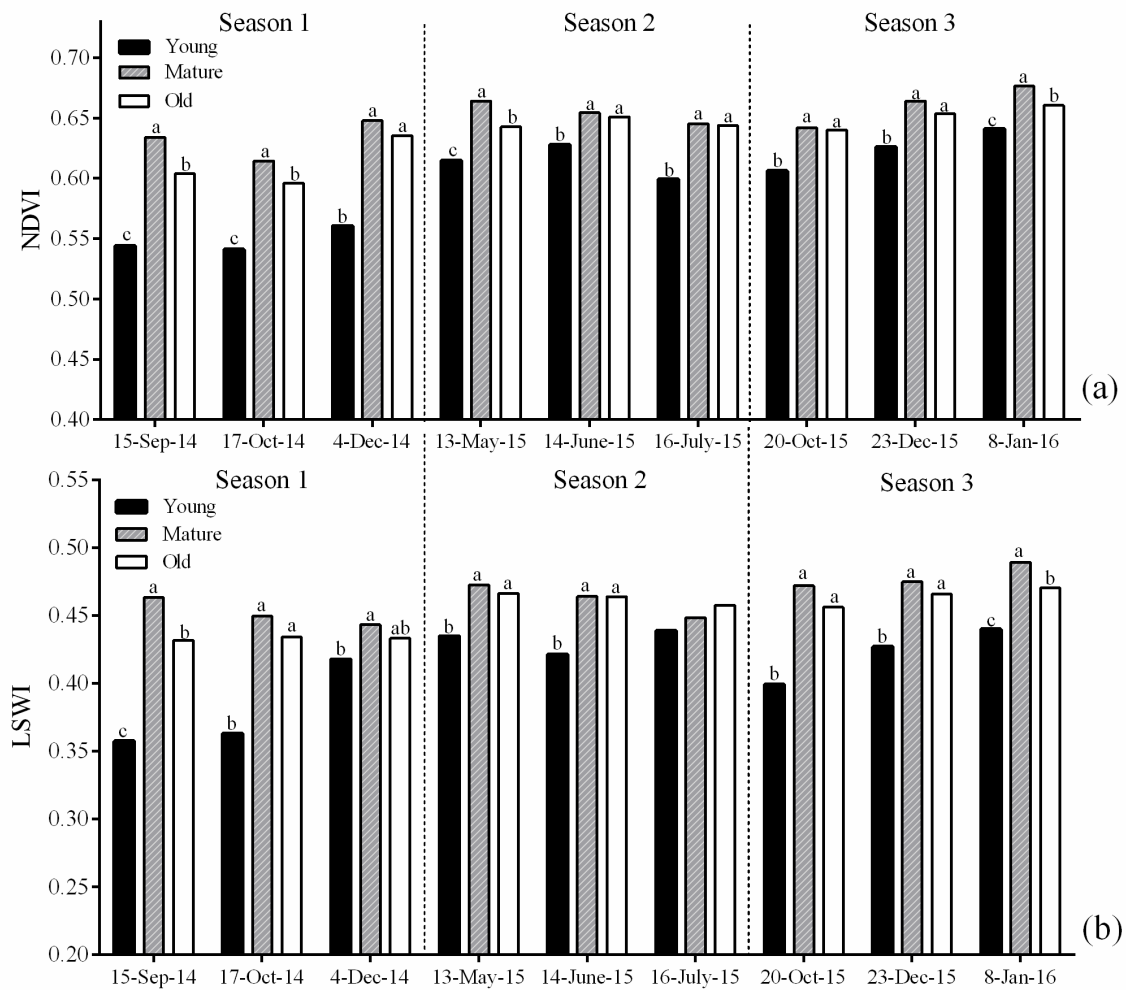


Figure 4.3: Effect of age on mean (a) Landsat 8 Scene NDVI and (b) Landsat 8 Scene LSWI extracted from sample points for all the nine scenes. Means followed by the same letter are not significantly different after the Tukey HSD test ( $\alpha \leq 0.05$ ).

#### 4.3.2 Determining anomalies

The percentage difference of coffee NDVI and LSWI values from the age-expected mean for one image scene is shown in Figure 4.4. The deviation of the pixel values for the mature age class were mostly around zero (Figure 4.4b and 4.4e) indicating that the coffee in this class was more uniform compared to the young (Figure 4.4a and 4.4d) and old age classes (Figure 4.4c and 4.4f). It was therefore possible to identify areas that deviated immensely from their expected global and age-class expected NDVI and LSWI means both positively and negatively. Since it was assumed that negative percentage difference indicate incongruent patches, the focus of the analysis was on red areas on the maps, which had lower NDVI values (Figure

4.5a-e) and LSWI (Figure 4.6a-e) than their expected mean. Between the 15<sup>th</sup> of September 2014 and the 8<sup>th</sup> January 2016, there is evidence that some fields and parts of fields were either getting closer to their age-class or deviating from the mean over time. For example, in the south-eastern part of the maps (Figure 4.5 and Figure 4.6), some young coffee fields began as red, indicating that they were below their age-expected mean (Figure 4.5a-4.5d) but by the end of the 8 Jan 2016 these patches had moved closer to their age-expected means (Figure 4.5e-4.5i).

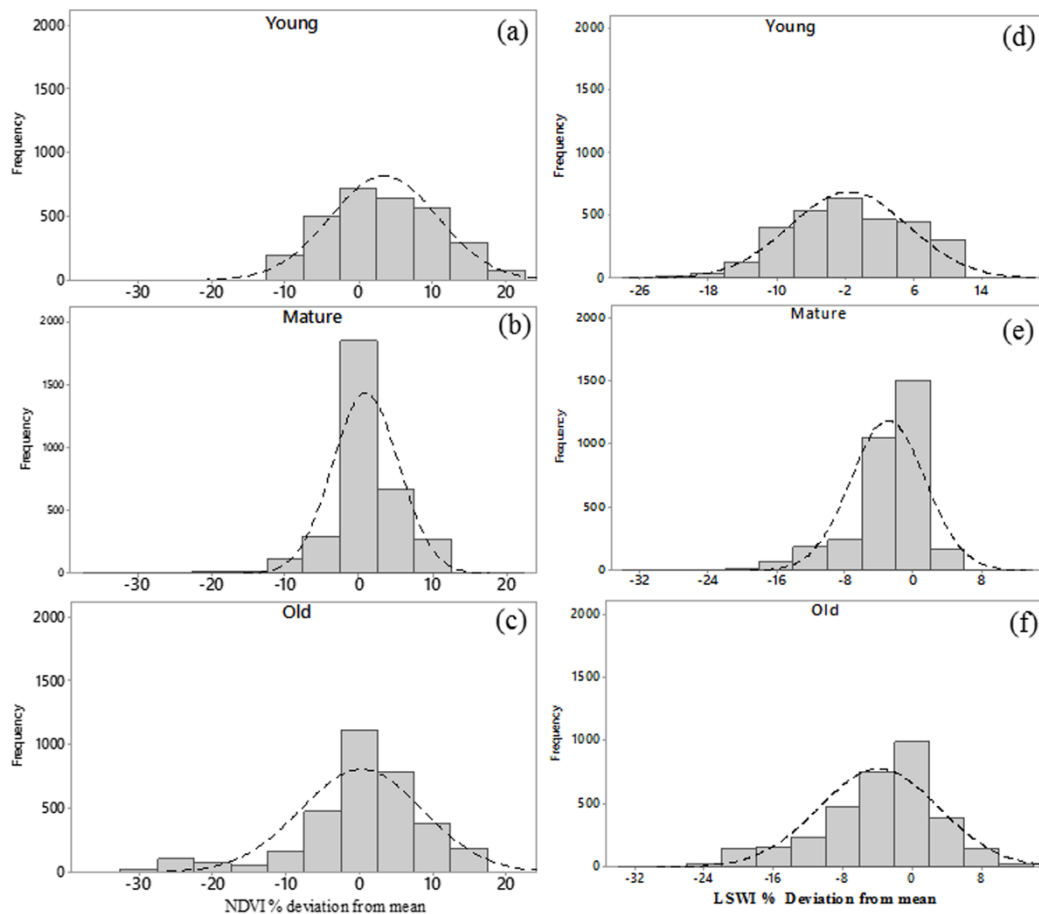


Figure 4.4: Histogram plots of the percentage differences of coffee NDVI (a-c) and LSWI (d-f) values from their age-expected mean values for the 4<sup>th</sup> of December image scene.

The map for NDVI deviation from the age-adjusted means (Figure 4.5) and that of LSWI deviation from the age-expected means (Figure 4.6) showed temporal trends in coffee condition. For example, as from the 16<sup>th</sup> of July 2015, one of the fields in the northern part of the study area deteriorated in condition, and this further extended to neighbouring fields by January 2016. There are more pronounced temporal and spatial variations in deviations from NDVI (Figure 4.5) than those from LSWI (Figure 4.6) indicating that NDVI captured more variations than LSWI.

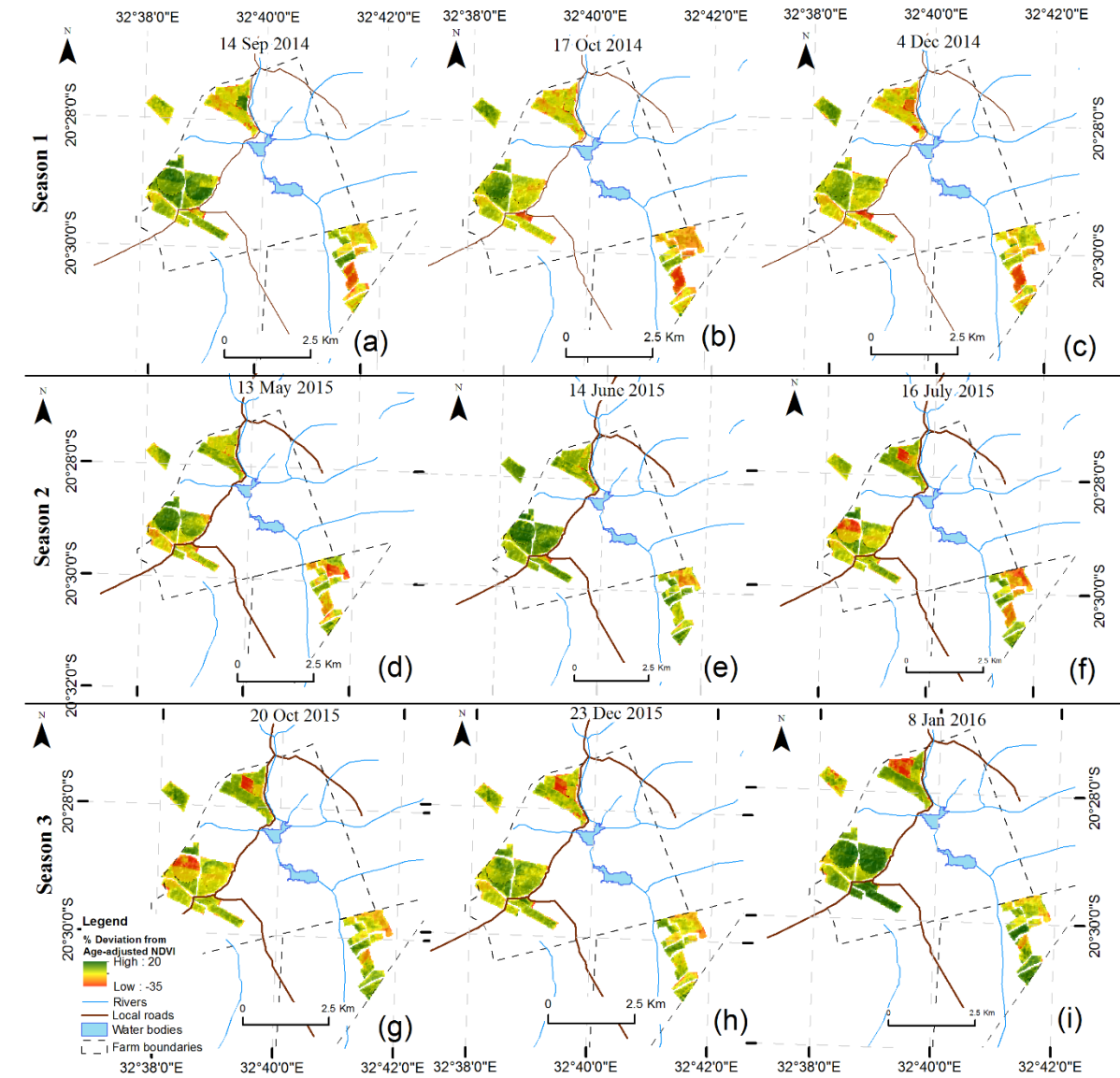


Figure 4.5: Spatial distribution of percentage deviation from age-expected NDVI mean over the study area across the nine Landsat 8 OLI image dates.

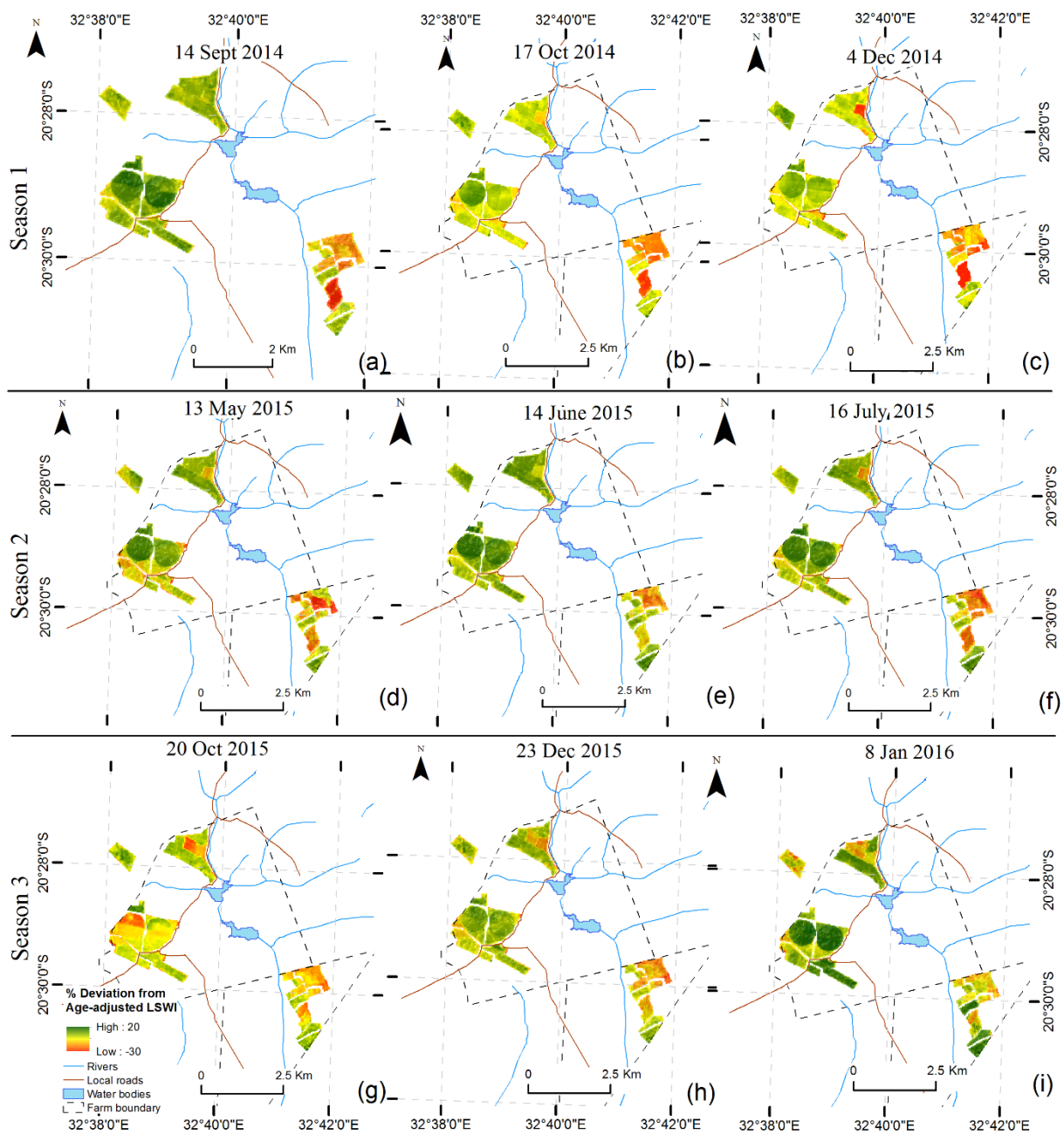


Figure 4.6: Spatial distribution of percentage deviation from age-expected LSWI mean over the study area across the nine Landsat 8 OLI image dates

By applying the ICDF to the distribution curves of percentage difference of NDVI values (global and age-adjusted) and LSWI values (global and age-adjusted), it was possible to identify areas with excessive NDVI anomalies that could be regarded as incongruent for a specified image scene date. These maps were then used to evaluate the accuracy of the mapping and the output from the age-adjusted NDVI anomalies, the best performing in terms of accuracy metrics, is shown in Figure 4.7. As expected, the areas that greatly deviated from the mean

NDVI for their age class on the first three Landsat Image scenes were incongruent (Figure 4.7a-c). From season two onwards, many of these were no longer incongruent as there were now within the range of their age-expected NDVIs for the age adjusted means (Figure 4.7d-i).

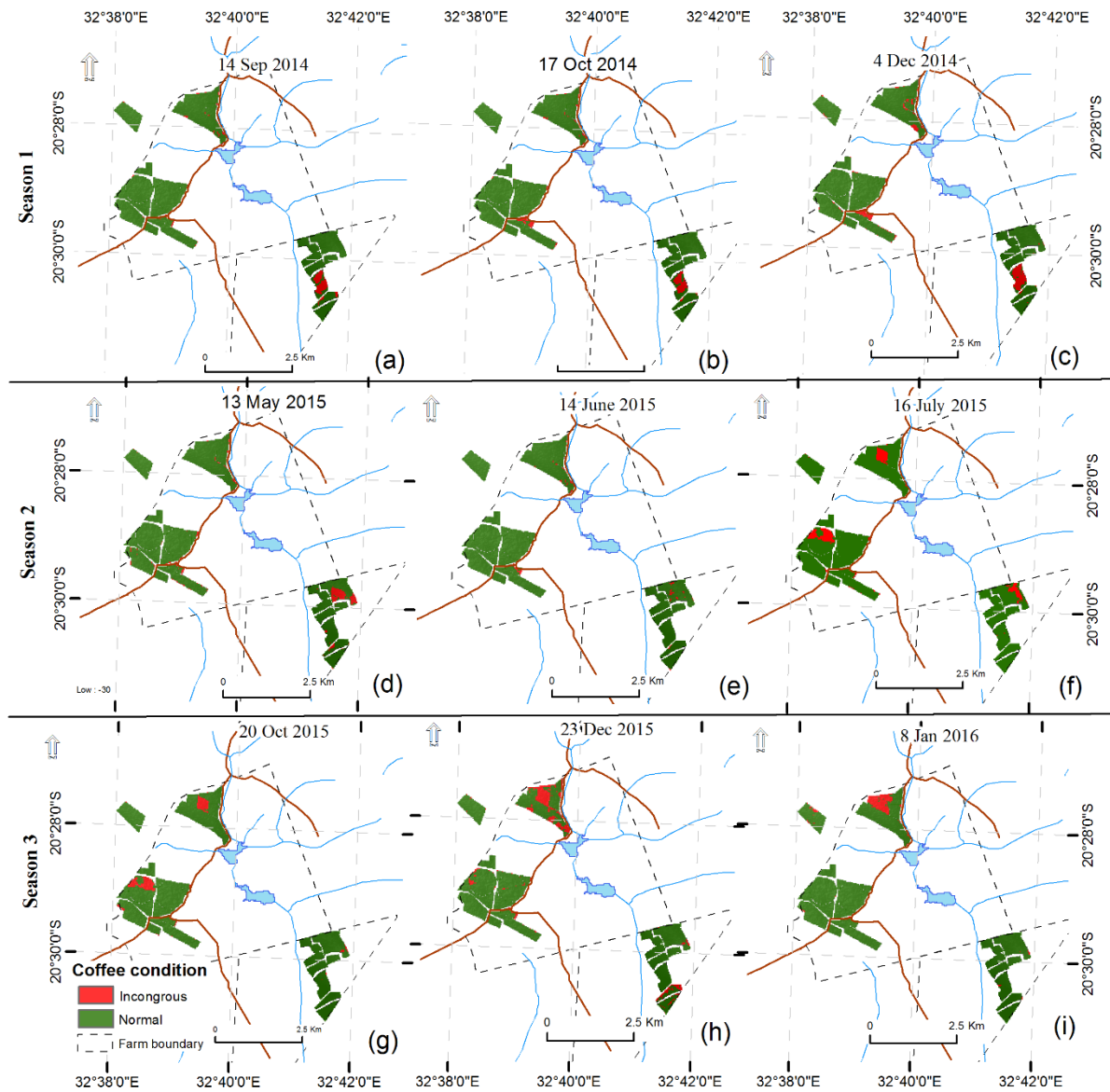


Figure 4.7: Map showing the distribution of incongruent coffee patches because of excessive age-specific NDVI deviation from the mean for the nine Landsat 8 OLI Scenes.

### 4.3.3 Accuracy assessment

Comparison of model determined anomalous and healthy coffee patches with reference data showed plausible model performance for age-adjusted anomalies compared to the anomalies based on the global coffee mean (Table 4.2). Age-adjusted NDVI anomalies had the highest accuracy metrics (overall accuracy, prevalence and AUC) while having lower error metrics (FPR and FNR) compared to using the global coffee mean in identifying incongruous patches



(Table 4.2). Class based accuracy assessment showed that age-adjusted NDVI anomalies had the highest producer's and user's accuracies for normal coffee patches and the highest user's accuracy for incongruous patches (Figure 4.8). Age-adjusted LSWI showed good performance for normal coffee patches in terms of user's and producer's accuracies but had lower user's accuracy for incongruous patches compared to both global NDVI and age-adjusted NDVI.

Table 4.2: Accuracy assessment of model performance

Accuracy Metric	Global NDVI	Age Adjusted NDVI	Global LSWI	Age-Adjusted LSWI
Overall				
Accuracy	66.4	77.2	49.2	66.0
Prevalence	0.72	0.83	0.68	0.85
AUC	0.59	0.67	0.41	0.54
FPR	0.54	0.25	0.80	0.57
FNR	0.26	0.23	0.37	0.30

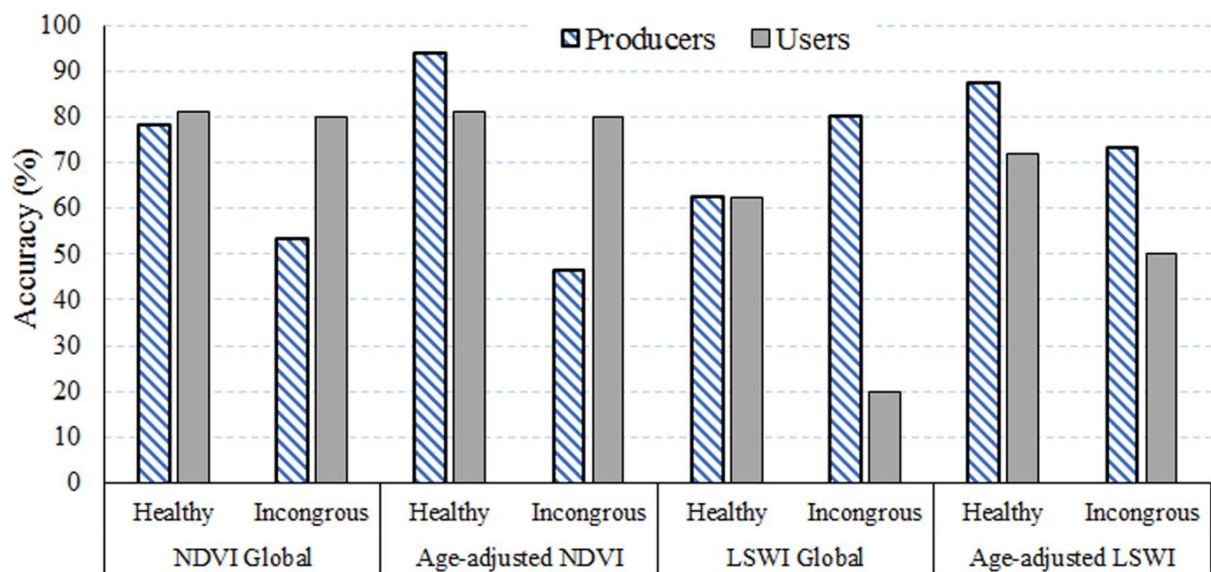


Figure 4.8: Producer's and user's accuracy for normal and incongruous patches for all different types of anomalies.

#### 4.3.4 Quantification of incongruous areas

Given that the age-adjusted NDVI anomalies had the best performance, they were used to determine the location, size and proportion of coffee areas that are performing below their expected conditions at each scene date. The results indicated that the proportion of the coffee

fields that was considered incongruent decreased with time for the young coffee age category while it increased for the mature coffee class (Figure 4.9). From the results of the identified incongruent patches, it was obtained that 56.3ha (6.6%) of the total area under coffee in the study area was excessively below their age-expected NDVI on the 14<sup>th</sup> of September 2014. This had increased to 74.7ha (8.8%) of the area by the 8<sup>th</sup> of January 2016.

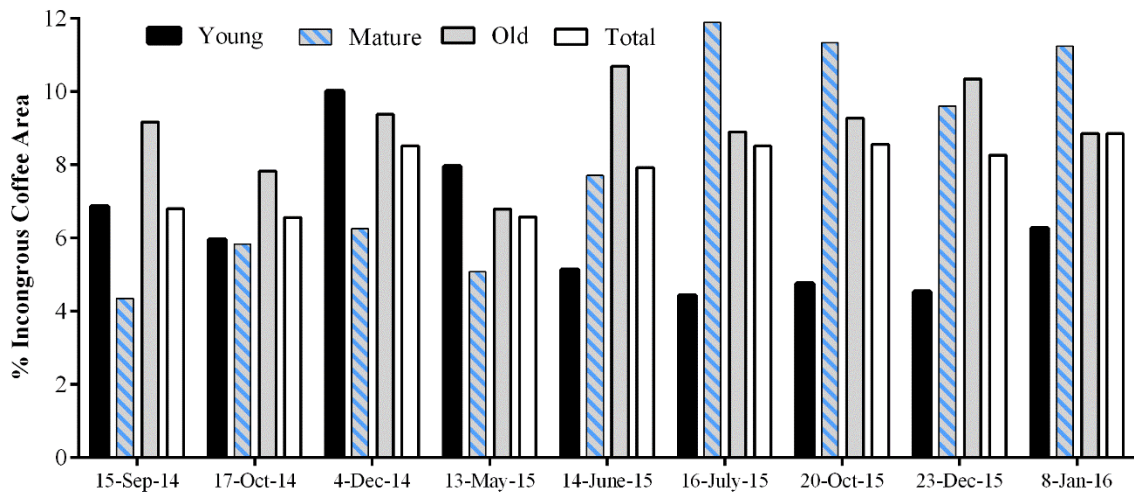


Figure 4.9: Percentage area determined as incongruent from age-adjusted NDVI anomalies for all the nine Landsat 8 OLI image scenes.

#### 4.4 Discussion

The aim of the study was to demonstrate a robust method for evaluating and incorporating age class in identifying incongruent patches in large coffee plantations by using plantations in Eastern Zimbabwe as a case study. To achieve this objective, NDVI and LSWI anomalies with and without age-adjustments derived from the Landsat 8 OLI data were used because of their improved sensor characteristics. These sensor characteristics were perceived to have the capability to improve the identification of healthy and anomalous areas and therefore map spatial heterogeneity of a perennial tree crop by using between and within field NDVI and LSWI variations.

##### 4.4.1 Effect of age on NDVI and LSWI values

It was observed from the study that age significantly influences coffee NDVI and LSWI values. This indicates that there is a transition in terms of spectral signatures of coffee over time, although this is less pronounced between mature and old coffee. The age-classes used in this study have a known influence on various management and productivity aspects of coffee. This is because young coffee is mainly in the gestation period before reasonable yields are achieved, mature coffee represent the most productive and profitable stage and old coffee represent the



drooping stages of the plant when productivity, LAI and photosynthetic efficiency are decreased (Logan & Biscoe, 1987). Moreover, younger coffee have a significant soil background influence, due to open canopies, which explains the observed significantly lower NDVI and LSWI values across all image scene dates (Chemura & Mutanga, 2016). Using high resolution QuickBird imagery, Campos et al. (2005) characterised the fractional components of different coffee fields to explain the observed significant effects of growth stage on NDVI and LSWI.

Landsat 8 NDVI performed better than LSWI in terms of distinguishing between age classes and in terms of accuracy of the mapping of incongruous areas. NDVI is known to correlate well with plant biophysical characteristics that are affected by age, condition and other factors (Ke et al., 2015). It was expected that NDVI will have problems in handling mature coffee which tend to have more biomass per unit area and canopy cover. However, this was not so possibly because coffee and other plantation crops are systematically planted in rows and therefore do not exhibit the dense canopies in natural forests and grasslands or in high density crops where the problems of saturation are reported (Mutanga & Skidmore, 2004; Wang et al., 2004). NDVI also performed better than LSWI for coffee possibly because LSWI is specific to water stress when the conditions of anomalous patches in this study were not only limited to water stress. This is despite the fact that some studies have linked water stress to crop stressors such as pests and diseases infestations, soil nutrient deficiency, and used water-sensitive vegetation indices to map these conditions (Peñuelas et al., 1994; Mutanga & Ismail, 2010; Oumar & Mutanga, 2014).

#### **4.4.2 Remote sensing-based identification of incongruent patches**

Transforming coffee NDVI and LSWI values into age-based deviation proved to be able to show and emphasize extreme patches that are either below or above their expected growth stages. For example, it was very clear how some patches moved from being incongruent to be in the range of their age-expected means for young coffee, indicating growth. This shows that although the age classes are wide (four years), they are able to provide a spatial and quantitative idea of crop performance for every Landsat 8 pixel per scene date. In addition, the results demonstrate that the change in characteristics of a particular pixel either naturally or after management intervention can be monitored using this approach. Although at this stage it is not possible to provide an indication of what is causing particular areas to be incongruous, identifying them provides not only the opportunity for managers to identify areas requiring

attention in large plantations, but provide opportunities for further application of remote sensing methods to determine causes at particular sites. Many hyperspectral, and high spatial and spectral resolution sensors can be focused on these specific areas to determine the cause of the below-average performance. For example, approaches demonstrated by Li et al. (2014) for citrus greening disease detection and Delalieux et al. (2009) for leaf biotic stress of apple plants can be targeted, making these more effective and cheaper. Methods such as linear discriminant analysis on hyperspectral data can then be used to detect levels of infection on specifically identified patches using satellite data, field spectroscopy UAVs or a combination of sensors (Zhang et al., 2012). However, not all pixels identified as incongruent are diseased or necessarily lagging in growth. This means that this method provides a first step in crop condition assessment, which can be followed up either by fieldwork or by other sensing approaches.

Relationships between NDVI and physical characteristics in croplands is confounded by many other factors to be directly determined. Venter et al. (2015), correlated NDVI anomalies of field crops with cumulative crop moisture index (CCMI) and concluded that the relationships were complex. This is possibly because it is not only soil moisture that explains NDVI variation of a pixel and this explains why LSWI did not perform better than NDVI in this study. Therefore, a two-stage process maybe the most appropriate for practical field application of remote sensing crop stress where the first method identifies incongruent patches (as demonstrated in this study) and the other methods identifies and quantifies the causes of the observed incongruence. Higher spatial and spectral resolution data, such as Worldview 2 and Sentinel-2 data could be used in the second stage of identifying what exactly is responsible for the observed departures from age-expected vigour.

Scene-based anomaly detection is more applicable in tropical areas because of problems of clouds and other surface characteristics that are dependent on sharp variations in seasons. For instance, many factors such as dry weather conditions and their effects on LAI explain the lower NDVI during the winter. The natural phenological cycle of coffee and field operations such as harvesting may have contributed to this (Brunsell et al., 2009; Bernardes et al., 2012). A general identification of anomalous areas could be more appropriate in plantation crops where crop conditions are a function of multiple stressors through opportunistic infections. For example, a short period of soil moisture stress could create opportunities for coffee white stem borer (*Monochamus leuconotus*) infestation which predisposes plants to pathogens, such as cercospora leaf spot (*cercospora coffeicola*) among others (Logan & Biscoe, 1987; Nelson,

2008; Kutuywayo et al., 2013). In addition, crop insurance companies can also use this approach to objectively determine compensation for farmers by setting a threshold of farm area determined incongruent.

It was expected that using age-adjusted anomalies would outperform use of the global mean in identification and mapping of incongruous patches in coffee. This is because Bausch (1993) pointed out that when the soil background effect is not factored in NDVI-based crop coefficient determination, there could be estimation errors of magnitudes exceeding 20%. In other studies, it was concluded that regression functions of plant biophysical characteristics with NDVI were significantly different at different growth stages (Sembiring et al., 2000; Freeman et al., 2007). From these it was concluded therefore that growth specific calibration is required when using NDVI for crop condition assessments but then this is impossible with coarse data. This points to the increasing requirement for age-mapping agricultural plantations to feed the information into precision phenotyping and characterisation that are required for many applications (Thenkabail et al., 2004b; Bhojaraja et al., 2015). The challenge is therefore in having age-disaggregated spatial datasets to enable applications of approaches like this. Merely separating a crop from other land use/cover types is still a challenge and thus extending the separation into within class classification brings further challenges. However, with improvements in sensor characteristics and availability of robust methods such as random forests, objected-oriented classification and sub-pixel classification, this is becoming possible (Tan, 2013; Chemura & Mutanga, 2016).

#### **4.4.3 Potential limitations and future improvements**

The approach presented in this study could have some limitations, which provide opportunities for future development. Despite promising results, there are still several limitations of the proposed method that need to be considered. The study is dependent on a pre-determined age-map also produced from Landsat 8 OLI data. An approach that could produce the age-map and concurrently identify anomalies could be easier to implement because not all areas have these age maps. The spatial resolution of the medium resolution Landsat 8 OLI data is large, meaning that it is not possible to identify specific individual plants that are anomalous due to infection, fertility and/or soil moisture deficit. Therefore, there is already considerable spread before the source of the anomalous plant is actually identified by this approach. However, it is known that causes of anomalous crop conditions, such as pest attacks, disease infections and fertility cause considerable damage to the plant system and even spread before any visual signs could be detected even by field methods (Carter & Miller, 1994; Eitel et al., 2011). This therefore means

that an anomalous pixel could actually have a mixture of problems which individual stress identification methods may give the error of omission, making the approach presented here a better option. In addition, since the method is not dependent on field samples, once expected mean values have been established or sample positions to collect them are set, there is huge opportunity for development of automated online platforms and apps for use in coffee monitoring from this approach.

#### **4.5 Conclusions**

In this study, the use of multi-temporal age-adjusted Landsat 8 OLI anomalies to identify and quantify within and between fields spatial heterogeneity of a perennial tree crop for informing management decisions was demonstrated. The following conclusions from this study were made:

- Age class is a significant factor affecting coffee NDVI and LSWI and should be incorporated in using anomalies for identifying incongruent patches.
- NDVI performs better in discriminating coffee age classes compared to LSWI and in accuracy of identifying incongruous patches.
- Age-adjusted anomalies perform better than using anomalies from the global mean in detecting crop conditions for coffee plantations.
- Although this method is promising for long-term monitoring of perennial tree crops which are in the field for decades, further studies are still required to validate the applicability of this method in other plantation crops.

#### **4.6 Link to next chapter**

This chapter demonstrated the ability of multispectral level remote sensing data to identify and quantify anomalous areas in coffee fields using age adjusted NDVI anomalies. The next two chapters experimentally evaluates the ability of the multispectral remote sensing data to identify specifically the cause of the anomalies using case studies of abiotic stressor (plant water stress) and biotic stressor (coffee leaf rust).

**SECTION 3: REMOTE SENSING INDIVIDUAL COFFEE  
PLANT STRESSORS**

## CHAPTER 5: MULTISPECTRAL LEVEL REMOTE SENSING OF PLANT WATER CONTENT IN COFFEE



Photo credits: A. Chemura (2017)

**This chapter is based on:**

1. **Chemura, A.** and Mutanga, O. (2015) Comparing physiological and spectral response of coffee (*Coffea Arabica*) to induced soil moisture stress for assessing potential for remote sensing abiotic stressors in coffee management, **Proceedings of the 15<sup>th</sup> WaterNet/SADC/GWP Symposium**, le Meridien Resort, Mauritius, 27-30 October 2015.
2. **Chemura, A.,** Mutanga, O. and Dube, T (2017) Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests, **Physics and Chemistry of the Earth ABC**.

## **Abstract**

Water management is an important component in agriculture, particularly for perennial tree crops such as coffee. Proper detection and monitoring of water stress therefore plays an important role not only in mitigating the associated adverse impacts on crop growth and productivity but also in reducing expensive and environmentally unsustainable irrigation practices. Current methods for water stress detection in coffee production mainly involve monitoring plant physiological characteristics and soil conditions. In this study, the ability of selected wavebands in the VIS/NIR range to predict plant water content (PWC) in coffee using the random forest algorithm was tested. An experiment was set up such that coffee plants were exposed to different levels of water stress and reflectance and plant water content measured. In selecting appropriate parameters, cross-correlation identified 11 wavebands, reflectance difference identified 16 and reflectance sensitivity identified 22 variables related to PWC. Only three wavebands (485nm, 670nm and 885nm) were identified by at least two methods as significant. The selected wavebands were trained (n=36) and tested on independent data (n=24) after being integrated into the random forest algorithm to predict coffee PWC. The results showed that the reflectance sensitivity selected bands performed the best in water stress detection ( $r = 0.87$ , RMSE = 4.91% and pBias = 0.9%), when compared to reflectance difference ( $r = 0.79$ , RMSE = 6.19 and pBias=2.5%) and cross-correlation selected wavebands ( $r = 0.75$ , RMSE = 6.52 and pBias = 1.6). These results indicate that it is possible to reliably predict PWC as part of coffee condition assessment using wavebands in the VIS/NIR range that correspond with many of the available multispectral scanners using random forests and further research at field and landscape scale is required to operationalize these findings.

**Keywords:** Plant water content, water stress monitoring, remote sensing, coffee, random forests

## 5.1 Introduction

Plantation agriculture uses more water per unit area per year than other agricultural enterprises (Prosser & Walker, 2009). This is because perennial tree crops are usually large plants with high leaf areas that consequently transpire more water, their tree canopy intercepts more rain, they often have deeper roots and use water over the whole year (O'Loughlin & Nambiar, 2001; Farley et al., 2005; Gordon et al., 2005). Since coffee (*Coffea arabica*) evolved from the highlands of Ethiopia (Coste, 1992), it is commonly produced in headwater catchments of river systems where its total moisture requirements are considerably influenced by the annual evaporation which, in turn, is influenced by the relative humidity, temperature, lengths of overcast weather and the amount and nature of rainfall (Logan & Biscoe, 1987; Hess et al., 1998). Consequently, irrigation is a necessity for successful coffee production in many areas due to the heightened increase in temperatures and reduced rainfall as a result of climate change and variability (ICO, 2009; Schroth et al., 2009). However, even in areas where a reliable source of good quality water supply is available, irrigation comes with increased costs of installation, water pumping and maintenance (Masarirambi et al., 2009). A water management program is therefore used in coffee production. Water availability is a critical factor for coffee plant survival, development and productivity and the monitoring of plant water status has important implications for profitability and sustainability.

Coffee plants affected by water stress go through a variety of physiological processes, such as damage or removal of the waxy cuticle, destruction of cell walls, reduced stomatal conductance and retarded rates of net carbon assimilation (DaMatta, 2004). This does not only affect growth and productivity, but predisposes the plant to infestation by opportunistic insect pests and diseases such as coffee white stem borer and coffee leaf rust (Kutywayo, 2002a; Gay et al., 2006; Ghini et al., 2011). Current plant water management and monitoring approaches in coffee production include general calendar-based irrigation scheduling, use of tensiometers, use of soil moisture probes and reliance on experienced farm managers who occasionally stroll the fields looking for signs of moisture stress (Logan & Biscoe, 1987). This is because precise laboratory based plant water content (PWC) quantification is time consuming, expensive and reliant on sampling, which is usually not very representative of the entire field condition, above being prone to instrumental and human errors (Glenn et al., 2008).

Remote sensing, particularly the use of hyperspectral sensors provides very promising options for early, objective and spatially resolved water stress detection for coffee management. Hyperspectral remote sensing data consists of many, very narrow contiguous spectral



wavebands located from the visible, near infrared, mid-infrared and thermal infrared portions of the electromagnetic spectrum. Many studies have confirmed the existence of strong water absorption bands in specific channels of the mid-infrared region of the electromagnetic spectrum specifically at around 1175nm, 1450nm, 1650nm, 1940nm, 2250nm and 2500nm (Carter, 1991; Glenn et al., 2008; Jeger & Pautasso, 2008). In addition, to the mid-infrared region having primary bands associated with plant water content, the visible to near infrared (VIS/NIR) region (400-1300 nm) has secondary water content related bands that can be used in water stress detection (Carter & Miller, 1994; Lin, 2011). This is possible because plant water content has several primary and secondary effects on leaf characteristics, which in turn influences leaf, canopy and top-of-the-atmosphere scale spectral reflectance.

Primary effects of water absorption are associated with water absorption that occurs in the mid infrared range, where they result in water mediated decrease in leaf reflectance in turgid plants, when compared to leaves with less water content (Carter, 1991). Secondary effects of water content on the spectral reflectance of leaves are influenced by the transmissivity of water that occur because of the leaf internal structure, cell size and cell shape and besides they are less pronounced when compared to primary effects (Carter & Miller, 1994; Oumar & Mutanga, 2014). These secondary effects are therefore very attractive from a remote sensing perspective, since relying on the full width reflectance of the electromagnetic spectrum for detecting and quantifying water deficit stress in field crop management present greater challenges in data handling, dimensionality, costs and noise, all of which dissuade potential users to rely on this range of data for determining water deficit stress in crops.

Using secondary effects of water absorption on leaf reflectance, Peñuelas et al. (1994) developed the water band index (WBI) which is based on the ratio between the water band at 970 nm and reflectance at 900 nm which has been widely used in plant water content assessments with remote sensing. For example, Oumar and Mutanga (2014) observed that the WBI was the most important variable that uses reflectance from the optical domain in estimating plant water thickness (EWT). Similarly, Eitel et al. (2006) found significant correlations between WBI and PWC ( $r = 0.90$ ), and WBI and EWT ( $r = 0.88$ ). This indicates that secondary effects of water absorption that occur in the VIS/NIR region of the spectrum could be useful in estimating PWC and EWT in crop plants, without using the primary effects found beyond 1000nm. This is important because many of the identified wavebands and consequently resultant vegetation indices for water stress detection that use primary effects of water absorption coincide with very intense atmospheric water vapour absorption (Lillesand &

Kiefer, 2008; Kahn et al., 2011). This limits operational remote sensing using these in tropical areas that are producing perennial tree crops such as coffee. In addition, many of the studies on PWC estimation using remote sensing, focused on detecting fuel moisture content (FMC) in fire risk modelling across large natural ecosystems (Omont et al., 2006; Lin, 2011; Ortega-Huerta et al., 2012; Dixon & Garrity, 2014), in timber plantations (Jeger & Pautasso, 2008; Mutanga & Ismail, 2010; Oumar & Mutanga, 2014) and annual crops (Liu et al., 2003; Venteris et al., 2015) and yet these are significantly different from tree crops, such as coffee in terms of leaf area index (LAI), leaf angles, canopy structure and photosynthetic channels.

The ability to predict plant water stress using wavebands in the VIS/NIR region allows for the early detection of unnoticeable stress that is caused by water stress associated with soil moisture, disease infestation and pest attack, while opening opportunities for stress detection using handheld scanners, multispectral satellite scanners and simple unmanned aerial vehicles (UAV). The objective of this study therefore was to evaluate the effectiveness of wavebands in the VIS/NIR region in predicting coffee PWC. This objective is achieved by identifying the most significant water stress related wavebands that can be integrated into a modelling algorithm for efficient PWC estimation in coffee in order to ensure economic sustainability while reducing unnecessary water losses and related financial and environmental costs.

## **5.2 Materials and methods**

### **5.2.1 Study area**

The study was carried out at Coffee Research Institute (CoRI) in Chipinge, Zimbabwe. CoRI is located at coordinates 32°37.523'E and, 20°12.474'S at an altitude of 1100 m.a.s.l. The average annual rainfall is 1180mm of which 80% falls in five months from November to March. The mean maximum temperature is 20°C and minimum is 14°C. Most of the soils in this area are leached and strongly weathered and in the Orthoferralitic group derived from Umkondo quartzite and sandstone (Chemura, 2014). Chipinge is in the main coffee production zone in Zimbabwe.

### **5.2.2 Experimental materials and design**

Six months old healthy coffee seedlings (variety Catimor 129) were used in the study. These seedlings were transplanted into black polythene pots (29cm x 13.5cm) at four months and were allowed two months to acclimatize the conditions of the greenhouse before

experimentation. This age represented the minimum age at which coffee can be transplanted into the field. The seedlings were grown in the nursery in recommended growth medium for raising coffee seedlings and all other routine nursery management activities were based on nursery recommendations from the Coffee Handbook (Logan & Biscoe, 1987).

### **5.2.3 Moisture stress treatments**

The six months old seedlings were subjected to water stress treatments by withholding and varying water supply to the plants to obtain a range of soil moisture. Three treatments were used for inducing plant water stress; each replicated 20 times (60 plants). The first batch had no moisture stress where the seedlings were provided with required irrigation twice a week (not stressed). The second had plants being provided irrigation once a week (moderate stress) while the third had plants that were not provided with irrigation water for two weeks prior to spectral assessments and PWC measurements (severe stress). This allowed for a gradient in plant water content.

### **5.2.4 Spectral reflectance measurements**

On each plant, spectral reflectance was measured on one of the leaves on the third node from the top. This was done to sample mature coffee leaves that are representative of those in the field. Reflectance was measured using an Apogee VIS-NIR spectrometer (Apogee Instruments, Inc., Logan, UT, USA) which has an effective spectral range of 400-900 nm and a spectral resolution of 0.5 nm. Each reading consisted of an average of three spectral scans, taken at 15 cm above the coffee leaf of interest at 30° angle. A white polytetrafluoroethylene (PTFE) reflectance standard was used as a reference, and reflectance by wavelength was calculated as the ratio of scene reflectance to the reflectance of the standard. The reflectance was averaged to 10nm to reduce dimensionality and a moving Savitzky–Golay filter with a frame-size of 5 data points and a 2<sup>nd</sup> order polynomial was employed to smooth the spectra (Savitzky & Golay, 1964).

### **5.2.5 Leaf water content measurements**

The physiological characteristics (leaf length, width, fresh weight and dry weight) of the leaves on each plant that was used in spectral sampling was recorded. Fresh weight (FW) was measured soon after reflectance measurements, using an electronic balance on site. For determining dry weight, the leaves were placed in tagged containers and placed in an oven set at 70°C for 8 hrs and afterwards their mass was measured using the digital balance. Plant Water Content (PWC) was calculated after Liu et al. (2003) as in equation 5.1.

$$\text{PWC}(\%) = \frac{\text{FW}(\text{g}) - \text{DW}(\text{g})}{\text{FW}(\text{g})} * 100 \quad [5.1]$$

### **5.2.6 Identification of water stress related wavebands**

Hyperspectral data are high dimensional and exhibit a high degree of inter-band correlation, leading to data redundancy that can cause convergence instability in models (Thenkabail et al., 2004a; Blackburn, 2007). Therefore, the use of fewer wavebands is preferable for more stable modelling of plant biophysical parameters and chemistry with hyperspectral data. It is also easier to implement in field applications. Variable selection methods attempt to capture the maximum information present in the original assorted data, while concurrently making sure that the selected data remain fit for purpose (Fodor, 2002; Demšar et al., 2013). Many variable selection methods have been reported for use with remote sensing data. Each has unique data processing capabilities and potential applications in selecting useful wavebands. The objective of all variable selection methods is to produce a list of features arranged by their discriminatory ability and therefore provides a means by which an optimal feature subset can be used in modelling from remote sensing data. Since no single best approach was available to determine the optimal number of bands required for coffee PWC modelling, three variable selection methods were used in this study, which are cross-correlation threshold, reflectance difference and reflectance sensitivity.

#### **5.2.6.1 Cross-correlation threshold**

Cross correlation is a standard measure of degree of similarity between two band matrices and is optimized for faster calculation and /or more accurate results (Ahmed et al., 2012). This method is dependent upon covariance calculation between the two wavebands, and can be adjusted for brightness and contrast using normalization (Wang et al., 2008). The higher the correlation between the bands, the more the similarity in their spectral characteristics for a feature of interest and vice-versa. Therefore, in the correlation matrix all areas of low coefficient of determination ( $R^2$ ) values are the waveband regions with the least redundancy and the highest information content and therefore should be retained (Thenkabail et al., 2004a). As such, a cross-correlation threshold should be set, beyond which datasets should be considered unsuitable for modelling. As expected, the adjacent bands of the hyperspectral data had very high inter-correlations, meaning that they are redundant. In this study this threshold was set as  $R^2=0.95$ , meaning that for wavebands with a coefficient of determination of more than 0.95, only one was retained.

### 5.2.6.2 Reflectance difference

The reflectance difference of the stressed treatments to the unstressed to calculate the spectral difference between the irrigated, moderate and severely stressed plants was also used. The objective of reflectance difference was to identify a few important bands based on the peak value of the reflectance difference across the three treatments, but not to seek the variation in the peak value itself. The mean reflectance values for each water deficit stress treatments was considered instead of that of individual replicates. Relative difference in reflectance, due to water stress was measured by calculating the reflectance difference at each wavelength after Riedell et al. (2003):

$$\text{Reflectance Difference(RD)} = (\text{Reflectance of Stressed} - \text{Reflectance of Unstressed}) \quad [5.2]$$

### 5.2.6.3 Reflectance sensitivity

The reflectance sensitivity method is an extension of the reflectance difference method where the difference between the stressed and the non-stressed leaf reflectance is considered as a proportion to the reflectance of the non-stressed samples. This method therefore normalizes the differences in spectral reflectance to the origin. The wavebands with the highest reflectance sensitivity represented the areas of the spectrum with the most important information and were therefore selected. Reflectance sensitivity was calculated after (Riedell et al., 2003):

$$\text{Reflectance Sensitivity(\%)} = [(\text{RD}/\text{Reflectance of Unstressed}) * 100] \quad [5.3]$$

### 5.2.7 Modelling approach

The Random Forest (RF) algorithm was used for modelling PWC from selected wavebands. RF is an ensemble machine learning algorithm developed by Breiman (2001) to solve classification and regression problems through a multitude of decision trees. RF employs an iterative bagging (bootstrap aggregation) operation where a number of trees (*n*tree) are independently built, using a random subset of samples from the training samples. Each tree is then independently grown to a maximum size based on a bootstrap sample of about two-thirds the training dataset. Each node is then split using the best, among a subset of input variables (*m*try). The ensemble then classifies the data that are not in the trees as out-of-bag (OOB) data, and by averaging the OOB error rates from all trees, the RF algorithm gives an error rate called the OOB error for each input variable (Gislason et al., 2004; Breiman & Cutler, 2007). In many applications, this algorithm produces one of the best accuracies to date and has important advantages over other techniques in terms of ability to handle highly non-linear data, robustness to noise and tuning simplicity (Rodriguez-Galiano et al., 2012; Lebedev et al.,

2014). The default number of trees (ntree) of 500 was used while mtry is automatically determined as the square root of the total number of variables used (Breiman, 2001). The r package randoForest was used for running the RF modelling (Liaw et al., 2009).

### 5.2.8 Model evaluation

The field data was randomly partitioned into 60:40 for model training and validation respectively. Several error indices are commonly used in model evaluation and some of them were applied to compare RF model performances and to assist in identifying the best performing variable selection methods. All model evaluation metrics were performed on independent data. The correlation coefficient (r) and coefficient of determination ( $R^2$ ) was used to assess the goodness of fit of the predicted PWC and measured PWC values. In terms of performance, the best model should be identified as the one with the largest r and  $R^2$ . In addition, mean absolute error (MAE, Equation 5.4) root mean square error (RMSE, Equation 5.5), normalized root mean square error (nRMSE(%), Equation 5.6) and percent bias (pBias (%), Equation 5.7) were used to determine the errors of the model in predicting PWC from selected variables. For MAE, RMSE, and pBias values of 0 indicate a perfect fit between measured and predicted PWC (Ghini et al., 2011). MAE is the average of the absolute values of the differences between predicted and measured values. RMSE is one of the commonly used error index statistics and the lower the RMSE the better the model performance. pBias measures the average tendency of the simulated data to be larger or smaller than their observed counterparts and positive values indicate overestimation whereas negative values indicate model underestimation (Ghini et al., 2011).

$$\text{MAE} = \left( \frac{1}{n} \sum |y_i - \hat{y}_i| \right) \quad [5.4]$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - \hat{y})^2} \quad [5.5]$$

$$\text{nRMSE} = \frac{\text{RMSE}}{y_{i \max} - y_{i \min}} \quad [5.6]$$

$$\text{pBias} = \left( \frac{\sum (y_i - \hat{y}_i) * 100}{\sum y_i} \right) \quad [5.7]$$

where for all cases n is the number of data points,  $y_i$  is the measured PWC (%) at that data point and  $\hat{y}_i$  is the model predicted PWC (%) at that data point.

### 5.3. Results

#### 5.3.1 Coffee plant water content and reflectance

The distribution of the PWC for all the samples is shown in Figure 5.1. The descriptive statistics of the plant water content for each of the three treatments used in the study are shown in Table 5.2. PWC ranged between 27.9% and 76.8% in all treatments, with an average of 53%. The results showed that reflectance from the stressed plants was higher than reflectance of the plants that had constant water supply (Figure 5.2a). Correlating the PWC with reflectance of each wavelengths showed that the PWC was negatively correlated with reflectance and that highest correlations were in the NIR bands (Figure 5.2b).

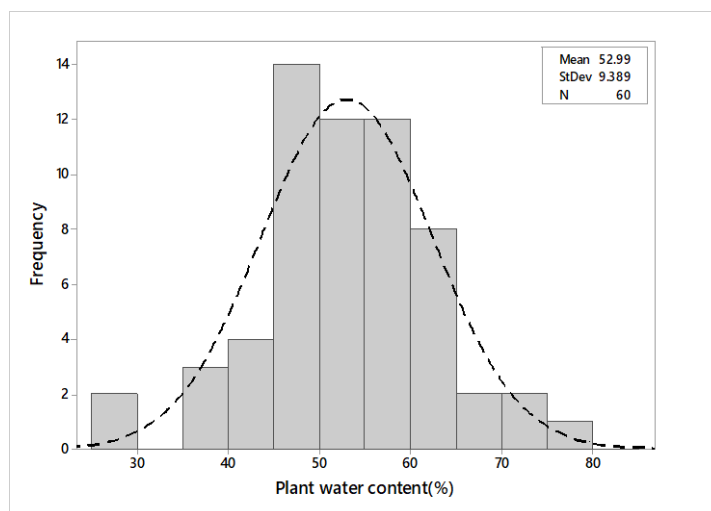


Figure 5.1: Distribution of plant water content of coffee leaves used in the study.

Table 5.1: Descriptive statistics of the PWC for the three stress levels.

Treatment	Min	Max	Mean	SD
Irrigated	45.67	76.84	55.33	4.9
Moderate Stress	37.89	65.06	54.92	11.2
Severe stress	27.86	64.26	48.73	9.8

#### 5.3.2 Variable selection

##### 5.3.2.1 Cross-correlation threshold

The results showed that the bands in the same regions were highly correlated and therefore redundant in modelling PWC. Inter-region cross correlation was particularly low between the NIR region and the red-edge region of the spectrum (Figure 5.2a). Lower correlations were

also observed between the NIR range and the visible parts of the spectrum. The application of the cross correlation threshold on the correlation matrix resulted in eleven wavebands being selected as having unique and independent information and therefore could be used in modelling PWC. Five of these wavebands were located in the red-edge position (between 600 and 700nm) while four were in the NIR region with only two bands being in the visible region of the spectrum (Figure 5.2b).

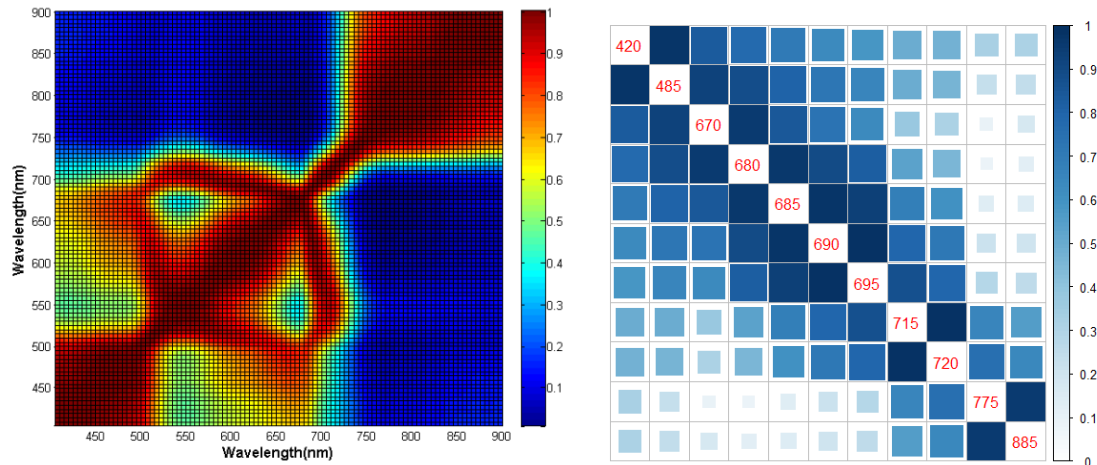


Figure 5.2: Coefficient of determination for (5.2a) all wavebands between 400 and 900 and (5.2b) selected wavebands using the threshold  $R^2$ .

### 5.3.2.2 Reflectance difference and reflectance sensitivity

The reflectance difference and reflectance sensitivity functions were applied on the spectra for variable selection. Reflectance difference selected 16 variables mainly in the NIR range between 830nm and 900nm. Two difference inflection points between the water stressed and the irrigated treatments were identified at 705nm and 735nm as important variables (Figure 5.3a). These indicate the wavelength at which reflectance difference is highest between the irrigated, the moderate stress and the severe stress treatments and thus, the wavelengths at which the PWC can be successfully separated. Normalising the differences by calculating the ratio of the reflectance difference to the FW produced a different set of variables than that obtained from reflectance differences. Variables selected were in the visible range and in the red-edge region of the spectrum with no variables selected in the NIR region (Figure 5.3b).



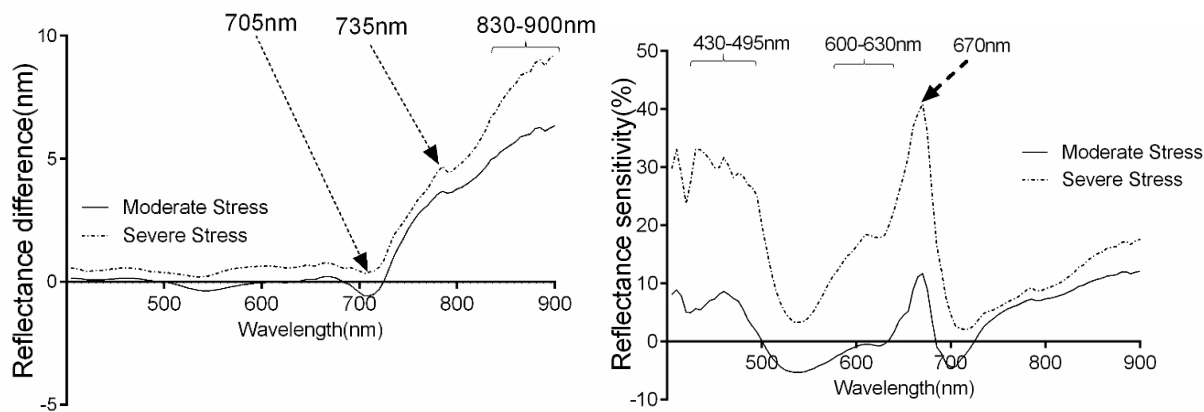


Figure 5.3: The application of the reflectance difference and reflectance sensitivity for variable selection. Figure 5.3a shows the variables selected using reflectance difference and Figure 5.3b shows the variables selected using reflectance sensitivity.

### 5.3.3 Variable importance

The variable importance obtained from the internal RF variable ranking for each method are shown in Figure 4. Two variables (the 775nm and 885nm) are distinctively significant in PWC estimation for the variables selected through cross-correlation (Figure 5.4a). 830nm and 890nm parameters were observed to be very important of the variables selected by reflectance differencing (Figure 5.4b). There were many variables considered important in predicting PWC in the variables selected by reflectance sensitivity, indicating a combined influence of parameters on the output rather than individual significance of parameters (Figure 5.4c).

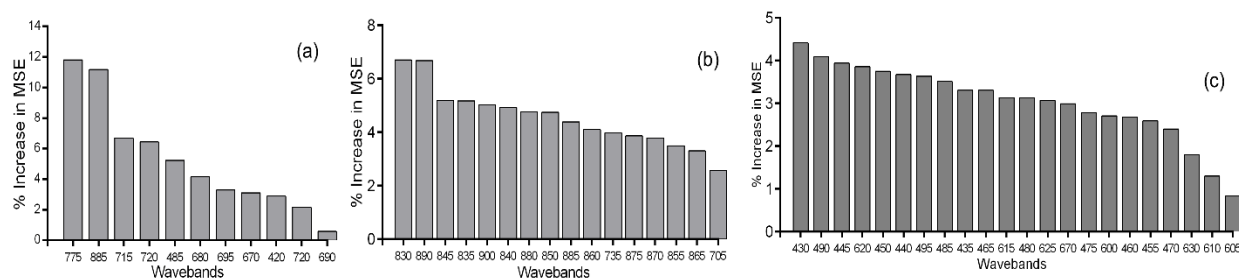


Figure 5.4: Variable importance of wavebands used in modelling plant PWC as selected by (5.4a) Cross-correlation (5.4b) Reflectance Difference and (5.4c) Reflectance Sensitivity.

### 5.3.4 Model performance evaluation

#### 5.3.4.1 Cross-correlation threshold

The wavebands selected by cross-correlation were trained using the training dataset ( $n=36$ ) and their coefficients were then used to predict PWC on independent validation dataset ( $n=24$ ). The use of variables selected by cross-correlation resulted in a good fit between RF predicted and measured PWC ( $r=0.75$ ,  $RMSE=6.52$ , Figure 5.6a). The RF model had a positive bias meaning that it was overestimating water content.

### 5.3.4.2 Reflectance difference

The fit between measured and predicted PWC from reflectance difference selected variables was better than that obtained from cross-correlation threshold selected variables. The correlation coefficient between measured and RF predicted PWC was 0.79 with a reduced RMSE to 6.19 (Figure 5.6b). Reflectance selected variables; however, had the highest positive bias (2.5%), showing that there was an overestimation of when predicted values are compared to the measured values.

### 5.3.4.3 Reflectance sensitivity

Reflectance sensitivity selected variables had the best fit to data, when compared to other variable selection methods. It had the highest correlation coefficient of 0.87, the lowest RMSE of 4.91% and the least bias of 0.9% (Figure 5.5c). Other model performance values used in evaluating the performance of the three variable selection methods are shown in Table 5.2.

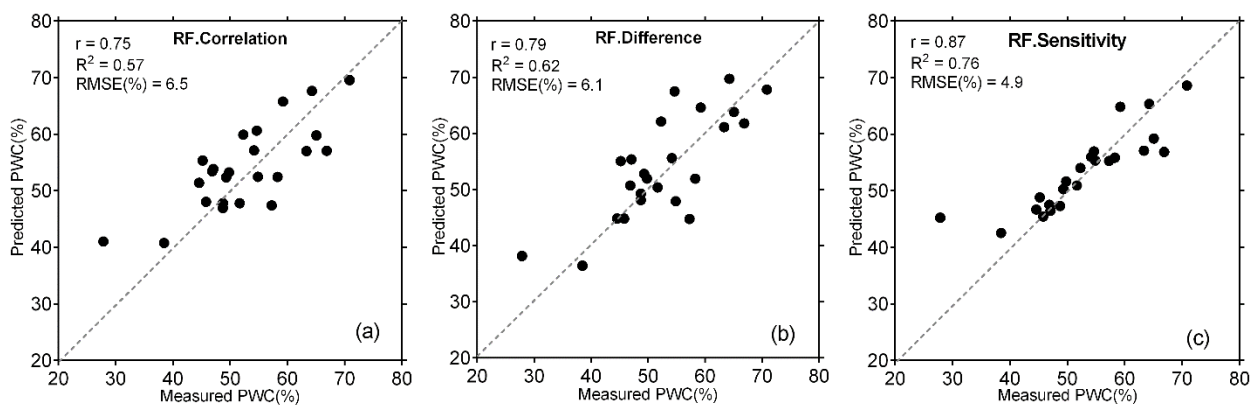


Figure 5.5: One-to-one plots showing performance of the RF model for predicting PWC from wavebands for selected through (5.5a) Cross-Correlation, (5.5b) RD and (5.5c) RS.

Table 5.2: Summary model performance evaluation using test dataset (n=24)

Selection Method	N	r	R <sup>2</sup>	MAE (%)	RMSE (%)	nRMSE (%)	pBias (%)
Cross-Correlation	11	0.75	0.57	5.65	6.52	67.5	1.6
Reflectance Difference	16	0.79	0.62	4.83	6.19	64.2	2.5
Reflectance Sensitivity	22	0.87	0.76	3.21	4.91	50.9	0.9

## **5.4 Discussion**

This study aimed at evaluating the effectiveness of wavebands in the VIS/NIR region in modelling PWC in coffee, a perennial tree crop of economic importance. This was achieved by identifying the most significant water stress related wavebands, using three methods and these were integrated into random forests, a novel modelling algorithm. The prediction of plant water content in plantation crops, such as coffee is very important for maintaining productivity, profitability, crop health and in safeguarding long-term investments that farmers make in their plantations. In coffee production, reliable estimates of plant water content are not just required for keeping growth vigour but for necessary agronomic practices, such as inducing uniform and high level of flowering in a short window of time, without which yield and quality, two important parameters in determining farmers' returns, are significantly reduced (Masarirambi et al., 2009). Therefore, the ability of remote sensing to provide accurate plant water content estimates for coffee as shown in this study bring opportunities for improved production and profitability in the sector, while safeguarding water resources.

### **5.4.1 Comparison of variable selection methods**

This study has identified secondary wavebands in the VIS/NIR region that can be used for estimation of PWC in coffee plants for field level decision making. Sensitivity could have produced very good predictive performance for PWC because unlike reflectance difference, it magnifies points where the effects of the water stress are most significant relative to their origin. Therefore, the overall difference in reflectance between the irrigated and water stressed may not be as important as the percentage change in reflectance. This is important in order to normalize the effect of water stress across the spectrum that is known to have significant patterns in vegetation reflectance, particularly in the NIR region. The cross-correlation threshold produced good results, but were inferior to the other methods. This is possibly because this method is not in any way related to the plant condition, as it just considers the characteristics of one waveband's reflectance in relation to another waveband, and not its behaviour as caused by the stress.

### **5.4.2 Range of identified bands for PWC estimation**

It is worth noting that the variables selected by reflectance sensitivity were mainly in the red-edge and the visible, particularly the blue region of the spectrum, with none in the NIR region. This result was rather unexpected because vegetation reflectance in this region is strongly influenced by chlorophyll, dry matter and leaf internal structure (Lin, 2011; Ortega-Huerta et al., 2012). Since this study is focused on one plant species, coffee, the influence of these factors

could be discounted, as they are more apparent in PWC estimation over multiple plant species whose biophysical parameters could be widely different. What is very surprising is the conspicuous absence of application of the blue region in PWC in order to make comparisons with this study. Many of the studies on PWC estimation effectively excluded the blue region of the spectrum (Lin, 2011; Ortega-Huerta et al., 2012; Dixon & Garrity, 2014), which is commonly regarded as the “water band” in remote sensing. When Kahn et al. (2011) evaluated the significance of wavebands related to effective leaf water thickness (EWT), they noted that 425nm was important in plant water estimation, which concurs with findings in this study. In this study, 430nm waveband was ranked as the most important band for all the bands selected by RS in predicting PWC.

These results compare well with studies that applied the known water absorption bands in the mid-infrared spectra such as at 1450nm, which obtained correlations of 0.78 (Mutanga & Ismail, 2010; Oumar & Mutanga, 2014). However, the seemingly better correlation observed in this study could stem from the fact that there was a direct sensing of water content in this study in healthy plants while in other studies the focus was on predicting water content in disease/insect infected crops. Other studies used both raw and first-derivative reflectance to predict plant water content and got very promising results (Kumar, 2007) while other used vegetation indices to predict PWC (Eitel et al., 2006).

#### **5.4.3 Potential applications and limitations**

The results presented in this study are important for coffee, which unlike other field crops, has an added water requirement for stimulating flowering (Logan & Biscoe, 1987). This additional role of water management in breaking bud dormancy and subsequent floral development has a direct influence on yield capacity and is dependent on the continuation of satisfactory plant water relations. This is additional to the water required to support maximum vegetative growth and berry development. If water is not supplied in adequate amounts, water stress will result in a drop in the rate of vegetative growth and if this is prolonged it will result in leaf fall and dieback, and if during bean filling, it will result in bean abortion, which reduces yield of coffee (Nair, 2010).

In addition, the ability of wavebands in the visible and NIR range to predict PWC is very attractive for application of multispectral scanners, particularly of high spatial resolution sensors and simple UAVs in PWC estimation. Using high spatial resolution sensors, such as GeoEye, WorldView-2 and 3, RapidEye and Sentinel series is very important for predicting

PWC as part of precision agriculture. This is because, as explained by Lin (2011) it is not possible to use low-resolution satellite data for PWC estimation at species level because one pixel will usually contain a mixture of several species and soil backgrounds which will affect the results. In addition to their ability to focus on specific crop fields and super-spectral bands strategically positioned across important regions of the spectrum, there is therefore potential for use of wavebands in the VIS/NIR region for PWC prediction.

The study, however, relied on healthy coffee leaves and under field conditions; other factors may influence the accuracy of the results. For example, plant diseases and insect infestations (Mutanga & Ismail, 2010; Oumar & Mutanga, 2014) tend to have an influence on plant water content and these are not accounted for in this study. In many areas around the world, coffee is often produced under shade, which may also limit the application of remote sensing approaches to detect PWC. Consideration of other factors that can affect PWC estimation such as age of the coffee plants can be handled by prior age-mapping before a method such as this is implanted (Chemura & Mutanga, 2016). Nonetheless, this study is important in proving a method for detecting PWC using a few selected bands in the VIS/NIR region making possible for wider application of PWC prediction using multispectral sensors and UAVs.

## **5.5 Conclusions**

This study presents findings on the ability of spectral features in the multispectral region of the spectrum to predict coffee PWC. It was concluded that hyperspectral wavebands in the VIS/NIR region selected by sensitivity analysis produce the best accuracy and the least errors in PWC estimation in coffee. Overall, this study provides a basis for application of remote sensing in precision irrigation planning, which has benefits in terms of crop productivity and in reducing unnecessary losses associated with excessive irrigation.

## **5.6 Link to next chapter**

This chapter presented the ability to model plant water content as an individual coffee stressor from multispectral remote sensing data. Plant water stress can explain the anomalies identified in Chapter 4. The next chapter presents an assessment of remote sensing to identify an individual biotic stress factor, coffee leaf rust which can also specifically explain coffee plant condition.

## CHAPTER 6: MULTISPECTRAL LEVEL REMOTE SENSING-BASED DISCRIMINATION AND SEVERITY MODELLING OF COFFEE LEAF RUST



Photo credits: A. Chemura (2017)

### **This chapter is based on:**

1. **Chemura, A.,** Mutanga, O and Dube, T (2016) Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions, *Precision Agriculture* **18**(5), 859–881.
2. **Chemura, A.,** Mutanga, O. Sibanda, M. and Chidoko, P., (2017) Machine learning prediction of coffee rust severity on leaves using spectroradiometer data, *Tropical Plant Pathology*, **In Press**.

## Abstract

Coffee leaf rust (CLR), which is caused by the fungus *Hemileia vastarix*, is a devastating disease in coffee producing countries and remote sensing has the potential to detect and monitor the disease. This study evaluated the detection, discrimination and quantification of CLR infection levels to inform on possibility of using Sentinel-2 data for disease monitoring. Field spectra was resampled to the band settings of the Sentinel-2 MSI sensor. The ability of remote sensing to discriminate levels of infection was evaluated using the random forest (RF) and partial least squares discriminant analysis (PLS-DA) algorithms with and without variable optimization. For modelling severity, the non-linear radial basis function-partial least squares regression (RBF-PLS) was employed. The results showed that using all variables, Sentinel-2 MSI derived vegetation indices achieved higher overall accuracy of 76.2% when compared to 69.8% obtained using raw spectral bands. Using the RF OOB scores, 4 spectral bands and 7 vegetation indices were identified as important variables in CLR discrimination. Using the PLS-DA VIP score, 3 spectral bands (B4, B6 and B5) and 5 vegetation indices were found to be important variables. Use of the identified variables improved the CLR discrimination accuracies for both the RF and the PLS-DA. The RBF-PLS derived models satisfactorily modelled CLR severity ( $R^2=0.92$  and  $RMSE=6.1\%$  with all bands and  $R^2=0.78$  and  $RMSE=10.2\%$  with selected bands) when compared to PLS ( $R^2 = 0.27$  and  $RMSE = 18.7\%$  with all bands and  $R^2 = 0.17$  and  $RMSE = 19.8\%$  with model optimally selected bands). Specifically, four bands; Band 2 (490nm), Band 4 (665nm), Band 5 (705nm) and Band 7 (783nm) were identified as the most important spectral bands in modelling CLR severity. Better accuracy was obtained in modelling severe levels of CLR ( $R^2=0.71$  with all variables) when compared to moderate infection levels ( $R^2=0.38$  using all variables). Overall, this study underscores the applicability of Sentinel-2 MSI spectral settings for accurate disease monitoring and modelling in producer countries as part of crop condition assessment, which can be confounded by co-occurrence of other crop characteristics that affect the spectral signature and spatial resolution of the bands.

**Keywords:** disease discrimination, red-edge, precision agriculture, variable optimization, resampled data, swath-width

## 6.1 Introduction

There are increases in severity and frequency of coffee biotic and abiotic stressors such as diseases, pests and drought that are limiting yield and quality of coffee produced, with consequences on livelihoods and economies dependent on coffee production. Regular and systematic monitoring of these is therefore required for sustainable and profitable coffee production. Coffee leaf rust (CLR), caused by the fungus *Hemileia vastatrix*, is the most destructive disease and threat to coffee production the world-over (Ghini et al., 2011; Cressey, 2013). The disease is most severe on *Coffea arabica*, which accounts for about 70% of the world coffee production and supply (Dinesh et al., 2011). Unlike other plant fungal diseases, CLR is not necrotic and its symptoms appear only on the underside of the leaves where the pathogen penetrates through the stomata, resulting in small yellowish lesions, which grow and coalesce to form uredospore with a distinct yellow or 'rustic' colour (Gouveia et al., 2005; Belan et al., 2015). In the absence of early detection and proper management, CLR results in up to 50% loss of leaves and 70% yield reduction in coffee through premature leaf drop, dieback and debilitation of trees, which will eventually lead to death of coffee plants (Avelino et al., 2004).

Current CLR disease monitoring methods rely on occasional field surveys by a team of specially trained and experienced personnel. However, besides being the largely adopted approach, particularly in resource limited areas, the technique is strenuous and subjective, with the results being confirmatory, mostly once the disease has fully established with already severe inflicted economic damages on the crop. Remote sensing therefore, offers unlimited opportunities for instantaneous, timely and spatially explicit objective assessment of plant condition throughout the growing season (Sankaran et al., 2010). The urgent adoption of these technologies is perceived to have the potential to increase crop productivity through the provision of accurate and up-to-date crop information, as well as reduce unplanned costs on disease control and environmental contamination associated with the over-application of pesticides. The success of potential crop protection methods is highly dependent on early disease detection (Rumpf et al., 2010; Martinelli et al., 2015).

Previous work demonstrates that remote sensing approaches can be reliably used in the detection of plant diseases in many crops. For instance, Huang et al. (2007) demonstrated that the photochemical reflectance index (PRI) developed from hyperspectral remote sensing data can be applied to identify and quantify yellow rust in winter wheat ( $r^2=0.91$ ), providing a basis for development of a proximal, airborne or spaceborne imaging sensors for its monitoring. In



addition, Mahlein et al. (2013) developed specific spectral disease indices (SDIs) for the detection and discrimination of healthy sugar beet leaves from those infected with cercospora leaf spot (*Cercospora beticola*), sugar beet rust (*Uromyces betae*) and powdery mildew (*Erysiphe betae*) and achieved high accuracy and sensitivity of over 85%. It is therefore clear from these studies that species-specific disease indices derived from hyperspectral bands located in the narrow contiguous parts of the red-edge and NIR regions have the strength and capability to significantly enhance crop disease detection, identification and monitoring.

Reflectance in the red-edge and NIR regions is able to identify changes in internal leaf structure, content and processes that affect absorption of radiation and thus reveal physiological stress in plants caused by the disease or other stress (Coops et al., 2003; Eitel et al., 2011). These detected changes are associated with changes in the quality and quantity of chlorophyll and chemical properties of the affected leaves, when compared to their previous stress-free condition or unstressed counterparts (Carter & Knapp, 2001; Mutanga & Skidmore, 2007). Many of these specific wavebands were previously only available in hyperspectral sensors, which are known to have many challenges, such as high costs, high dimensionality and above all poor spatial coverage. So far, there are no prospects that there will be global coverage of affordable hyperspectral data in the near future at sufficient temporal resolution (Rulinda et al., 2012). This therefore means that there is need to shift towards harnessing the opportunities provided by the new generation of multispectral imaging sensors.

New generation multispectral space-borne earth observation instruments, such as WorldView-2, RapidEye and Sentinel-2 multispectral imager have incorporated narrow wavebands including those in the red-edge position that were not available in predecessor sensors (i.e. Landsat series, MODIS, SPOT, ASTER etc). These technological advancements therefore provide an opportunity for timely landscape or farm-based assessment of crop condition (i.e. health status and yield estimation). Unlike the hyperspectral sensors, multispectral sensors have a huge swath-width and are currently available at low or no costs for many developing countries where coffee is produced. There has been a lot of interest in the upcoming Sentinel-2 multispectral imager data in terms of its potential applications as a game changer in multispectral remote sensing (Wang et al., 2004; Funk & Budde, 2009; Rulinda et al., 2012; Atzberger, 2013; Rembold et al., 2013; Ding et al., 2014). This is because it is freely available with relatively high resolution (10m for some bands), as well as strategically positioned bands

(such as two red-edge bands), which makes it useful for many applications, including vegetation characterization and mapping.

The Sentinel-2 multispectral satellite capitalizes on the technology and the vast experience acquired with SPOT and Landsat series data over the past decades. Sentinel-2 multispectral imager which is a polar orbiting, super-spectral high resolution imaging mission (Hansen & Loveland, 2012), has a huge swath-width of about 290 km with thirteen unique spectral bands. These spectral bands range from the visible and near infrared (VNIR) to the shortwave infrared (SWIR) regions of the spectrum. Of these thirteen bands, four are provided at 10 m spatial resolution, six bands at 20 m spatial resolution and three bands at 60 m spatial resolution (Funk & Budde, 2009). Most importantly, the sensor provides data about the earth's surface every five days under cloud-free conditions, and typically every 15–30 days in cloudy areas, making it attractive for temporal feature analysis (Dangwal et al., 2016).

One of the most attractive features of the Sentinel-2 multispectral imager is that it incorporates three new bands in the red-edge region, which are centred at 705, 740 and 783 nm specifically designed for vegetation characterization and quantification (Frampton et al., 2013; Hedley et al., 2012). Because of these advanced sensor characteristics, Sentinel-2 multispectral imager is therefore hypothesized to be capable of providing timely data for the generation of high-level operational products. These include the generation of spatially explicit estimation and monitoring of important plant biophysical variables, (i.e. chlorophyll, LAI and leaf water content and crop health) in addition to producing generic land-cover, land-change detection and crop disease maps.

The operationalization of using the Sentinel-2 sensor in coffee and other plantation crops may, however, be limited. This is because the effect of disease infection is not only confused by soil background effect as influenced by age but also by other confounding factors such as co-infection with other diseases, nutrient deficiency issues, water stress among many others. This therefore means that what the sensor detects may not necessarily be the effect of the crop disease, but something else. It therefore becomes of paramount importance to determine the possibility of separating disease levels using the Sentinel-2 sensor in the absence of other potential stressors. The aim of this work was therefore to evaluate the potential of the Sentinel-2 multispectral imager derived band settings and vegetation indices in discriminating coffee leaf rust infection levels using the random forest (RF) and partial least squares discriminant

analysis (PLS-DA) algorithms at leaf level. A secondary objective was to model CLR severity using spectral settings of the Sentinel 2 MSI sensor.

## **6.2 Materials and methods**

### **6.2.1 Study area**




The study was carried out at Coffee Research Institute, Chipinge, Zimbabwe (32°37.523'E, 20°12.474'S and altitude 1,100 m above sea level). The climate in Chipinge is subtropical with two distinct seasons: the dry season and the wet season. Average total annual rainfall is 1800 mm of which 80 % fall in 5 months from November to March. The mean maximum temperature is 20°C and minimum is 14°C.

### **6.2.2 Coffee leaf rust inoculation**

Eight months old coffee seedlings (variety Yellow Catuai) were inoculated with CLR in a greenhouse for the study. CLR spores were collected from naturally infected coffee plants from a coffee field that is maintained at the station as disease reservoir in June 2015. The spores from the infected leaves were scrapped into petri dishes, using a razor blade. These spores were then used to make a spore suspension of  $8 \times 10^6$  spores/ml, using sterile deionized water as counted by a hemocytometer. Inoculation of spores was done by brushing the spore suspension at the underside of leaves, using pen brushes. The inoculated plants were incubated for 72 hours in dark incubation chambers with ~100% relative humidity. After three days in the incubation chamber, seedlings were removed, and laid on benches to allow completion of infection. Eighty plants were inoculated in two batches of 40 each in a space of two weeks, in order to get different levels of infection.

Twenty-one days after the first inoculation, coffee plants were visually scored and grouped by a plant pathologist, using visual signs into severely infected, moderately infected and healthy seedlings (no inoculation). Reflectance from twenty-one coffee plants was measured for each level of infection (healthy, moderate CLR and severely infected level ( $n = 63$ , Table 6.1). Only 21 samples were used for each class because of the poor success of inoculation associated with CLR under controlled conditions. For severity modelling, diseased area was measured using a graduated transparent polythene plate and converted to proportion infected area (%) by dividing over leaf area. The distribution of the leaf area of the leaf samples is shown in Figure 6.1a and the distribution of area diseased for the healthy, moderate and severe leaf samples are shown in Figure 6.1b.

Table 6.1: Description of levels of CLR infection levels, sample images and number of samples used in the study

Disease levels	Sample picture	Description of class
Healthy (n=21)		Healthy leaves from plants that were left non-inoculated.
Moderate infection (n=21)		Infected leaves with early or sparse spores of CLR visible on the underside of the leaves. Estimated covered area less than 10% of the leaf area
Severe infection (n=21)		Severely infected leaves with typical CLR yellowing on the underside of leaves. Covered area more than 10% of the leaf area.

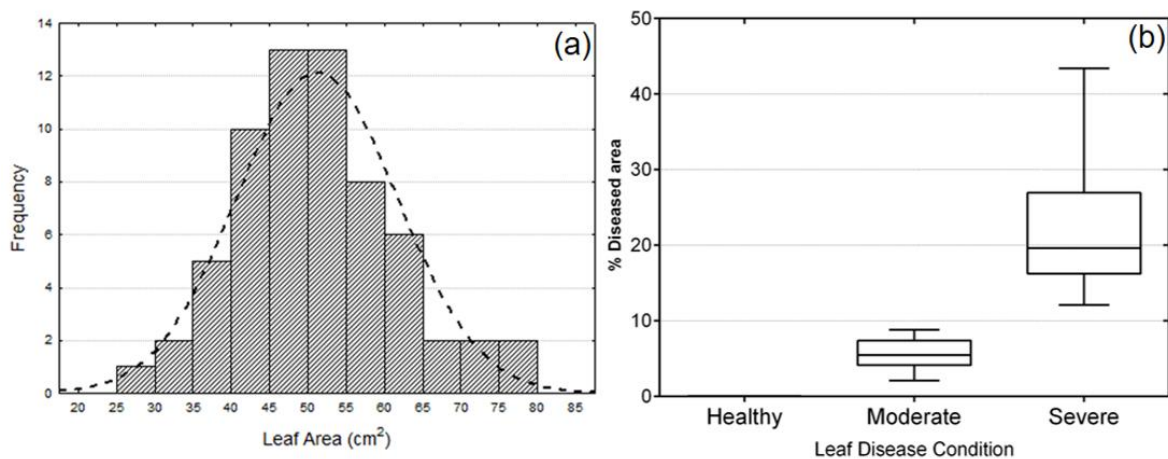


Figure 6.1: (a) The histogram showing distribution of leaf area of samples and (b) box plots of percent diseased area of leaves as measured on the day of reflectance measurements (N=63).

### 6.2.3 Reflectance measurements and resampling

Reflectance was measured using an Apogee VIS-NIR spectrometer (Apogee Instruments, Inc., Logan, UT, USA) with an effective spectral range of 400-900 nm and a spectral resolution of

0.5 nm. Each reading consisted of an average of three spectral scans, taken at 15 cm above the coffee leaf of interest at 30° angle. A white polytetrafluoroethylene (PTFE) reflectance standard was used as a reference. Reflectance by wavelength was calculated as the ratio of scene reflectance to the reflectance of the standard. The reflectance was averaged to 5 nm to reduce dimensionality.

The collected reflectance measurements were resampled to simulate the Sentinel-2 satellite sensor’s reflectance (Table 6.2). The resampling of the field spectra was done in ENVI 4.7 (ITT, 2008) software. The resampling method used applies a Gaussian model with a Full Width at Half-Maximum (FWHM) equal to the band spacing provided. The technique uses the field spectral data from the spectrometer and resamples it to the spectral width of the sensor being simulated. Only eight Sentinel-2 MSI land management bands were used, because the other bands were considered unnecessary for plant biophysical studies, had a higher spatial resolution for application in coffee or were outside the range of the spectroradiometer used in this study (Table 6.2).

Table 6.2: Specifications of the Sentinel-2 Multispectral Instrument (MSI)

Spectral band	Centre wavelength (nm)	Band width (nm)	Spatial resolution (m)
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20
B8	842	115	10

#### 6.2.4 Vegetation Indices

Seventeen spectral vegetation indices were computed and applied in evaluating transformation of spectral bands of the Sentinel-2 MSI data’s ability to detect and discriminate the different CLR infection levels (Table 6.3). These vegetation indices were selected based on their reported ability to discriminate different vegetation characteristics and conditions from remotely sensed data. The three red-edge bands contained in Sentinel-2a multispectral imager were used to generate vegetation indices that were indexed according to the particular red-edge band used (B5, B6 and B7). Vegetation indices were used in discrimination studies only and not in severity modelling.

Table 6.3: Selected vegetation indices (VIs) evaluated in the study.

Name	Formula	Sentinel-2 Bands	Source
Normalized Difference Vegetation Index	$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$	B8, B4	Rouse et al. (1973)
Simple Ratio	$SR = \frac{\rho_{NIR}}{\rho_R}$	B8, B4	Baret and Guyot (1991)
Green Chlorophyll Index	$GCI = \left( \frac{\rho_{NIR}}{\rho_{GREEN}} \right) - 1$	B8, B3	Gitelson et al. (2005)
Green Normalized Difference Vegetation Index	$GNDVI = \frac{\rho_{NIR} - \rho_{GREEN}}{\rho_{NIR} + \rho_{GREEN}}$	B8, B3	Gitelson et al. (1996)
Renormalized Normalized Difference Vegetation Index	$RNDVI = \frac{\rho_{NIR} - \rho_{RED}}{\sqrt{\rho_{NIR} + \rho_R}}$	B8, B4	Gitelson and Merzlyak (1994b)
Normalized Difference Red-edge Index	$NDVI.RE1 = \frac{\rho_{NIR} - \rho_{RE1}}{\rho_{NIR} + \rho_{RE1}}$	B8, B5	Gitelson and Merzlyak (1994b)
	$NDVI.RE2 = \frac{\rho_{NIR} - \rho_{RE2}}{\rho_{NIR} + \rho_{RE2}}$	B8, B6	Gitelson and Merzlyak (1994b)
	$NDVI.RE3 = \frac{\rho_{NIR} - \rho_{RE3}}{\rho_{NIR} + \rho_{RE3}}$	B8, B7	Gitelson and Merzlyak (1994b)
Simplified Canopy Chlorophyll Content Index	$SCCCI1 = \frac{NDVI.RE1}{NDVI}$	B8, B4, B5	Barnes et al. (2000)
	$SCCCI2 = \frac{NDVI.RE2}{NDVI}$	B8, B4, B6	Barnes et al. (2000)
	$SCCCI3 = \frac{NDVI.RE3}{NDVI}$	B8, B4, B7	Barnes et al. (2000)
Red-edge Chlorophyll Index	$CI_{RE1} = \left( \frac{\rho_{NIR}}{\rho_{RE1}} \right) - 1$	B8, B5	Gitelson et al. (2005)
	$CI_{RE2} = \left( \frac{\rho_{NIR}}{\rho_{RE2}} \right) - 1$	B8, B6	Gitelson et al. (2005)
	$CI_{RE3} = \left( \frac{\rho_{NIR}}{\rho_{RE3}} \right) - 1$	B8, B7	Gitelson et al. (2005)
Normalized Red-edge Difference Index	$NREDI1 = \frac{\rho_{RE3} - \rho_{RE1}}{\rho_{RE3} + \rho_{RE1}}$	B7, B5	Gitelson and Merzlyak (1994a)
	$NREDI2 = \frac{\rho_{RE3} - \rho_{RE2}}{\rho_{RE3} + \rho_{RE2}}$	B7, B6	Gitelson and Merzlyak (1994a)
	$NREDI3 = \frac{\rho_{RE2} - \rho_{RE1}}{\rho_{RE2} + \rho_{RE1}}$	B6, B5	Gitelson and Merzlyak (1994a)

## **6.2.5 Statistical approaches for CLR discrimination**

### **6.2.5.1 Random forest algorithm.**

The random forest (RF) ensemble algorithm developed by Breiman (2001) was one of the two robust machine learning approaches used for CLR discrimination with raw spectral bands and vegetation indices. The classification version of RF (Breiman & Cutler, 2007) together with in-built variable optimization was used in this analysis. The concept of the random forest algorithm is described in detail by Breiman (2001) and Lebedev et al. (2014). The default number of trees (501) and the square root of the number of variables as *mtry* were used as RF settings in the randomForest library in R (Liaw et al., 2009). The RF variable importance measure was used to select bands and vegetation indices that were used in CLR discrimination. The RF algorithm assesses the importance of each input variable to the outcome by comparing how much the OOB error increases when a variable is removed, while all others are left unchanged (Gislason et al., 2004; Breiman & Cutler, 2007). This way, the RF ranks the variables according to the mean decrease in error when that variable is included in the modelling and those variables with higher mean decreases in error are the most important variables for the modelling and should therefore be retained.

### **6.2.5.2 Partial least squares discriminant analysis**

In addition to the RF approach, partial least squares discriminant analysis (PLS-DA) algorithm was used in CLR discrimination. PLS-DA is the classification version of PLS regression and is a powerful multivariate supervised pattern recognition method that uses a training routine to assign class membership to variables based on their known statistical parameters projected into latent variables (Wang et al., 2011; Venteris et al., 2015). It is very important in PLS-DA to determine the appropriate number of components for the model to avoid overfitting and this was done using cross-validation (Wang et al., 2011). The appropriate number of components was identified as 2 for spectral and 3 for vegetation indices and these were used accordingly. Variable optimization in PLS-DA was done through the Variable Importance in Projection (VIP) scores produced in the discrimination. The VIP is a quantitative estimate of the discriminatory power of each individual variable used in the model and as such can be used to interpret the outcome in relation to the input variables (Vancutsem et al., 2009).

## **6.2.6 Statistical approaches for severity modelling**

The radial basis function partial least squares regression (RBF-PLS) was used to model CLR severity from Sentinel-2 spectral bands. The RBF-PLS is part of the kernel learning family of

algorithms. These learning methods are based on mapping the originally observed data into a high-dimensional feature space where simple linear models are then constructed (Rosipal, 2010). The concept of the RBF-PLS is described in detail in literature (Orr, 1996; Yan et al., 2004; Jia et al., 2010; Jiang et al., 2013). The RBF-PLS was implemented in Matlab using the TOMCAT toolbox (Daszykowski et al., 2007).

## **6.2.7 Accuracy assessment**

### **6.2.7.1 CLR discrimination**

In order to assess the performance of CLR discrimination, k-fold cross validation with 10 folds was used since the sample number was relatively small ( $n = 63$ ) for sub-setting the data into training and test data. A confusion matrix, defined as a table that describes the performance of a discriminating model on a set of test data for which the true values are known, was used to evaluate the overall accuracy of the random forest discrimination of CLR infection levels (i.e. healthy, moderate and severe CLR infection). Overall accuracy, Kappa ( $k$ ) and related class user's and producer's accuracies were calculated to evaluate the performance of Sentinel-2 MSI derived spectral bands and vegetation indices in discriminating CLR, using the RF and PLS-DA algorithms. The effect of variable optimization was determined by the McNemar's test (Foody, 2004) of confusion matrices of discrimination with all variables versus discrimination with optimized variables.

### **6.2.7.2 Severity modelling**

In order to assess the performance of CLR severity predictions, k-fold cross validation with 100 folds was used since the sample number was relatively small ( $n = 63$ ) for sub-setting the data into training and test data. The correlation coefficient ( $r$ ) and coefficient of determination ( $R^2$ ) were used to assess the goodness of fit of the predicted and measured CLR severity values. In addition, Mean Absolute Error (MAE, Equation 5.4) Root Mean Square Error (RMSE, Equation 5.5), and percent bias (pBias, Equation 5.6) were used to determine the errors of the model in predicting CLR severity from variables.

## **6.3 Results**

### **6.3.1 Spectral Resampling**

Figure 6.2 shows the mean reflectance of Sentinel-2 bands for the healthy, moderate and severe CLR samples used in the study. As expected with vegetation, the results show that coffee leaf



reflectance was higher in the NIR region of the spectrum when compared to the visible spectral region. Healthy leaves produced higher reflectance in B3 (green) and the least in two of the three Sentinel-2 red-edge bands (B5 and B6) where the severely CLR infected leaves produced the highest reflectance. The reflectance of the severely infected CLR was least in the red-edge 3 band (B7) and the NIR band (B8).

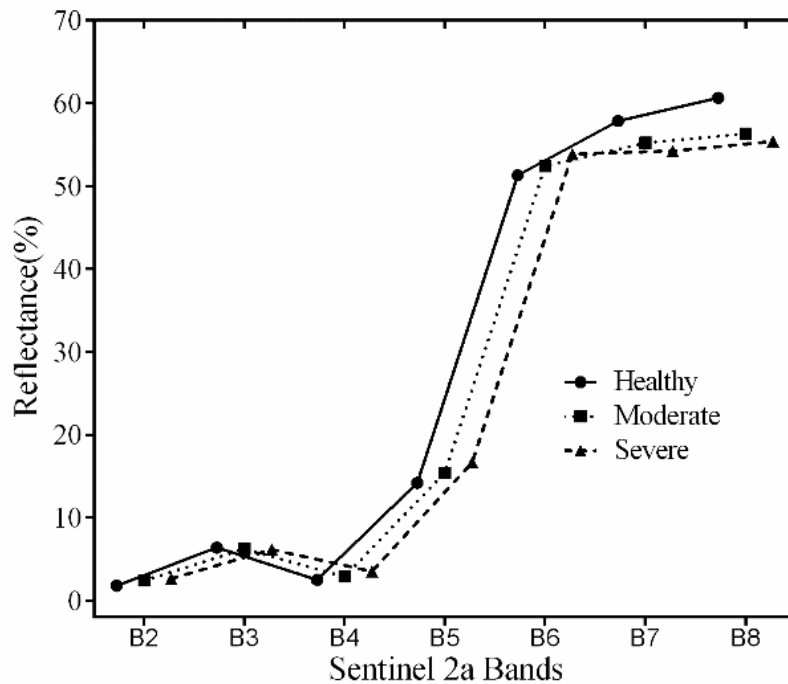


Figure 6.2: Mean spectral reflectance of CLR infection levels across the Sentinel-2 bands obtained from resampling hyperspectral imagery.

## 6.3.2 CLR discrimination

### 6.3.2.1 Discrimination with all variables

The results from CLR discrimination with all variables ( $N = 7$  for bands and  $N = 17$  for vegetation indices) are shown in Table 6.4. When Sentinel-2 MSI seven bands were used with the PLS-DA, an overall accuracy of 63.5% ( $k=0.45$ ) was achieved while this increased to 69.8% ( $k = 0.55$ ) from the use of the RF algorithm. Results indicate spectral confusion in discriminating between the moderate and severe CLR when compared to the discrimination between healthy and moderate CLR for both algorithms. Higher accuracies in discriminating CLR were achieved by using all vegetation indices (PLS-DA =68.3% and RF = 76.2%) when compared to that obtained from use of all spectral bands (PLS-DA =63.5% and RF = 69.8%). However, notable misclassifications were observed when vegetation indices were used between moderate CLR (producer accuracy = 66.7% for RF and 47.6% for PLS-DA) and severe CLR classes (producer accuracy = 71.4% for RF and 66.7% for PLS-DA).

Table 6.4: CLR discrimination accuracies obtained using Sentinel-2 MSI derived spectral bands and vegetation indices

All Bands (N=7)	RF					All Bands (N=7)	PLS-DA				
	Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy		Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy
Healthy	17	3	1	21	81.0	Healthy	19	2	0	21	90.5
Moderate CLR	3	13	5	21	61.9	Moderate CLR	4	9	8	21	42.9
Severe CLR	2	5	14	21	66.7	Severe CLR	2	7	12	21	57.1
Total	22	21	20	63		Total	25	18	20	63	
User's Accuracy	77.3	61.9	70.0			User's Accuracy	76.0	50.0	60.0		
Overall Accuracy	69.8					Overall Accuracy	63.5				
Kappa	0.55					Kappa	0.45				

All VIs (N=17)	RF					All VIs (N=17)	PLS-DA				
	Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy		Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy
Healthy	19	0	2	21	90.5	Healthy	19	2	0	21	90.5
Moderate CLR	3	14	4	21	66.7	Moderate CLR	4	10	7	21	47.6
Severe CLR	2	4	15	21	71.4	Severe CLR	0	7	14	21	66.7
Total	24	18	21	63		Total	23	19	21	63	
User's Accuracy	79.2	77.8	71.4			User's Accuracy	82.6	52.6	66.7		
Overall Accuracy	76.2					Overall Accuracy	68.3				
Kappa	0.64					Kappa	0.52				

### 6.3.2.2 Variable optimization

The results in Figure 6.3 illustrates the most important selected model variables (i.e. bands and vegetation indices) derived using the RF OOB and variable importance in projection (VIP) classification ensembles. From the results in Figure 6.3a and 6.3b, it can be observed that the RF model identified four spectral bands (B4, B6, B3, B7) whereas PLSA-DA selected only three (B4, B6 and B5). Comparatively, both models managed to select two identical variables (i.e. B4 and B6 bands) as the most important variables for discriminating CLR levels (Figure 6.3a and 6.3b). Further, when the variable optimization was implemented using the seventeen derived vegetation indices, only seven were selected as important by RF and most of these were transformations of Sentinel-2 red-edge bands and the NIR (Figure 6.3c). On the other hand, the PLS-DA algorithm identified only five vegetation indices (i.e. SCCCI3 (B8, B7 & B4), CI<sub>RE1</sub> (B8 & B5), RNDVI (B8 & B4), NREDI2 (B7 & B6), GCI (B8 & B3)) as having a VIP above 1. Only four of the vegetation indices were identically selected by both the methods (SCCCI3, CI<sub>RE1</sub>, RNDVI and GCI).

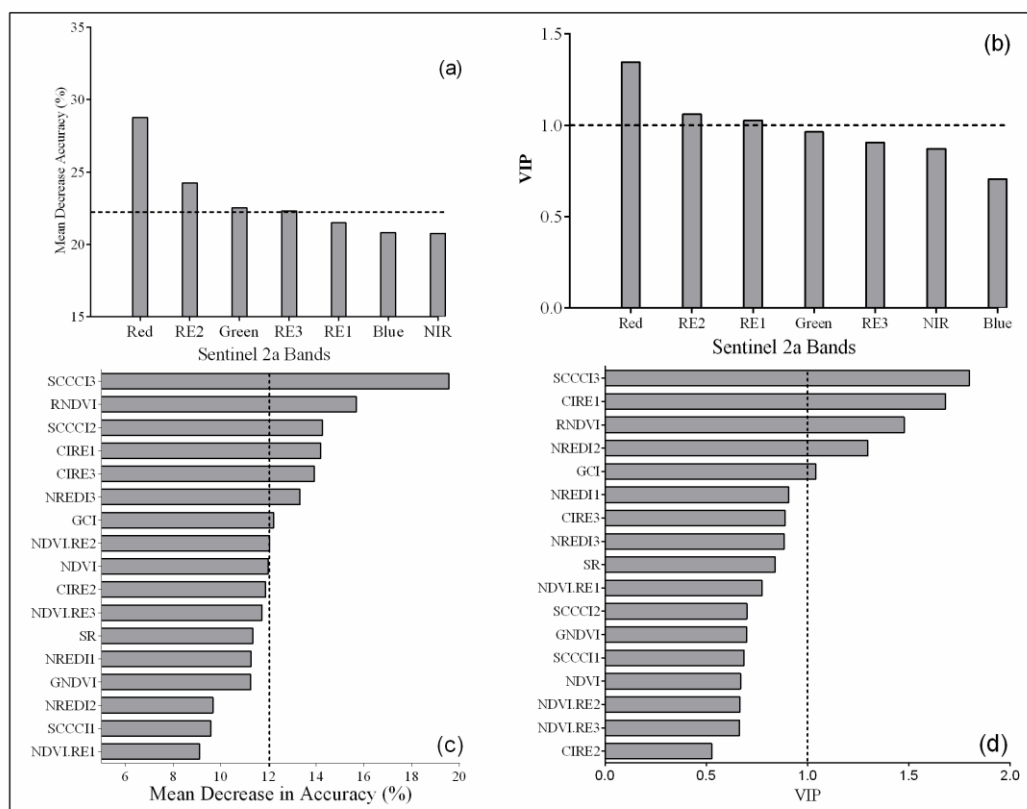


Figure 6.3: Optimization of Sentinel-2 variables for CLR discrimination through (a) RF-OOB error for spectral bands (b) PLS-DA for spectral bands (c) RF-OOB error for vegetation indices and (d) PLS-DA VIP for vegetation indices. The dotted line shows the cut-off point for variables.

### 6.3.2.3 CLR discrimination with optimized variables

Table 6.5 shows the results of CLR discrimination using optimized variables. The results indicate that there is an improvement in accuracy of CLR discrimination when optimized variables are used, with the highest magnitude of change observed in optimized vegetation indices (Table 6.5). Overall, optimized Sentinel-2 derived vegetation indices achieved the highest overall CLR discrimination accuracies of 82.5% ( $k = 0.74$ ) and 71.4% ( $k = 0.57$ ) using the RF and PLS-DA algorithms respectively. It can be noted that CLR discrimination accuracies increased by about 10% for RF algorithm. On the other hand, optimization of Sentinel-2 derived vegetation indices from 17 to 5 variables improved the performance of the PLS-DA model in CLR discrimination, particularly in terms of the producer accuracy of the class severe and all user accuracies.

In general, the finding of this work further demonstrates that the RF outperformed the PLS-DA in CLR discrimination using Sentinel-2 with and without model optimization (Figure 6.4). The differences in accuracy due to optimization was also assessed through McNemar's test (Table 6.6), which showed that variable optimization significantly improved the accuracy of CLR discrimination with spectral bands run through the RF algorithm ( $\chi^2 = 4.17$ ,  $p < 0.05$ ).

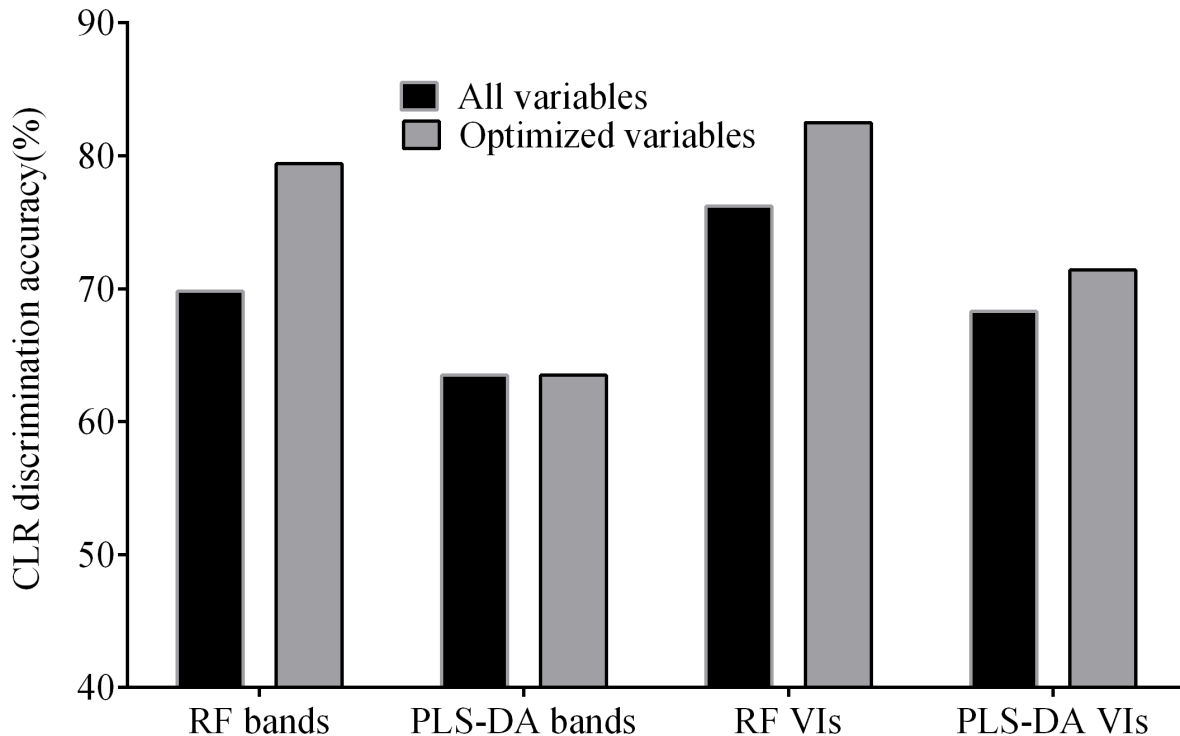


Figure 6.4: Effect of RF and PLS-DA model variable optimization on CLR discrimination accuracies using Sentinel-2 bands and vegetation indices.

Table 6.5: CLR discrimination accuracies derived using the most important selected model variables (i.e. spectral bands and optimized vegetation indices).

Optimized Bands (N=4)	RF					Optimized Bands (N=3)	PLS-DA				
	Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy		Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy
Healthy	20	1	0	21	95.2	Healthy	19	1	1	21	90.5
Moderate CLR	3	13	5	21	61.9	Moderate CLR	3	9	9	21	42.9
Severe CLR	1	3	17	21	81.0	Severe CLR	3	6	12	21	57.1
Total	24	17	22	63		Total	25	16	22	63	
User's Accuracy	83.3	76.5	77.3			User's Accuracy	76.0	56.3	54.6		
Overall Accuracy	79.4					Overall Accuracy	63.5				
Kappa	0.69					Kappa	0.45				

Optimized VIs (N=7)	RF					Optimized VIs (N=5)	PLS-DA				
	Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy		Healthy	Moderate CLR	Severe CLR	Total	Producer's Accuracy
Healthy	19	2	0	21	90.5	Healthy	18	3	0	21	85.7
Moderate CLR	1	14	6	21	66.7	Moderate CLR	2	11	8	21	52.4
Severe CLR	1	1	19	21	90.5	Severe CLR	1	4	16	21	76.2
Total	21	17	25	63		Total	21	18	24	63	
User's Accuracy	90.5	82.4	76.0			User's Accuracy	85.7	61.1	66.7		
Overall Accuracy	82.5					Overall Accuracy	71.4				
Kappa	0.74					Kappa	0.57				

Table 6.6: Comparison of the performance of CLR discrimination accuracy results derived using RF and PLS-DA models with and without variable optimization

Parameters	f <sub>11</sub>	f <sub>12</sub>	f <sub>21</sub>	f <sub>22</sub>	Total	$\chi^2$	P-value (95%)
RF Bands	44	0	6	13	63	4.17	0.041
RF VIs	48	0	4	11	63	2.25	0.133
PLS-DA Bands	37	3	3	20	63	0.00	1.000
PLS-DA VIs	41	2	4	16	63	0.17	0.683

### 6.3.3 CLR severity modelling

#### 6.3.3.1 Relationship between Sentinel-2 MSI spectral bands and CLR severity

The results showed that there were weak correlations between CLR severity and Sentinel-2 MSI spectral parameters (Table 6.7). Only three out of the seven bands were significantly ( $p < 0.05$ ) correlated with CLR severity (Blue, Red and RE1). In both of these cases, the correlations were positive. Results showed very poor linear relationships between CLR severity and Sentinel-2 MSI spectral settings as all bands had weak correlations with CLR severity ( $r < 0.5$ , Figure 6.5). From the correlation plots of the significant variables in Figure 6.5, it is difficult to use linear modelling to relate CLR severity with Sentinel-2 MSI spectral parameters. Leaf size did not have any significant influence on reflectance as leaf area was not significantly correlated ( $p > 0.05$ ) with any of the Sentinel-2 MSI bands. Interestingly, all spectral bands were significantly correlated to each other, presenting potential challenges of collinearity in the model if linear methods are applied (Table 6.7).

Table 6.7: Correlation coefficients (r) and significance of correlation ( $\alpha < 0.05$ ) between Sentinel-2 MSI bands and CLR severity.

	Severity (%)*	Leaf Area	Blue	Green	Red	RE1	RE2	RE3	NIR
Severity (%)	1	-0.127	<b>0.264</b>	-0.102	<b>0.331</b>	<b>0.254</b>	-0.049	-0.101	-0.106
Leaf Area	0.320	1	0.026	0.118	0.047	0.116	0.145	0.126	0.096
Blue	<b>0.036</b>	0.839	1	0.835	0.954	0.749	0.550	0.468	0.517
Green	0.426	0.355	<0.001	1	0.755	0.858	0.760	0.632	0.616
Red	<b>0.008</b>	0.711	<0.001	<0.001	1	0.751	0.497	0.422	0.473
RE1	<b>0.044</b>	0.363	<0.001	<0.001	<0.001	1	0.607	0.453	0.474
RE2	0.697	0.256	<0.001	<0.001	<0.001	<0.001	1	0.974	0.931
RE3	0.432	0.323	<0.001	<0.001	<0.001	<0.001	<0.001	1	0.973
NIR	0.409	0.454	<0.001	<0.001	<0.001	<0.001	<0.001	0<0.001	1

\*Correlation coefficients with bold letters were significant ( $\alpha < 0.05$ )

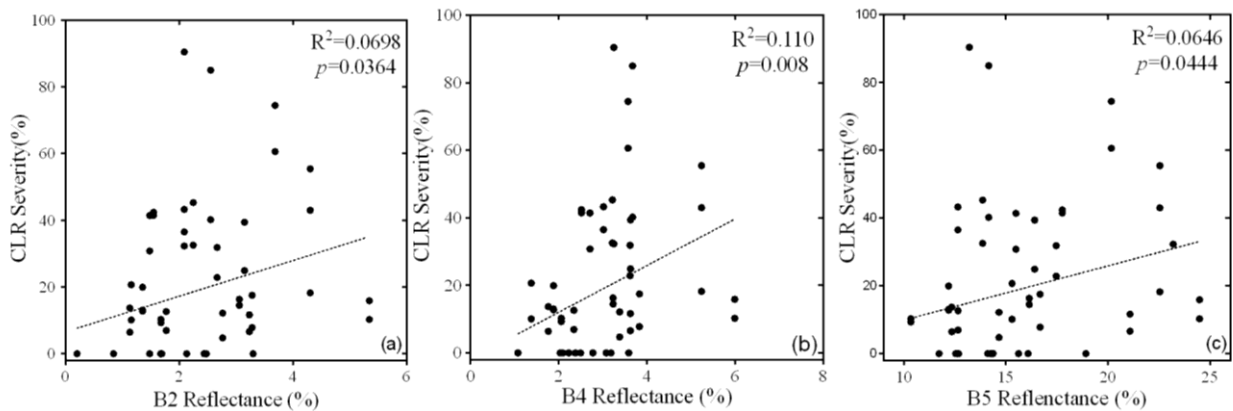


Figure 6.5: Correlation between CLR severity and Sentinel-2 MSI (a) Band 2 reflectance, (b) Band 4 reflectance and (c) Band 6 Reflectance.

### 6.3.3.2 Determining gaussian widths

Figure 6.5 shows the relationship between cross-validated RMSE and  $\sigma$  representing Gaussian widths. The results show that the best Gaussian width for the RBF-PLS model is 0.3 as it produces the least cross-validated training RMSE of 6.7 for all variables (Figure 6.6a). The model therefore used this as the  $\sigma$  for the model. For selected variables, however, the best Gaussian width was lower at 0.2 but with a minimum cross-validated RMSE of 6.8 being higher than that of all variables (Figure 6.6b). The learning process thus produced different results for developing the model with all variables and with selected variables.

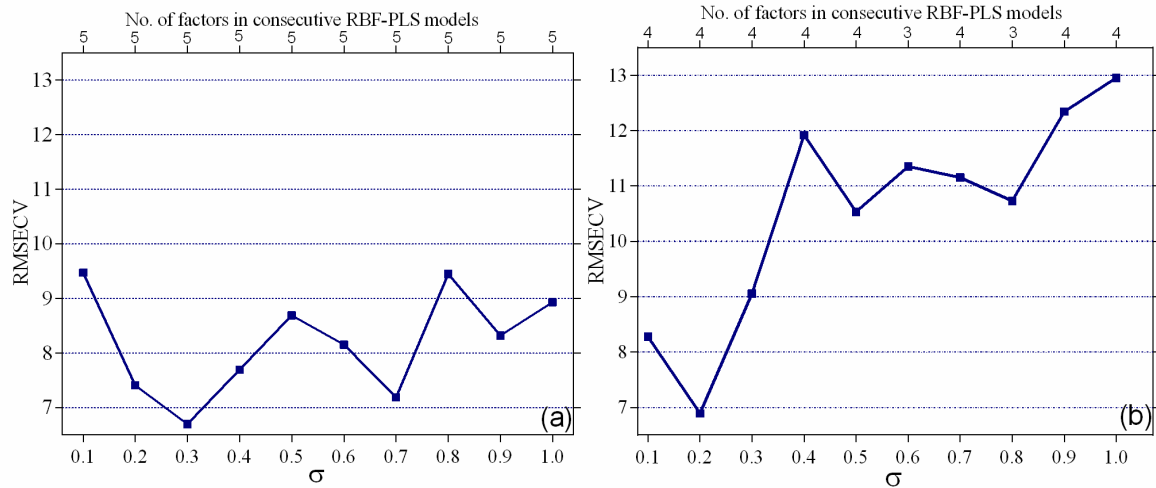


Figure 6.6: Determination of  $\sigma$  through cross-validated RMSE for use as Gaussian widths in the non-linear RBF-PLS models for (a) All Sentinel-2 MSI bands and (b) Selected Sentinel-2 MSI variables.

### 6.3.3.3 Modelling CLR severity

The RBF-PLS was able to satisfactorily model CLR severity at Sentinel 2 spectral settings. Using all Sentinel-2 MSI bands as variables was able to explain 92% of the variance in CLR severity using RBF-PLS (Figure 6.7a, Table 6.8). On the other hand, the use of model selected variables reduced the accuracy of the modelling ( $R^2=0.78$ ,  $RMSE=10.2$ , Figure 6.7b, Table 6.8). A comparison shows that using all Sentinel-2 MSI variables outperforms the use of a few model selected variables in modelling CLR ( $p<0.05$ ). Although the results show that both models are good, there is a general indication that moderate levels of infection are more difficult to predict than severe using all Sentinel-2 MSI bands and only significantly correlated Sentinel-2 MSI bands ( $R^2=0.71$  for severe and  $R^2=0.38$  for moderate bands, Table 6.8).



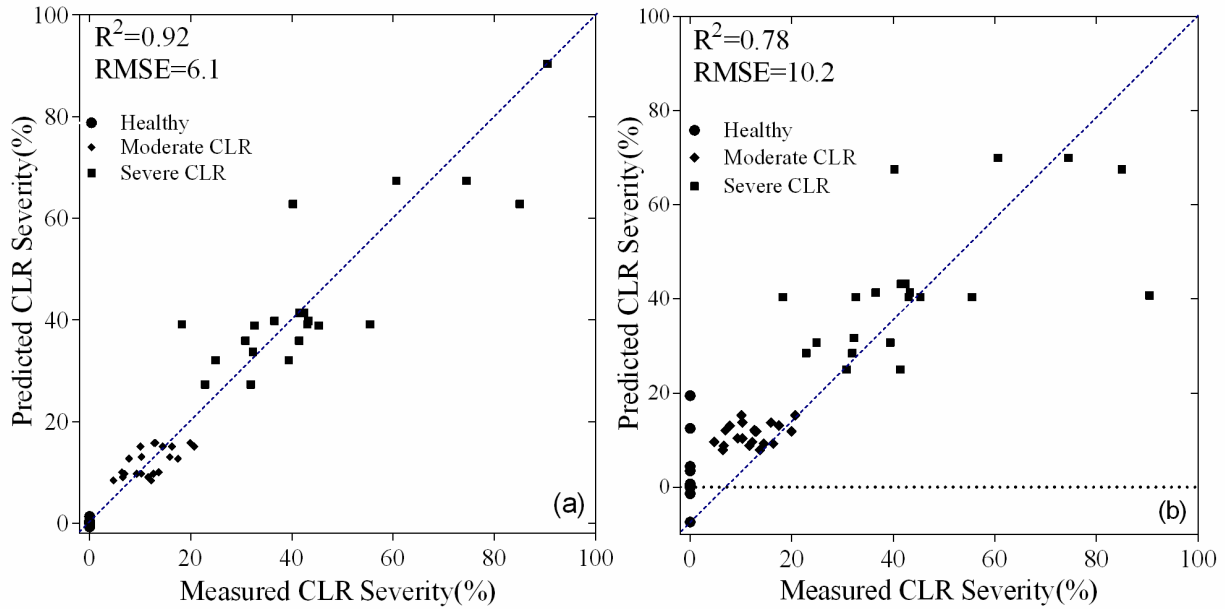


Figure 6.7: Relationship between measured and modelled CLR severity with RBF-PLS with (a) All Sentinel-2 MSI bands and (b) Selected Sentinel-2 MSI variables.

Table 6. 8: Error metrics for all models in predicting CLR severity from Sentinel-2 MSI variables with RBF-PLS

Method	Infection Levels	MAE	RMSE	pBias	R <sup>2</sup>
RBF-PLS: All	Moderate CLR	3.05	3.37	-0.1	0.38
	Severe CLR	7.43	10.06	0.0	0.71
	All levels	3.58	6.13	-	0.92
RBF-PLS: Selected	Moderate CLR	3.56	4.2	-6.8	0.11
	Severe CLR	10.3	15.31	-4.9	0.37
	All levels	6.18	10.18	-	0.78

## 6.4 Discussion

This study aimed to evaluate the spectral separability of coffee leaf rust infection levels and predict its severity using machine learning methods and spectral data at Sentinel-2 MSI resolutions as a basis for field disease monitoring and modelling for coffee condition assessment with remote sensing. To achieve this objective, two machine-learning algorithms were applied on un-optimized and optimized Sentinel-2 MSI spectral bands and vegetation indices for CLR discrimination while one was applied for CLR severity modelling. Accurate discrimination of

CLR infection levels and prediction of severity provides the most required knowledge for precise crop diseases monitoring and controlling with remote sensing, especially if high crop productivity is to be realized.

#### **6.4.1 Spectral resampling**

The resampled spectra demonstrated the general response of vegetation to electromagnetic radiation with lower reflectance in the visible region and higher reflectance in the NIR regions. This research confirms the effect of biotic stress on vegetation reflectance observable by the blue shift in the red-edge region of the spectrum where infected samples produced higher reflectance, when compared to healthy samples. The blue shift as indicated by the shift of the slope of stressed plants is associated with a decline in leaf chlorophyll quantity and quality (Ustin et al., 2009) and can be used as an early indicator of disease infection, allowing for control measures to be more effective. This phenomenon has been reported due to plant diseases (Prabhakar et al., 2013), nutrient stress (Mutanga & Skidmore, 2007), pest attacks (Stone et al., 2001) and other physio-chemical plant disorders (Carter & Knapp, 2001; Roy et al., 2014).

#### **6.4.2 Discrimination**

##### **6.4.2.1 Performance of Sentinel-2 spectral bands in CLR discrimination**

This study has for the first time managed to demonstrate the utility and strength of the new generation Sentinel-2 MSI sensor in discriminating disease infections in commercial crops such as coffee- a previous challenging task from broadband multispectral sensors (e.g. Landsat series). For example, the results of this study have successfully managed to identify Sentinel-2 MSI's most important individual spectral bands (i.e. B4, B6 and B5) required for the accurate discrimination of CLR infection levels. The results from the RF OOB score show that the Sentinel-2 RE3 band (B7) is the most significant variable for CLR discrimination. This confirms the importance of this region of the spectrum in fungal disease detection and discrimination in vegetation as reported in previous studies (Mahlein et al., 2013; Bhojaraja et al., 2015).

The finding that the red band (B4) was ranked by both the RF OOB score and the PLS-DA VIP score as the most significant band was rather surprising. Although all red-edge bands (B5, B6 & B7) were identified as important by RF OOB score, they were behind the red band in significance while only RE2 (B6) and RE1 (B5) were selected by the PLS-DA VIP score. This could be because CLR is different from other diseases and plant stress conditions in that it is not necrotic

(Belan et al., 2015) and also result in distinctive asymptomatic rustic pustules on the underside of the leaf, which perhaps are responsible for the red band being significant in its discrimination. While studies have generalized the approaches for remote sensing plant diseases, it is clear from this work that the application of spectral bands in disease discrimination could be disease and plant specific, given the observed peculiarities. Most interestingly, in vegetation reflectance studies, the red band is associated with high absorption of radiation because of pigments, such as phycocyanin and phycoerythrin and not its reflectance. More studies are therefore required to verify the effects of CLR infection on the leaf content of these pigments in order to explain the observed significance of the Sentinel-2 red band.

#### **6.4.2.2 Performance of Sentinel-2 vegetation indices in CLR discrimination**

The general finding from this study was that the use of spectral vegetation indices produced plausible CLR discrimination results, when compared to the use of the sensor's raw spectral bands. This observation is in line with previous findings that concluded that the spectral transformation that occurs in converting spectral bands to vegetation indices enables more information to be obtained resulting in better performance from vegetation indices. For example, Dube and Mutanga (2015) achieved higher model fit in biomass estimation using vegetation indices ( $R^2=0.53$ ) than by spectral bands ( $R^2=0.40$ ) with the RF algorithm. Similarly, other studies reported better performance of vegetation indices than spectral bands (Wang et al., 2004; Dube & Mutanga, 2015).

In addition, the better performance of vegetation indices could also be attributed to the ability of Sentinel-2 MSI derived vegetation indices to deal with confounding factors, such as reflectance saturation, leaf area, roughness and moisture in the leaf and canopy level that reduce the performance of raw spectral bands. It is known that narrow band vegetation indices as the ones obtainable from Sentinel-2 MSI sensor are capable of reducing the effects of asymptotic saturation common in raw reflectance and broadband vegetation indices (Mutanga & Skidmore, 2004; Baret & Buis, 2008; Glenn et al., 2008). This could be because Sentinel-2 MSI derived vegetation indices are more sensitive to plant biochemical and biophysical properties, as they are a combination of two or more strategically positioned spectral bands solely designed for vegetation condition assessment. For instance, the simplified canopy chlorophyll content index (SCCCI3) that was considered most useful for CLR discrimination is a second stage

transformation of the normalized difference vegetation index and the normalized difference red-edge index that are based on the NIR (B8) and red-edge (B7) Sentinel-2 bands.

#### **6.4.2.3 Effect of variable optimization and algorithm on CLR discrimination**

Results have also shown that RF and PLS-DA algorithms variable optimization improves the accuracy of CLR discrimination, when compared to model implementation without variable selection. There are many reasons why variable optimization is able to achieve better results than using all available variables. These multispectral sensors have been developed for wide purposes ranging from forestry, agricultural, water and urban applications (although Sentinel-2 has a bias towards vegetation) and therefore, a few may be fit for purpose. Although the wavebands in multispectral data have been optimally selected, there remains a need to find the optimal model parameters from the feature subset, as there is no guarantee that the parameters for the full feature set are equally optimal for the optimal feature subset, which are the selected multispectral bands and vegetation indices (Saeys et al., 2007).

The effect of variable selection/optimization was even more remarkable in the RF algorithm that is eulogized for its robustness to noise, ability to handle non-linear data and deal with huge numbers of variables (Gislason et al., 2004; Mutanga et al., 2012; Rodriguez-Galiano et al., 2012). These characteristics of the RF explain the observed superior performance of the RF over PLS-DA, making it a better candidate for use in CLR discrimination with recently launched Sentinel-2 MSI data. This is important because in order to improve the effectiveness of remote sensing applications, retrieval algorithms should be accurate, fast, robust, and sufficiently flexible to make use of the Sentinel-2 spectral bands and vegetation indices (Fassnacht et al., 2015). It appears from this study that the RF algorithm is fit for this purpose with an added advantage of explaining how the output is obtained through ranking the weights of input variables.

#### **6.4.3 Severity modelling**

##### **6.4.3.1 Relationship between Sentinel-2 MSI spectral bands and CLR severity**

The finding that CLR severity as measured by proportion of leaf diseased area was not significantly correlated to most spectral bands and weakly so for bands where results were significant, was not at all surprising. This could be because of the non-linear influence of disease infection on spectral reflectance (Zhang et al., 2012; Mahlein et al., 2013). Thus, when the

reflectance is averaged across a spectral width, this direct influence is lost, resulting in non-linear relationships. Even for the specific narrow-band indices obtained from hyperspectral indices, this linear relationship is also not always obvious because of other confounding factors that may influence the reflectance. For example, it is known that the CLR infection has significant influence on nutrient distribution within the leaf structure (Belan et al., 2015). The effect of the disease on the nutrient distribution will then result in nutritional composition having direct influences on reflectance. Other studies have also shown that disease and pest incidence have significant influences on leaf water content (Mutanga & Ismail, 2010; Oumar & Mutanga, 2014), which in turn influences water absorption features. The influence of CLR severity on water content may explain the significance of the blue band in this study. More studies are however required to determine the exact relationship between CLR severity and leaf water content as this relationship is disease specific. The role of the red and red-edge bands that were found important in this study in condition assessment has been reported widely (Rumpf et al., 2010; Eitel et al., 2011).

#### **6.4.3.2 Modelling CLR severity with spectral bands**

Results of this study showed that using only spectral bands on the RBF-PLS showed a high accuracy ( $R^2 = 0.92$ ), which is an interesting achievement. This is so because much of the reported good results in vegetation condition modelling have been from vegetation indices or at least a combination of vegetation indices and spectral bands, even with hyperspectral data. Although vegetation indices produce good results, many vegetation indices are derivatives of the NIR or the red-edge bands and putting more than one index in a model results in serious over-fitting which eliminates transferability of the model. For example, the normalised difference vegetation index, simple ratio, renormalized normalized difference vegetation index and simplified canopy chlorophyll index all use the Sentinel MSI B8 and B4. Thus developing a model with all these vegetation indices will likely produce an unstable model. Therefore, having a band-based model producing this level of accuracy is important as each individual parameter is unique in their contribution to the outcome. Ramoelo et al. (2015) also used only spectral bands to model nitrogen in rangelands and obtained a high accuracy ( $R^2=0.90$ ,  $RMSE=0.04$ ), confirming that the spectral settings of Sentinel-2 MSI spectral settings are good for modelling vegetation condition. Similar bands (except for B2) were identified as important in CLR

discrimination with RF and PLS-DA, confirming that these are the bands that can be used to perform both discrimination and modelling of CLR severity with Sentinel-2 MSI data.

Results showed that the correlation between modelled and measured values was higher for severe levels of CLR compared to moderate levels when the RBF-PLS was used. This indicates better model performance for severe CLR infection level than on moderate levels. These observations can be explained by the fact that on more serious levels of infection, there is little spectral confusion as effects will be distinct compared to moderate levels as observed in the discrimination study. However, from a practical application viewpoint, it is more useful to be able to predict the CLR at moderate levels because control measures can be implemented successfully (Carter & Miller, 1994; Rumpf et al., 2010). At severe levels, the leaves will most definitely be shaded and thus economic impact would have been inflicted.

#### **6.4.4 Multispectral level remote sensing of biotic stress in coffee**

CLR discrimination and severity modelling with Sentinel-2 data could inform farm managers and plant phytosanitary regulators on occurrence and levels of severity of plant diseases, reducing the subjectivity, costs, spatial singularity and inconveniences associated with field methods. In addition, the high temporal resolution of Sentinel-2 can provide opportunities for modelling CLR rate and direction of spread by combining analysis in both time and space over fields. Therefore, Sentinel-2 MSI can freely bring information on plantation crop condition that the current crop of multispectral sensors could not. Despite the fact that these results reported here are based on resampled data, they actually provide the required impetus for field application of the Sentinel-2 data in disease discrimination, providing opportunities for the required regionalized data in coffee condition assessment. Notwithstanding these findings, there is need for additional research to bridge the gap between the handheld device used in this study and the satellite platform that it is mimicking. It is a gigantic leap of faith to assume that leaf level studies such as these could be interpreted in terms of actual satellite data performance at field level.

#### **6.5 Conclusion**

The aim of this work was to explore the utility of the Sentinel-2 multispectral imager in detecting, discriminating and modelling the severity of coffee leaf rust (*Hemileia vastatrix*) at leaf level. Specifically, this work has demonstrated that Sentinel-2 MSI derived bands and vegetation indices computed using spectral information located the red-edge position are useful for crop

diseases discrimination and modelling that is the basis for coffee condition assessments. Also, the study has shown that optimized vegetation indices and spectral bands perform better in CLR discrimination when compared to the use of all variables as an independent dataset while this is not so for modelling the severity of CLR with RBF-PLS. This study therefore underpins the application of Sentinel-2 MSI data in crop and vegetation state assessment that can improve management of croplands and stewardship of the environment through reduced unnecessary use of crop protection chemical for disease control. Although the results are positive in indicating potential application as sensors in disease levels modelling, factors such as canopy structure and distribution of the disease across the canopy have to be considered for practical application. There is need for more field studies to apply the RF and RBF-PLS in modelling the biophysical (e.g. LAI reduced to defoliation) and biochemical (e.g. chlorophyll and foliar nitrogen) parameters of coffee that are affected by the disease. Notwithstanding the positive outcome of these studies, the transfer of models from leaf level to canopy level is not easy and direct. Therefore, these results should only be considered for leaf level assessment with more studies required to transfer leaf level assessments to canopy level applications.

## **6.6 Link to next chapter**

The two chapters in this section (Chapter 5 and Chapter 6) sought to experimentally evaluate the potential for remote sensing individual stressors that affect coffee condition at leaf level. Although the results are very positive, there are many challenges in operation upscaling of leaf-level models to canopy level models. There is evidence that coffee stressors have a significant bearing on coffee foliar biochemical properties particularly chlorophyll and nitrogen and therefore these can be used as proxies for coffee condition, whatever the individual factor contributing to them. The next chapter therefore assess the potential for modelling coffee leaf chlorophyll using multispectral remote sensing data at field level.

**SECTION 4: EMPIRICAL MODELLING OF COFFEE  
CONDITION USING MULTISPECTRAL REMOTE SENSING  
DATA**



## CHAPTER 7: LANDSCAPE SCALE MULTISPECTRAL REMOTE SENSING OF COFFEE TOTAL CHLOROPHYLL CONTENT



Photo credits: A. Chemura (2017)

### **This chapter is based on:**

**Chemura, A., Mutanga, O., and Odindi, J. (2017)** Empirical modelling of leaf chlorophyll content in coffee (*Coffea arabica*) plantations with Sentinel-2 MSI data: Effects of spectral settings, spatial resolution and crop canopy cover. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (**In Press**).

## **Abstract**

Coffee leaf chlorophyll (Chl) is an important proxy for coffee plant photosynthetic rates, nitrogen content, leaf health and yield potential. Therefore, approaches to monitor coffee leaf Chl are necessary for field, landscape, regional and national scale decision making to improve coffee stand health and productivity and to reduce yield losses. Leaf chlorophyll is among the many plant biochemical parameters that can be non-destructively mapped with remote sensing. Whereas the recently launched Sentinel-2 *multi-spectral instrument* (MSI) data has great potential for plant condition assessment, the value of its spectral settings at variable spatial resolutions in relation to crop canopy cover on Chl content prediction remains largely unexplored. In this study, an empirical model to estimate coffee leaf Chl with Sentinel-2 MSI data was applied. Specifically, the influence of spectral settings, spatial resolution and age-related crop canopy cover on modelling performance was investigated. The random forest algorithm was used to predict coffee leaf Chl using all 9 Sentinel-2 MSI bands at 20m, all 9 bands at 10m, 5 bands at 10m and 4 bands at 20m spatial resolutions for all coffee stands (3-8 years, N=72) and then for mature stands only (5-8 years, N=60). Results showed that coffee biophysical parameters (height and canopy cover) are significantly influenced by stand age while biochemical parameters (plant water concentration (PWC) and total Chl) are age invariant. Results further showed that the best modelling results ( $R^2=0.69$ , RMSE=64.4) were achieved when all the bands at 10m spatial resolution were used in modelling coffee leaf Chl for all coffee stands. The prediction accuracy improved ( $R^2=0.77$ , RMSE=56.5) when only mature coffee stands were used. The 20m bands (red-edge and SWIR) performed similarly in coffee leaf Chl estimation as using all bands at 20m and using only 10m bands for all coffee stands and for mature coffee stands. Accuracy metrics were highest and error values lowest when coffee leaf Chl was estimated for mature stands only compared to when both all coffee stands were used in the modelling. It was concluded that Sentinel-2 MSI is a valuable dataset for predicting coffee leaf Chl, however, based on our findings, it is suggested that finer spatial resolutions of 10m applied on mature coffee stands only should be adopted for better prediction results.

**Key words:** Leaf chlorophyll, Crop condition assessment, Random forest, Spectral resolution, Spatial resolution

## 7.1 Introduction

Coffee (*Coffea arabica*) is a perennial tree crop with significant input costs, and thus intra- and inter-seasonal crop condition assessments and productivity monitoring are necessary for sustained production and profitability. Plant biophysical (e.g. LAI) and biochemical (e.g. chlorophyll) parameters are good indicators of vegetation condition and directly related to productivity in crops and natural systems. Chlorophyll (Chl), for example, is important in agricultural crop management as it is a reliable indicator of photosynthetic rates, nitrogen content and leaf health. This is because leaf Chl is associated with plant CO<sub>2</sub> assimilation capacity, which is directly related to photosynthetic rate and quantum yield due to its association with leaf biochemistry such as Rubisco content (Lawlor, 1995). In coffee, Chl is closely associated with N content and yield levels, hence can be used in yield estimations (Reis et al., 2009). Since coffee plantations dominate landscapes in areas of production, Chl amounts in coffee leaves control carbon exchange and are therefore related to a wide range of ecosystem goods and services. In addition, Chl content is an indicator of crop condition as it varies with both biotic and abiotic stress factors such as drought, plant diseases, pest infestation and nutrient imbalances (Netto et al., 2005; de Oliveira et al., 2009; Pompelli et al., 2010). Hence predicting Chl content in coffee is necessary for coffee crop management and yield estimation at local, regional and national levels. This type of information is necessary in monitoring intra- and inter-seasonal variations in crop health required by extension officers, plantation managers, investors, insurers and other stakeholders interested in monitoring crop performance at various scales.

Many studies have shown that Chl can be reliably estimated using remotely sensed data at points, across landscapes and even at regional scales. For example, Kalacska et al. (2015) demonstrated that it is possible to model and map chlorophyll Chl distribution across mixed species peatlands at leaf and canopy levels ( $R^2 > 0.8$ ). Using multi-sourced imagery, Croft et al. (2013) mapped Chl across broadleaf and needle leaf canopies with MERIS and Landsat TM data ( $R^2 = 0.62$  and  $0.65$  respectively). Similar success in remote sensing based estimation of Chl has also been reported in agricultural crops. For instance, Darvishzadeh et al. (2012) successfully used ALOS AVNIR-2 image data and inverted coupled radiative transfer models to predict Chl in rice fields at 10m resolution using root mean square error (RMSE) as the cost function ( $R^2 = 0.65$ , RMSE=0.45), while Moharana and Dutta (2016) produced canopy Chl maps for rice using various vegetation indices and predictive models (highest  $R^2 = 0.77$  for a linear model and  $0.82$  for non-linear model).

Many other studies have applied remote sensing data in Chl estimation in agricultural crops e.g. (Wu et al., 2008; Hunt et al., 2011; Peng et al., 2011; Clevers & Gitelson, 2013). However, adoption of remote sensing in biochemical assaying of perennial tree crops such as coffee remains largely unexplored. In natural ecosystems the canopy structure as influenced by mixtures (or lack thereof) of tree and grassland species dictate absorbance and resultant remote sensing ability of Chl and other pigments (Baret & Buis, 2008; Malenovský et al., 2009). On the other hand, in annual crops such as maize, rice, wheat and soybean, plants follow a weather mediated phenological cycle which results in largely uniform-aged crop fields (Houborg et al., 2015). Unlike in natural systems and in annual cropping systems, perennial tree crops such as coffee are planted uniformly across fields, but with fields of different ages to ensure production continuity. Thus, coffee farms will have different age stands and probably different crop varieties at a time. These have significant influence on spectral characteristics (Chemura & Mutanga, 2016) and therefore complicating the potential use remote sensing data for plant pigment estimations (Chapter 3).

Evidence shows that the Sentinel-2 multispectral imager (MSI) spectral settings can be used in modelling leaf Chl content at canopy level. Clevers and Gitelson (2013) for example, showed that leaf Chl is strongly related to some Sentinel-2 MSI data in maize ( $R^2=0.94$ ), soybean ( $R^2=0.94$ ) potato ( $R^2=0.89$ ) and grass ( $R^2=0.80$ ). Also, Vincini et al. (2014) identified MERIS Terrestrial Chlorophyll Index (MTCI) and red-edge position (REP) as important Sentinel-2 MSI derived indices in predicting Chl in winter wheat canopies. Promising results have also obtained by Frampton et al. (2013) who related several vegetation indices to Chl in grapevines. These results are attributed to the sensor's thirteen unique and strategically positioned bands that are useful for vegetation characterization and mapping. These spectral bands range from the visible and near infrared (VNIR) to the shortwave infrared (SWIR) regions of the spectrum. However, these studies were limited to an evaluation of the sensor's spectral settings, disregarding the fact that the Sentinel-2 MSI data comes at different spatial resolutions for different bands. Knowledge on the interaction between spectral band settings and spatial resolution in Chl and other vegetation characteristics modelling with Sentinel-2 MSI data remains limited. Of the available Sentinel-2 MSI's thirteen bands, four are provided at 10 m spatial resolution, six bands at 20 m spatial resolution and three bands at 60 m spatial resolution (Frampton et al., 2013). This design

is unique to the sensor in that there is interaction of both spatial and spectral resolution in measured reflectance values of each band.

The three new bands (compared to predecessor sensors such as Landsat) in the red-edge region specifically designed for vegetation analysis are centered at 705 nm, 740 nm and 783 nm are available at a 20m spatial resolution while the traditional VIS/NIR are available at a 10m spatial resolution. It has not been established if the lower (20m) spectral bands with vegetation focused red-edge spectral settings are better in coffee (and other vegetation type) Chl modelling compared to the higher (10m) traditional VIS/NIR bands. Some results have also shown that performance of Sentinel spectral settings in modelling leaf Chl are dependent on dominant leaf angle distribution of crop canopies and other biophysical characteristics (Vincini et al., 2015). Hence this study sought to evaluate the performance of Sentinel-2 MSI data in modelling and mapping leaf Chl in coffee plantations at landscape scale. Specifically, the study sought to (i) determine the influence of age on coffee biophysical parameters affecting reflectance (ii) identify the influence of bands combinations and spatial resolutions in estimating leaf Chl with Sentinel-2 MSI data in coffee and (iii) assess the influences of age of coffee stands on coffee leaf Chl modelling performance.

## **7.2 Materials and methods**

### **7.2.1 Study area**

The study was conducted at Jersey Tea and Coffee Estates in Chipinge district, Zimbabwe. The site is located at longitude 32° 41'00E and 32° 42'00E, and latitude 20° 28'00S and 20° 31'00S (Figure 7.1). The area is characterised by a subtropical with two distinct dry and wet seasons, divided almost equally between months of the year i.e. October to March – rainy and April to September - dry season. The topography is undulating with a relief difference of over 100m. The area receives relatively high mean annual rainfall totals for a subtropical area(1200-1300 mm/year) with mostly warm temperatures, around 22.5°C (Lagerblad, 2010; Nicolin, 2011). With deep red clayey soils formed from mafic rocks, climatic conditions in the area make it suitable for good quality coffee production. Sun-coffee production, i.e. coffee plantations without tree shading is practiced in the area (Chemura et al., 2015a).

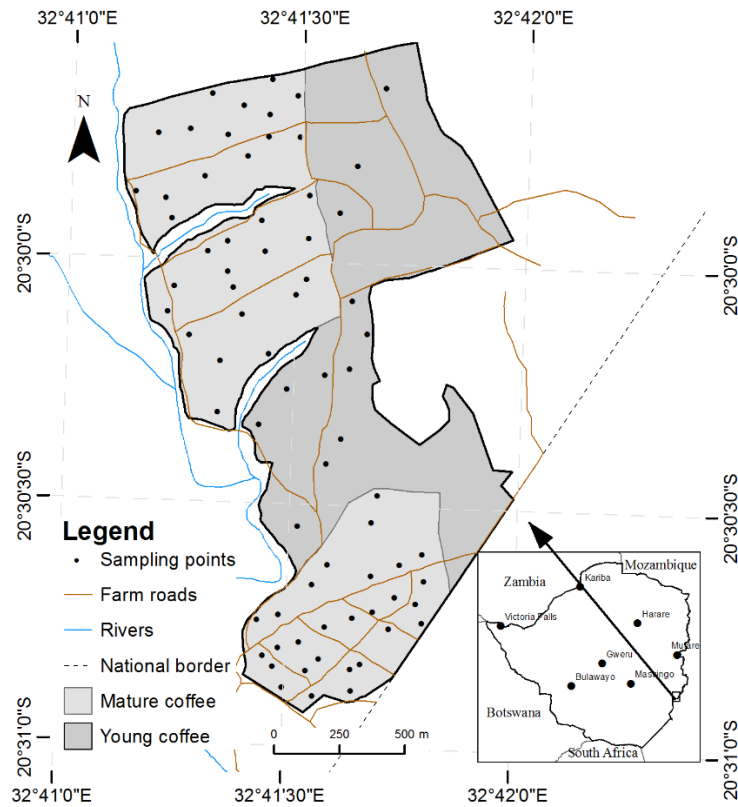


Figure 7.1: The study area showing the distribution of the coffee fields and stand ages.

### 7.2.2 Field data

Field data on various biophysical (coffee height, variety, canopy width, age since planting) and biochemical (Chl) characteristics of coffee that were deemed to be related to reflectance was carried out during the first week of January 2017. Sampling points were randomly selected across the coffee fields and a high-resolution imagery sampling map was generated for the identification of sampling sites. The total size of the sampled fields was 216 hectares with 64% under mature coffee. Sampling points were selected to cover all coffee age groups (mature and young) and to capture the varieties produced in the plantations (Catimor 129 and CR95). The sampling plots on which data was collected consisted of six coffee plants obtained from two rows of three adjacent plants each. The coffee trees had a sowing arrangement of 2.0 m x 2.5 m giving a sampling area of 15m<sup>2</sup> for each sampling point. The height (in cm) of each plant was measured using a graduated stick and averaged per plot.

Canopy area was determined by measuring canopy diameter (in cm) of each planting station in the plot, which was then used to calculate canopy area, assuming that the coffee canopy

approximates a circle. A SPAD-502 (Konica Minolta Sensing Inc., Japan) was used to measure absorbance that was converted into total coffee leaf Chl content. The SPAD reading of each coffee leaf was taken at approximately the same spot on selected leaves. The average of SPAD measurements from five leaf discs of each of the six coffee trees was used to determine the average SPAD readings of the plot. To convert the SPAD readings to total Chl, 34 coffee leaves collected from the same fields as SPAD measurements were frozen and their Chl a and b determined in the laboratory. The N,N-dimethylformamide (DMSO) solvent method was used for Chl extraction and a curvilinear function relating SPAD readings and total Chl ( $\mu\text{g}/\text{cm}^2$ ) was established and used.

### 7.2.3 Image acquisition and pre-processing

Sentinel-2 Level-1C (L1C) MSI data were downloaded from Remote Pixel, a Sentinel-2 Data Hosting portal (<https://remotepixel.ca/>). The Sentinel-2 data for the study area was taken on the 30th of November 2016, which was the closest day to field work with usable cloud-free images. The Sentinel-2 L1C product used in this study is characterized by individual bands of 100 km<sup>2</sup> tiles (ortho-images in UTM/WGS84 projection). The image had 0.02% cloud cover and the spectral bands, center wavelengths and band width of the Sentinel-2 MSI data used are shown in Table 7.1.

Table 7.1: Specifications of the Sentinel-2 Multispectral Instrument (MSI)

Spectral band	Centre wavelength (nm)	Band width (nm)
B2	490	65
B3	560	35
B4	665	30
B5	705	15
B6	740	15
B7	783	20
B8	842	115
B11	1610	90
B12	2190	180

Individual bands were stacked in four different ways according to the different spatial resolutions. To obtain a stack with all 9 bands at 20m resolution, the 10m bands were resampled to 20m using nearest neighbour in Sentinel Application Platform (SNAP) version 4.0. To obtain

all 9 bands at 10m resolution the 20m bands were pan sharpened in ENVI 5.3 (Exelis Visual Information Solutions, Boulder, CO, USA) using Gram-Schmidt pan-sharpening. The Gram-Schmidt method is generally recommended for most pan-sharpening applications as it is typically more accurate than other methods. This accuracy is attributed to its use of spectral response function of a given sensor to estimate what the higher resolution band would be.

All resampling and pan-sharpening were done after converting reflectance to top of the canopy reflectance (Level-2A) using the Sen2cor atmospheric correction module in SNAP (Muller-Wilm et al., 2013). The dark-object subtraction method was applied in QGIS 2.6 using Semi-Automatic Classification Plugin for QGIS as the reflectance values were low after atmospheric correction in SNAP. To model the influence of age on performance, data analysis was done firstly on all the data, then on mature coffee stands only, with the mature fields obtained from an age mask. The band combination, number of variables and spatial resolutions are shown in Table 7.2.

Table 7.2: Specifications of the Sentinel-2 Multispectral Instrument (MSI) band settings showing spatial/spectral combinations, number of bands and spatial resolution.

Set	Spatial/Spectral Combinations	# of bands	Sentinel-2 MSI bands	Spatial Resolution	Description
A	All MSI bands at 20m	9	B2, B3, B4, B5, B6, B7, B8a, B11, B12	20m	10m bands resampled to 20m
B	10m bands only	4	B2, B3, B4, B8a	10m	Original 10m bands only
C	20m bands only	5	B5, B6, B7, B11, B12	20m	Original 20m bands only
D	All MSI bands at 10m	9	B2, B3, B4, B5, B6, B7, B8a, B11, B12	10m	20m bands pan-sharpened to 10m resolution



#### 7.2.4 Machine learning modelling approach

The Random Forest (RF) algorithm was used for modelling total coffee chlorophyll from different band combinations and spatial resolutions. RF is an ensemble machine learning algorithm developed by Breiman (2001) to solve regression problems through a multitude of decision trees. RF employs an iterative bagging (bootstrap aggregation) operation where the number of trees (*ntree*) are independently built using a random subset of samples from the training samples. Each node is then split using the best, among a subset of input variables (*mtry*). In many applications, this algorithm produces one of the best accuracies to date and has important advantages over other techniques in terms of ability to handle highly non-linear data, robustness to noise and tuning simplicity (Rodriguez-Galiano et al., 2012; Lebedev et al., 2014). The default number of trees (*ntree*) of 500 was used while *mtry* is automatically determined as the square root of the total number of variables used (Breiman, 2001). Coffee chlorophyll modelling was done in R (R Core Team, 2013) using the R package randomForest to run the RF modelling (Liaw et al., 2009) for both training and prediction.

#### 7.2.5 Accuracy assessment and performance comparison

In order to assess the performance of all models in coffee leaf chlorophyll estimation, the field data was split into 60% for training and 40% for evaluation (43:29 for all stands and 36:24 for mature coffee stands). The correlation coefficient (*r*) and coefficient of determination ( $R^2$ ) were used to assess the goodness of fit of the predicted and measured coffee leaf chlorophyll. In addition, Mean Absolute Error (MAE, Equation 7.1) Root Mean Square Error (RMSE, Equation 7.2), and percent bias (pBias, Equation 7.3) were used to determine the errors of the model in predicting coffee leaf chlorophyll from variables.

$$\text{MAE} = \left( \frac{1}{n} \sum |y_i - \hat{y}_i| \right) \quad [7.1]$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad [7.2]$$

$$\text{pBias} = \left( \frac{\sum (y_i - \hat{y}_i) * 100}{\sum y_i} \right) \quad [7.3]$$

where for all cases  $n$  is the number of data points,  $y_i$  is the measured coffee leaf chlorophyll content at that data point and  $\hat{y}_i$  is the model predicted coffee leaf chlorophyll content at that data point (Moriassi et al., 2007; DeJonge et al., 2016).

### **7.3 Results**

#### **7.3.1 Effect of biophysical and age characteristics**

Tree height and canopy area increased with age, with results showing that Cat129 variety is taller than CR95 while not having wide canopies (Figure 7.2a and Figure 7.2b). Total coffee Chl was not significantly influenced by age or variety (Figure 7.2c). The results show that there was more variation in coffee tree canopy area per age group and variety compared to that of height, especially in mature coffee trees. There was great variation in total coffee leaf Chl across varieties and ages. These results indicate that there was no significant influence of variety ( $p>0.05$ ) on all key biophysical and biochemical parameters in coffee leaves. Conversely, age significantly affected coffee biophysical parameters i.e. height and canopy area ( $p<0.05$ ) and not the biochemical parameter i.e. Chl ( $p>0.05$ ). Young coffee stands had a mean Chl of  $43.4\pm 10.6 \mu\text{g}/\text{cm}^2$  while mature coffee stands had a mean Chl of  $49.3\pm 10.2 \mu\text{g}/\text{cm}^2$ .

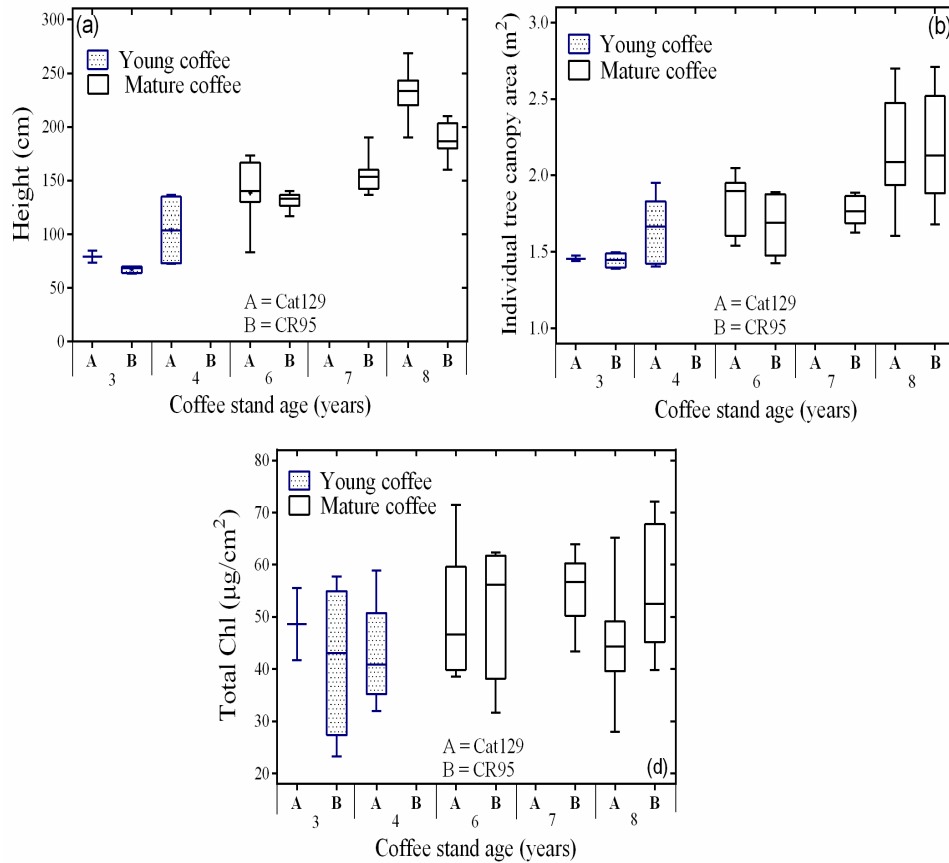


Figure 7.2 Influence of stand age and variety on (a) coffee height (b) tree canopy area (c) total Chl obtained from field work (n=72). The numbers (3,4,6,7 and 8) are the ages while letters (A and B) are coffee varieties.

### 7.3.2 Total coffee leaf chlorophyll estimation with all coffee stands

Results showed that better modelling performance was achieved using the 10m spatial resolution Sentinel-2 MSI data than 20m in the parameterized RF model. There was poor model performance when Chl modelling was done with all Sentinel-2 MSI bands at 20m resolution ( $R^2=0.60$ ,  $RMSE=7.0$ , Figure 7.3a, Table 7.3). The predictive performance of the RF further deteriorated when the 5 Sentinel-2 spectral bands originally at 20m were used in modelling coffee leaf Chl ( $R^2=0.58$ ,  $RMSE=7.4$ , Figure 7.3c, Table 7.3). Conversely, modelling coffee leaf Chl at 10m improved model performance. Better model performance was achieved when modelling all coffee stands with all bands at 10m resolution ( $R^2=0.69$ ,  $RMSE=6.8$ , Figure 7.3d), than 4 bands at 10m resolution ( $R^2=0.62$ ,  $RMSE=7.2$ , Figure 7.3b). Generally, the results showed overestimation of lower Chl values (below  $50\mu\text{g}/\text{cm}^2$ ) and underestimation of higher values (Figure 7.4)

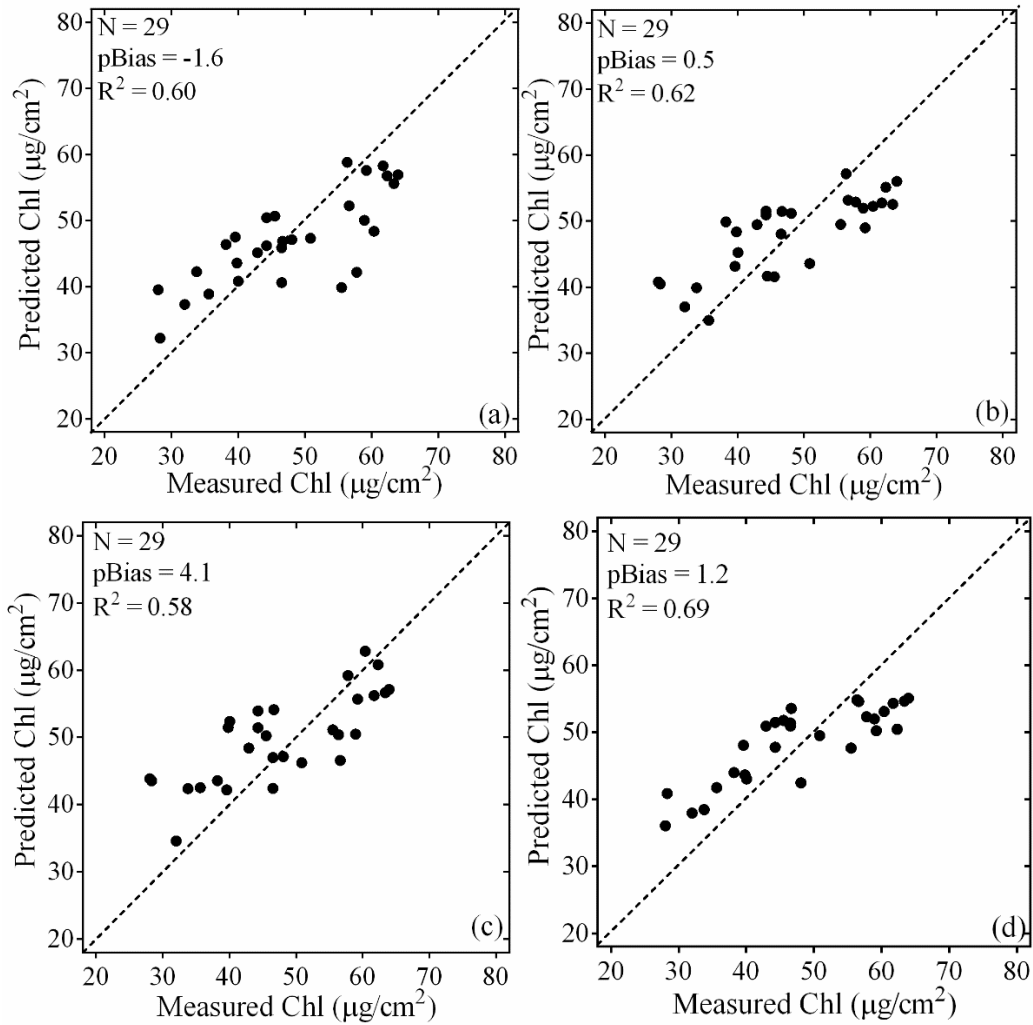


Figure 7.3: Relationship between measured and predicted coffee leaf chlorophyll using (a) all 9 bands at 20m resolution, (b) 4 bands originally at 10m resolutions (c) 5 bands originally at 20m resolutions and (d) all 9 bands at 10m resolutions.

Table 7.3: Performance evaluation of different band settings in predicting coffee leaf chlorophyll using all coffee stands.

Band settings	MAE	RMSE	r	R <sup>2</sup>
All bands at 20m	5.7	7.0	0.78	0.60
10m bands only	6.4	7.2	0.79	0.62
20m bands only	6.3	7.4	0.76	0.58
All bands at 10m	6.3	6.8	0.83	0.69

The total coffee leaf Chl maps produced from the best performing 20m RF model (all 9 bands at 20m resolution) and at 10m (all 9 bands at 10m resolution) are shown in Figure 7.4a and Figure 7.4b respectively. At 20m resolution, the maps show that the greater part of the younger coffee fields have lower total coffee leaf Chl (Figure 7.4a), which is somewhat unexpected, considering the insignificant influence of age on total coffee leaf Chl in Figure 7.2c. However, there is concurrence between the 20m and 10m RF models in predicting northern and central parts of the coffee fields as having lower total coffee Chl.

The slope of the relationship between measured and predicted total coffee leaf Chl was not significantly different between using all bands at 20m and using only the 10m bands. These results thus show that spectral settings of Sentinel-2 MSI data play a significant role in total coffee leaf Chl prediction performance beyond that played by spatial resolution. As a result, using all coffee stands, the prediction performance among all bands at 20m, 10m bands only and 20m bands only were close in terms of error metrics (Table 7.3).

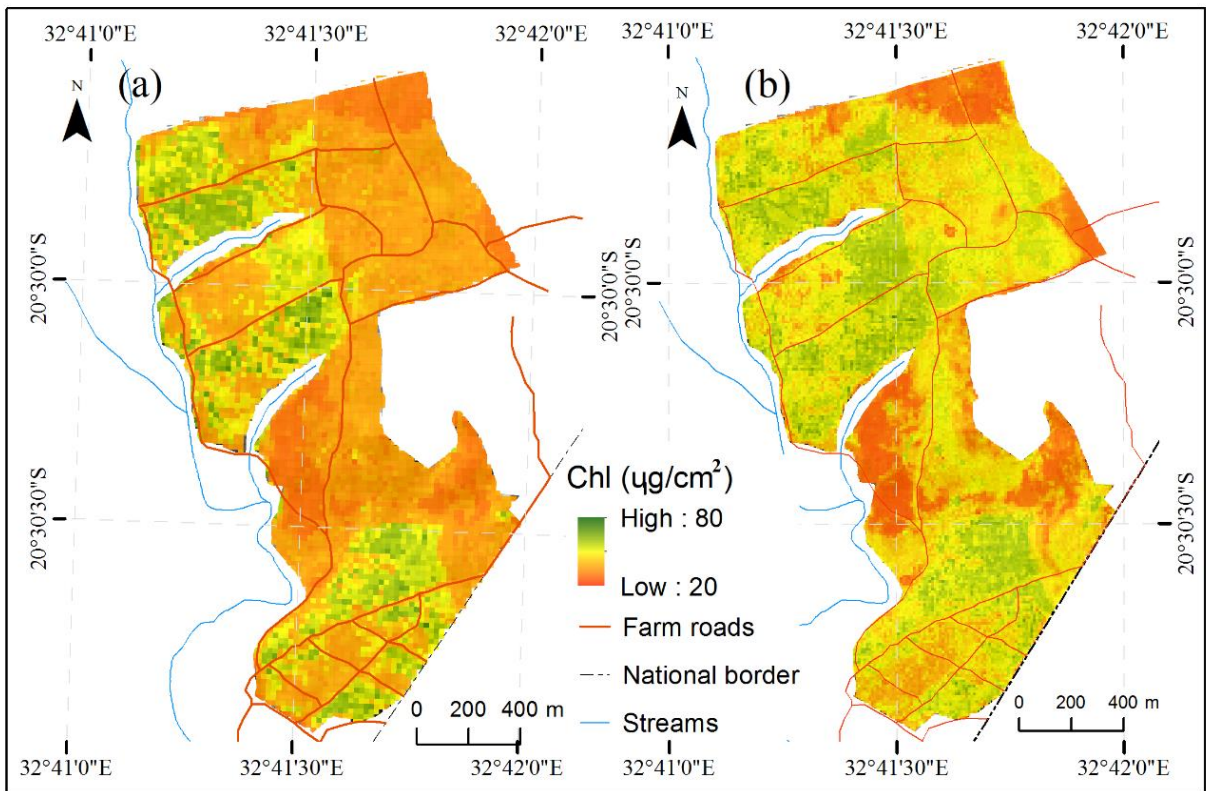


Figure 7.4: Total coffee leaf Chl distribution maps obtained using the RF on (a) all 9 bands at 20m spatial resolutions and (b) all 9 bands at 10m spatial resolutions for all coffee stands.

### 7.3.3 Total coffee leaf chlorophyll estimation for mature coffee stands only

Although there was no significant influence of age in coffee leaf Chl content (Figure 7.2c), modelling results show that age is important in modelling coffee leaf Chl as there is a general increase in prediction performance for mature coffee stands compared to when all coffee stands were used (Table 7.4). However, the trends in model performance remained the same where models built using 20m spatial resolution had lower predictive power (Figure 7.5a and 7.5c), compared to those built using 10m spatial resolution Sentinel-2 MSI data (Figure 7.5b and Figure 7.5d). As with using all coffee stands, the best performing model for mature coffee was for all bands at 10m spatial resolutions ( $R^2=0.77$ ,  $RMSE=5.9$ , Figure 7.6d, Table 7.3). All results showed that prediction of total coffee Chl had positive model bias with the least bias of 0.4 obtained from modelling total coffee Chl with all 9 bands at 20m (Figure 7.5a).

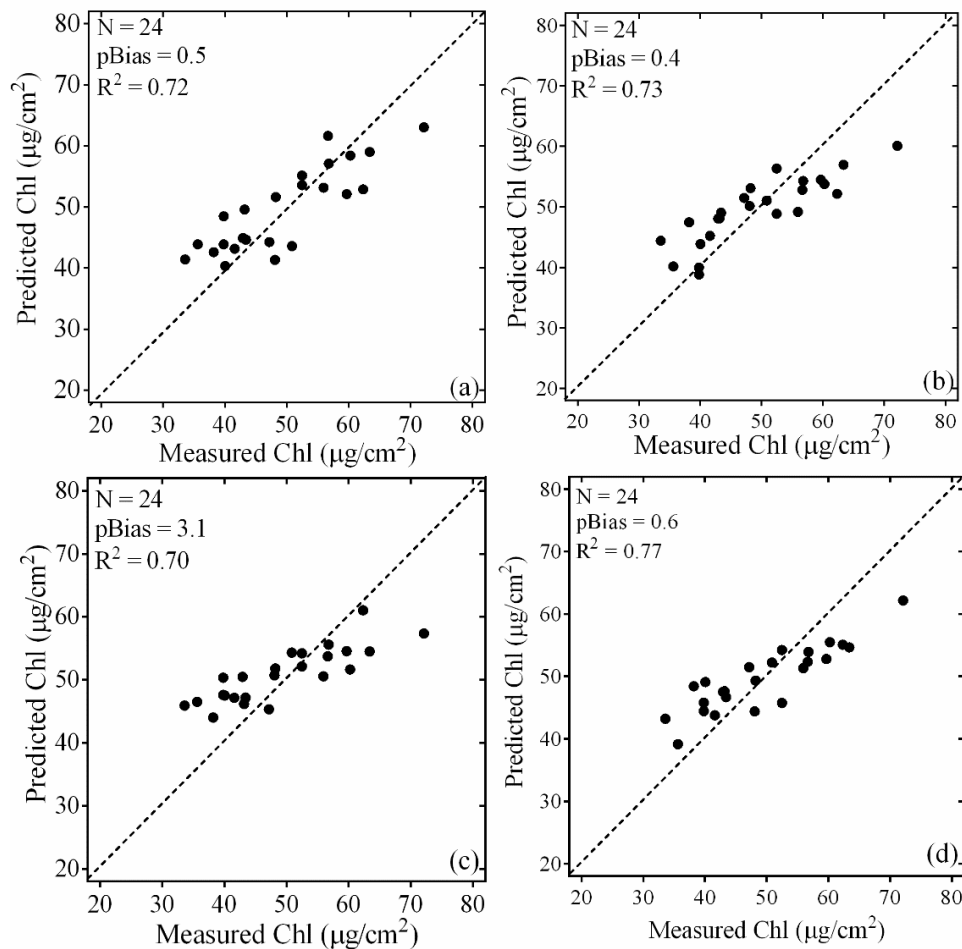


Figure 7.5: Relationship between measured and predicted mature coffee leaf chlorophyll using (a) all 9 bands at 20m resolution, (b) 4 bands originally at 10m resolutions (c) 5 bands originally at 20m resolutions and (d) all 9 bands at 10m resolutions.

Table 7. 4: Performance evaluation of different band settings in predicting coffee leaf chlorophyll using mature coffee stands only.

Band settings	MAE	nRMSE %	PBIAS %	r	R <sup>2</sup>
All bands at 20m	45.24	54.4	3.1	0.85	0.72
10m bands only	44.20	57.1	1.3	0.85	0.73
20m bands only	51.54	66.1	3.1	0.84	0.70
All bands at 10m	50.25	58.6	1.4	0.88	0.77

The distribution of the total coffee leaf Chl as predicted from the best performing 20m RF model (all 9 bands at 20m resolution) and at 10m (all 9 bands at 10m resolution) for mature coffee stands are shown in Figure 7.6a and Figure 7.6b respectively. There is an agreement in patterns of total coffee Chl distribution but the total coffee leaf Chl distribution from using all bands at 10m (Figure 7b) shows greater Chl variation than those from using all bands at 20m (Figure 7.6a). Using 20m Sentinel-2 MSI bands had the highest RMSE and the lowest R<sup>2</sup> value in total coffee leaf Chl estimation.

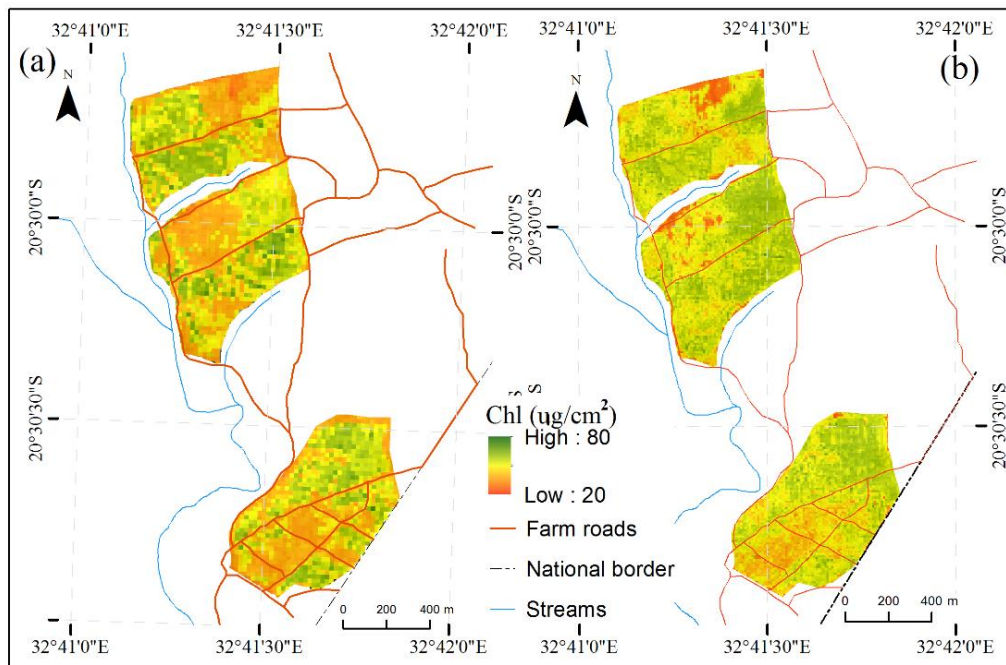


Figure 7.6: Total coffee leaf Chl distribution maps obtained from the RF using (a) all 9 bands at 20m spatial resolution and (b) all 9 bands at 10m spatial resolutions for mature fields.

### 7.3.4 Comparison of performances of different approaches in coffee chlorophyll estimation

Correlating the predicted total coffee leaf Chl from the different band settings showed that the highest correlations were between 10m spatial resolutions ( $r=0.82$ ) while the least was between 20m spatial resolution ( $r=0.59$ ) for all coffee stands (Table 7.5). A similar trend was observed for prediction of total coffee Chl in mature stands only. The influence of canopy area on total coffee leaf Chl prediction accuracy was further evaluated (Figure 7.7). The results indicated that for all spatial and spectral settings, there was overestimation of total coffee leaf Chl for smaller coffee canopy area and underestimation for large coffee canopies, with a higher significance at 20m spatial resolution (Figure 7.7a). Absolute differences show that prediction errors were higher for smaller coffee canopies (younger coffee), which decreased as canopy sizes increased (Figure 7.7b).

Table 7.5: Correlation coefficients ( $r$ ) between predicted coffee leaf chlorophyll using different band settings\*.

	All bands at 20m	10m bands only	20m bands only	All bands at 10m
All bands at 20m	1	0.70	0.68	0.73
10m bands only	<b>0.72</b>	1	0.67	0.85
20m bands only	<b>0.59</b>	<b>0.75</b>	1	0.70
All bands at 10m	<b>0.73</b>	<b>0.82</b>	<b>0.65</b>	1

\*Bold figures (below the diagonal) are for all coffee stands while plain numbers are for mature coffee stands only.

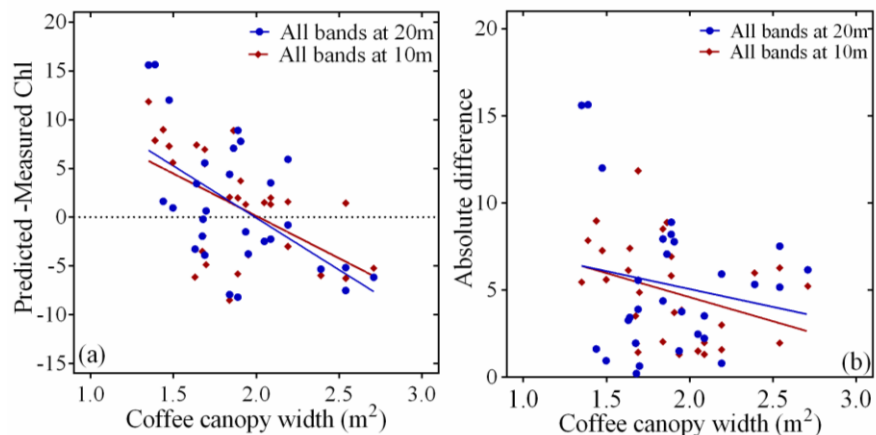




Figure 7.7: Effect of coffee canopy area on total coffee Chl estimation accuracy as shown by (a) prediction difference and (b) absolute difference with all coffee data (n=29).

## **7.4 Discussion**

The aim of this study was to assess the influence of Sentinel-2 MSI spectral settings and spatial resolution and coffee stand age in image-level coffee leaf Chl estimation with the empirical RF machine learning algorithm. This aim was pursued with an intention of developing landscape scale spatially-explicit crop condition assessment approach that is scientifically sound, easier to implement and cost effective. It is anticipated that adoption of such an approach will be valuable in coffee condition assessments for improving productivity as well as provision of environmental goods and services. This is based on the fact that total coffee leaf Chl is a valuable indicator of crop health, hence yield quality and quantity (de Oliveira et al., 2009; Reis et al., 2009). To achieve the aforementioned aim, it was hypothesized that understanding the influence of spectral settings and spatial resolution of Sentinel-2 MSI data, and of coffee stand age could provide useful insights into choice of bands and limits to practical applications in determining coffee stand Chl content.

### **7.4.1 Effect of biophysical and age characteristics**

As expected, coffee biophysical attributes (height and canopy area) were strongly related to coffee age. However, there was no relationship between these attributes and coffee variety. This is explained by the fact that the majority of coffee varieties originate from crosses of the genotype hibrido de timor (Chidoko et al., 2017), which possibly explains why they are not significantly different in their structural characteristics. Our findings are consistent with a large body of literature that has used remotely sensed data for age mapping or modelling of other perennial crop characteristics using the relationship between stand biophysical characteristics and age (Chemura et al., 2015b). However, our findings show that biochemical properties are not influenced by age or variety, as total coffee Chl was variety and age invariant. Thus, while biophysical coffee parameters are age-dependent, biochemical characteristics may be influenced by site conditions such as fertility, insects or pests or other local factors.

In remote sensing, measured canopy reflectance is determined by the plant's structural properties and biochemical parameters. For example, local factors such as soil fertility and or soil moisture have direct influences on both biophysical (canopy area, LAI, height) parameters related to age

and biochemical (total Chl, leaf N, PWC) crop parameters that may not be related to age. Whereas the influence of these parameters on reflectance is relatively complex, it has been established that biophysical vegetation parameters such as LAI and height are not always related to foliar pigment concentrations. For example, Lepine et al. (2016) found no relationships between leaf nitrogen and LAI in boreal and temperate forests of different ages. Based on our findings, it is concluded that variety specific models are unnecessary, as the coffee's biophysical and biochemical characteristics are similar between coffee varieties. Although age was not related to biochemical plant parameters, it is suggested that it should be incorporated in modelling as an age-related influence of biophysical characteristics on coffee canopy reflectance was established.

#### **7.4.2 Effect of spectral settings and spatial resolution on coffee leaf Chl estimation**

Based on our finding, it is concluded that total coffee leaf Chl empirical modelling performance with Sentinel-2 MSI data is more related to image's spatial resolution than spectral characteristics. This finding is important as the Sentinel-2 MSI data comes at 10m, 20m and 60m spatial resolutions, tied to unique spectral settings. There is much promise reported in the potential role of Sentinel-2 MSI red-edge bands in vegetation characterization and quantification (Herrmann et al., 2011; Clevers & Kooistra, 2012; Frampton et al., 2013; Sibanda et al., 2016). However, these studies relied on simulated spectra, ignoring the influence of spatial resolution. While simulation-based studies provide important information on the potential of a sensor to perform specific tasks, it is necessary that future studies consider all sensor factors that may influence its performance. This is particularly critical to Sentinel-2 MSI data as its VIS-NIR data comes in different spatial resolutions, a departure from previous sensor designs such as Landsat, Worldview and GeoEye that have a constant spatial resolution, except for the panchromatic band. Thus, our finding that spatial resolution is more important in influencing total Chl modelling in coffee casts aspersions on spectroscopic simulation studies using Sentinel-2 MSI data that did not consider spatial resolution. Studies in natural forests have also reported that leaf biochemical estimations are more accurate at finer resolutions than at coarser resolutions (Lepine et al., 2016). The finding that Sentinel-2 MSI data's spatial resolution is more influential than spectral settings in total coffee leaf Chl estimation may however be restricted to coffee for a number of reasons. Firstly, coffee is planted in hedgerows. Hence, there is always soil background effect due to open canopies. In this regard, the visible portion of the spectrum becomes an important parameter in

explaining variations in reflectance compared to closed canopies where the NIR and other vegetation specific bands (and resultant vegetation indices) are more significant (Kalacska et al., 2015). Secondly, the NIR (B8) band used in this study is available at 10m spatial resolution with VIS bands while the vegetation quality and quantity bands of the red-edge portions are available at 20m. Studies have shown that the NIR band alone can be used in plant pigment concentration estimation. For instance, Lepine et al. (2016) developed a model for N mapping using MODIS ( $R^2=0.80$ ) and Landsat ( $R^2=0.79$ ) NIR bands, with no red-edge bands. Many other studies (El-Shikha et al., 2008; Lee et al., 2008; Schlemmer et al., 2013a) have associated NIR reflectance with plant biochemical parameters related to total Chl.

The Sentinel-2 MSI NIR band has the largest spectral width of all MSI bands (115nm), which can be considered an advantage or a disadvantage in total coffee leaf Chl estimation. It is an advantage in that the wide spectrum can capture and average wide reflectance across the NIR spectrum including that of background effect, capturing influence of structural parameters such as canopy cover in the process. In this regard, Lepine et al. (2016) observed no effect of bandwidth on the strength of the relationship between leaf biochemistry and canopy reflectance, hence suggested that many of narrow wavelengths in imaging spectrometers may just be superfluous, a view incongruent with our findings. On the other hand, recorded NIR reflectance maybe be directly related to total coffee Chl through its interaction with photons that have an influence on spectral absorption features and indirectly through capturing effects of plant associations that affect light scattering and capture (Kruse et al., 2006; Ollinger, 2011). Unscrambling what, by how much and why and what is contributing to what is beyond the scope of this study. Nevertheless, since there was a similar performance between the use of 20m bands (red-edge bands) and 10m bands with all and mature coffee stands, it is concluded that there is an influence of these red-edge bands in plant biochemical analysis. However, in coffee this is eroded by spatial resolution.

#### **7.4.3 Influence of coffee stand age on coffee leaf Chl estimation**

Coffee leaf biochemical parameters did not significantly vary with age or variety. or variety. Unlike in multispecies studies, the Chl of coffee leaves is expected to not significantly vary with anomalous Chl conditions being caused by such factors as soil fertility, diseases, pest attack, water stress or other stressors (de Oliveira et al., 2009; Reis et al., 2009). However, modelling

results improved on mature coffee stands, compared to the use of both young and mature ages. This result is partly explained by the observation that the NIR reflectance is more reflective in mature coffee (Chemura & Mutanga, 2016) and thus able to capture variability in factors that contribute to the reflectance. The effects of canopy area (which is directly related to age) were also apparent in influencing both prediction patterns and magnitude of errors. From these results, it is suggested that the model for coffee leaf Chl mapping be applied only to mature coffee fields (over 4 years). In large scale modelling, use of pre-existing age masks is useful for practical applications (Chapter 3). This is despite the fact that the data used for mature stands was 84% of the data used in all stands. Therefore, the few young coffee stands (12 in this case) had an effect of reducing the modelling performance for the entire dataset when included in the modelling of total coffee Chl. Further studies using more young stands maybe required to confirm this and determine if this is the same beyond coffee plantations to other tree crops and plantation forests that may have different age stands.

When coffee leaf Chl was estimated for mature fields only, the best model (all bands at 10m resolution) only explained 77% of the variation in measured coffee leaf Chl. Potential sources of errors in this finding include the difference between the image and the field data collection dates, the strength of the pan-sharpening (where 20m bands were resampled to 10m) and the algorithm used. Despite these factors possible contribution to the limited results, our selection is considered to be the best possible for the study. Based on our findings and approach, it may be possible to develop Sentinel-2 MSI data-based near-real time online and mobile apps for coffee Chl assessment that can be used by field managers, insurance companies, extension officers and other interested stakeholders as Chl is an indicator of coffee condition. This is possible, especially once the twin Sentinel-2b satellite, with a five-day temporal resolution is in orbit.

## **7.5 Conclusions**

This study has evaluated the role of Sentinel-2 MSI spectral settings and spatial resolution together with stand parameters related to age in empirical estimation of coffee leaf Chl content using image data. From this study, it was concluded that:

- 1) Coffee biophysical parameters (height and canopy cover) are significantly influenced by stand age while biochemical parameters (PWC and total Chl) are age invariant. This is

despite the fact that total canopy reflectance is influenced by both biophysical and biochemical characteristics.

- 2) Fine scale (10m) spatial resolution Sentinel-2 MSI data produce better coffee leaf Chl estimation results, regardless of band settings. Thus, future applications should consider spatial resolution, in addition to the red-edge bands in coffee Chl estimations.
- 3) Models for coffee leaf Chl estimation with Sentinel-2 MSI data should be applied on mature coffee stands for better results as young coffee increases prediction errors.

### **7.6 Link to next chapter**

This chapter showed that it is possible to model coffee leaf chlorophyll with Sentinel-2 multispectral data at landscape scale. The key findings are that the spatial resolution is more important and that acceptable modelling results are obtained for mature coffee formed the basis of the next chapter in nitrogen modelling. In the next chapter, spatial variability of coffee foliar nitrogen is modelled at 10m spatial resolution and for mature stands only as informed from this chapter.

# CHAPTER 8: LANDSCAPE SCALE MULTISPECTRAL REMOTE SENSING OF COFFEE FOLIAR NITROGEN CONTENT



Photo credits: A. Chemura (2017)

## **This chapter is based on:**

1. Abel Chemura, Onesimo Mutanga, John Odindi and Dumisani Kutwayo (2017) Mapping spatial variability in foliar nitrogen in coffee (*Coffea arabica*) plantations with multispectral level Sentinel-2 MSI data. **The 38th Asian Conference on Remote Sensing**, October 23-27, 2017, New Delhi, India.
2. **Chemura, A.**, Odindi, J. and Mutanga, O., Kutwayo, D. (Accepted) Mapping spatial variability in foliar nitrogen in coffee (*Coffea arabica*) plantations with multispectral level Sentinel-2 MSI data, Submitted to **ISPRS Journal of Photogrammetry and Remote Sensing**.

## **Abstract**

Nitrogen (N) is the most limiting factor to coffee development and productivity. Therefore, development of rapid, spatially explicit and temporal remote sensing based approaches to determine spatial variability of coffee foliar N are imperative for increasing yields, reducing production costs and mitigating environmental impacts associated with excessive N applications. This study sought to assess the value of Sentinel-2 MSI spectral bands and vegetation indices in empirical estimation of coffee foliar N content at landscape level. Results showed that coffee foliar N is related to Sentinel-2 MSI B4 ( $R^2=0.32$ ), B6 ( $R^2=0.49$ ), B7 ( $R^2=0.42$ ), B8 ( $R^2=0.57$ ) and B12 ( $R^2=0.24$ ). Vegetation indices were more related to coffee foliar N as shown by the Inverted Red-Edge Chlorophyll Index - IRECI ( $R^2=0.66$ ), Relative Normalized Difference Index – RNDVI ( $R^2=0.48$ ), CIRE1 ( $R^2=0.28$ ), and Normalized Difference Infrared Index – NDII ( $R^2=0.37$ ). These variables were also identified by the random forest variable optimisation as the most valuable in coffee foliar N prediction. Modelling coffee foliar N using vegetation indices produced better accuracy ( $R^2=0.71$  with RMSE= 0.27 for all and  $R^2=0.73$  with RMSE= 0.25 for optimized variables), compared to using spectral bands ( $R^2=0.57$  with RMSE= 0.32 for all and  $R^2=0.58$  with RMSE= 0.32 for optimized variables). Combining optimized bands and vegetation indices produced the best results in coffee foliar N modelling ( $R^2=0.78$ , RMSE=0.23). All the three best performing models (all vegetation indices, optimized vegetation indices and combining optimal bands and optimal vegetation indices) established that 15.2 ha (4.7%) of the total area under investigation had low foliar N levels (<2.5%). This study demonstrates the value of Sentinel-2 MSI data, particularly vegetation indices in modelling coffee foliar N at landscape scale.

**Keywords:** Nutrient management, random forest, canopy nitrogen, precision agriculture.

## 8.1 Introduction

Coffee follows the C3 or Calvin cycle photosynthetic pathway with the plant biomass productivity, growth and yield being dependent on carbohydrates produced through photosynthesis (Wrigley, 1988). The capacity of the plant to produce these carbohydrates depends on nutrient supply with nitrogen (N) known as the single most limiting factor to coffee productivity (Coste, 1992). Mature coffee requires a total of 105 kg ha<sup>-1</sup> of N to achieve yield levels of 1 tonne ha<sup>-1</sup> per year, and yet some coffee farms report up to 6 tonnes clean coffee per ha<sup>-1</sup> per year (Chemura, 2014). N plays several interconnected roles in coffee plant development and productivity. Firstly, N determines plant establishment and root growth, which in turn influences other aspects of the plant's health. Secondly, in addition to enhancing berry productivity in the year of supply, N ensures production of fresh cropping wood frame, valuable for improved productivity in subsequent years. Furthermore, N plays a significant role in plant resilience to biotic and abiotic stresses. For example, it has been suggested that bushes with low nitrogen levels are more susceptible to coffee white stem borer attacks (Kutywayo, 2002b). In addition, sufficient N helps the coffee plant to tolerate higher levels of manganese by preventing the breakdown of leaf proteins under extremely hot and dry conditions. Coffee plants with higher N are also known to have a higher ability to withstand drought due to higher hydraulic conductivity (Logan & Biscoe, 1987; DaMatta, 2004).

Commonly, N fertilisers are used to supplement soil available nitrogen to increase growth and prevent leaf fall, hence optimise photosynthetic area. On the other hand, excessive or ill-timed N may induce the excessive production of side shoots from the nodes, rather than flower buds, reducing berry yield (Kutywayo et al., 2010). Concerns about leaching and volatilization of N fertilisers from excessive applications have also been raised (Bortolotto et al., 2012; Tully et al., 2012). Although N is essential for coffee plants throughout the year, its demand increases significantly during blossom and berry development (Coste, 1992). Hence, Coffee N management decisions are not just about N supply, but about achieving and maintaining a season long balance of nitrogen within the plant.

To achieve the fine N balance in coffee plants, plantation managers supply nitrogen as compound, organic or straight fertilisers applied as basal applications, foliar sprays and/or trickle or drip fertigation. These applications are based on calendar management program or occasional



lab-based soil and/or leaf sampling used in conjunction with established diagnostic norms. In extreme cases, nutrient deficiency symptoms such as the characteristic yellowing of leaves for N deficiency are used for reactive decision making. These methods are not only labour intensive, but also commonly adopted once economic damage has already been inflicted on the crop (Chemura et al., 2017). In the long term, these sampling-based methods are not ideal for balancing the crop nutrient needs, supply of nutrients from natural sources, and the short- and long-term fate of the fertilizer applied, resulting in unsustainable production and yield shortfalls.

Remote sensing methods have been demonstrated to successfully identify distributions of foliar N in natural vegetation and agricultural crops across different biomes. This is enabled by the fact that N is related to factors influencing spectral responses of vegetation across the visible, NIR and SWIR regions of the spectrum. Many authors, (e.g. Ollinger et al. (2008), Wang et al. (2016) and Lepine et al. (2016)) have reported that there are significant correlations between NIR reflectance (800–850 nm) and canopy foliar mass-based nitrogen concentration (N), which enable remote sensing-based N predictions. Whereas the reason for this correlation is still under speculation (Knyazikhin et al., 2013; Ollinger et al., 2013), there is sufficient ground for development of remote-sensing based foliar N prediction and mapping from remotely sensed data at various scales (Baret et al., 2007).

Leaf level N versus reflectance dynamics vary with canopy level interactions. For instance, Frampton et al. (2013) reported contrasting performance of vegetation indices at leaf and canopy levels. Consequently, Wang et al. (2016) suggested that canopy structure confounds the estimation of foliar nitrogen when using canopy spectral data because it is the main driver of canopy reflectance variations. This is interesting for application of remote sensing in coffee plantations because coffee plants have a distinct canopy, influenced by their physiology, growth cycle and planting arrangement. Thus, reported interactions between vegetation and reflectance in forests, grasslands, annual crops and other plantations may not hold for coffee. This is because (i) coffee is planted in hedgerows and there is always background effect which reduces potential for saturation (ii) the function and structure of the canopy that explain reported variation is the same in coffee fields (except where coffee is under shade) and, (iii) each pixel is usually on uniform aged plants, with similar height and canopy width, which influence adjoining cover reflectance.

Understanding the functional relationships between leaf N and plant growth and development, as well as rapid and less expensive diagnostic approaches for spatial and temporal estimation of plant N are necessary for efficient and sustainable management of coffee plantations. However, majority of reported approaches either use hyperspectral data, which is not easily accessible to many coffee producers or field spectroscopy, with similar limitations that characterise aforementioned sampling approaches. New generation satellite sensors such as WorldView-2 and 3, RapidEye and Sentinel-2 have incorporated new wavebands unavailable in similar predecessor sensors. The Sentinel-2 is particularly attractive due to its free availability, with relatively high spatial resolution, as well as strategically positioned bands, making it useful for many applications that include vegetation characterisation and mapping. It has a huge swath-width of about 290 km with thirteen unique spectral bands (Hansen & Loveland, 2012). These spectral bands range from the visible and near infrared (VNIR) to the shortwave infrared (SWIR) regions of the spectrum. Of these thirteen bands, four are provided at 10 m spatial resolution, six at 20 m spatial resolution and three at 60 m spatial resolution (Funk & Budde, 2009).

Due to these unique characteristics, the Sentinel-2 multispectral imager is hyped to be capable of providing timely data for the generation of high-level operational products. These include the generation of spatially explicit estimation and monitoring of important plant biophysical and biochemical variables such as chlorophyll, N, LAI, leaf water content and crop health. Given the uniqueness of coffee as a target and the spectral features of the Sentinel-2 MSI data, the aim of this study was to assess the potential for Sentinel-2 bands and vegetation indices to predict foliar N content in coffee (*Coffea arabica* L.). A secondary objective was to establish the relationships between spectral variables and coffee foliar N at canopy level. The latter objective was aimed at identifying the best performing variables in coffee N prediction for the purposes of quantifying and mapping N levels that can be operationally used to characterise and manage coffee fields to achieve optimum productivity.

## **8.2 Materials and methods**

### **8.2.1 Study area**

The study was conducted at Jersey Tea and Coffee Estates in Chipinge district, Zimbabwe. The site is located at longitude 32° 41'00E and 32° 42'00E, and latitude 20° 28'00S and 20° 31'00S

(Figure 8.1). The area is characterised by a subtropical climate with two distinct dry and wet seasons, divided almost equally between months of the year i.e. October to March – rainy and April to September – dry seasons. The topography is undulating with a relief difference of over 100m. The area receives relatively high mean annual rainfall totals for a subtropical area (1200-1300 mm/year) with mostly warm temperatures, around 22.5°C (Lagerblad, 2010; Nicolin, 2011). With deep red clayey soils formed from mafic rocks, climatic conditions in the area make it suitable for good quality coffee production.

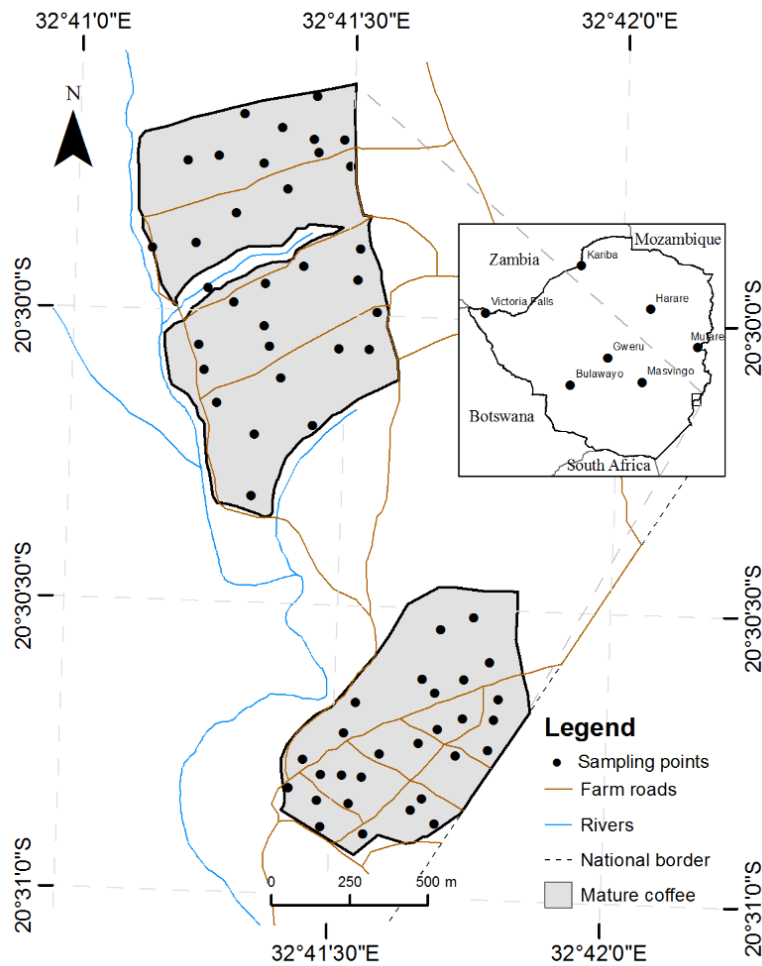


Figure 8.1: The study area showing the distribution of the coffee fields and stand ages

### 8.2.2 Field data

A field campaign for sampling coffee leaf samples was carried out during the first week of January 2017. Sampling points were randomly selected across the coffee fields and a high-resolution imagery sampling map was generated for the identification of sampling sites. The

sampling plots on which data was collected consisted of six coffee plants obtained from two rows of three adjacent plants each. On each tree on the plot, five leaf samples were collected from different canopy levels, making a total of 30 leaves from the six plants. The sampling site was marked with a handheld GPS with an accuracy of ~3m (eTrex 10, Garmin, Germany). The descriptive statistics of the data collected during field work is shown in Table 8.1.

### 8.2.3 Foliar nitrogen determination

Coffee leaves were obtained from the field and stored in a cooler box with ice packs during field work. The leaf samples were oven-dried at 70°C for 8hrs to remove moisture and properly crushed using mortar and pestle. The micro-Kjeldahl method was used for quantitative determination of foliar nitrogen content in coffee leaves using the procedures detailed by Okalebo et al. (2002). 0.5 g of each dried coffee leaf sample was taken for chemical analysis. The foliar nitrogen content of coffee leaf samples was calculated from the amount of ammonia present. The amount of ammonia was determined by titration with sulphuric acid solution with a methyl orange pH indicator. Percentage of foliar nitrogen concentration in coffee leaves was determined using the following formula:

$$Foliar\ nitrogen\ (\%) = \frac{0.014 \times Volume\ of\ H_2SO_4 \times Normality\ of\ the\ H_2SO_4}{Weight\ of\ leaf\ samples} \quad [8.1]$$

Table 8.1: Descriptive statistics of field data

	Age	Height	PWC (%)	N (%)
Min	5	143.3	48.9	1.7
Mean	5.6	165.8	56.7	2.8
Max	8	261.7	68.8	3.8
Std.Dev	1.9	59.5	4.2	0.4
Range	3	128.3	19.8	2.2

### 8.2.4 Image acquisition and preprocessing

Sentinel-2 Level-1C (L1C) MSI data were downloaded from Remote Pixel, a Sentinel-2 Data Hosting portal (<https://remotepixel.ca/>). The Sentinel-2 data for the study area was taken on the 30<sup>th</sup> of November 2016, which was the closest day to field work with usable cloud-free images.

The Sentinel-2 L1C product used in this study is characterized by individual bands of 100 km<sup>2</sup> tiles (ortho-images in UTM/WGS84 projection) detailed in Table 8.2.

Table 8.2: Specific details about the Sentinel-2 MSI image used.

Details	Characteristic
Date of image acquisition	2016-11-30
Day of the Year (DOY)	335
Quantization	12 bits
Cloud cover	0.02%
Data format	.jp2
10m bands	4
20m bands	5

The spectral bands, centre wavelengths, band width and spatial resolution of the Sentinel-2 MSI data used are shown in Table 8.3. To obtain all nine bands at 10m resolution, the 20m bands were pan sharpened in ENVI 5.3 (Exelis Visual Information Solutions, Boulder, CO, USA) using Gram-Schmidt pan-sharpening. The Gram-Schmidt method is generally recommended for most pan-sharpening applications as it is typically more accurate than other methods. This accuracy is attributed to its use of spectral response function of a given sensor to estimate what the higher resolution band would be. All resampling and pan-sharpening were done after converting reflectance to top of the canopy reflectance (Level-2A) using the Sen2cor atmospheric correction module in SNAP (Muller-Wilm et al., 2013). The dark-object subtraction method was applied in QGIS 2.6 using Semi-Automatic Classification Plugin for QGIS as the reflectance values were low after atmospheric correction in SNAP. Only mature coffee data was used, with the mature coffee being identified from an existing age mask.

Table 8.3: Specifications of the Sentinel-2 Multispectral Instrument (MSI) band settings showing centre wavelengths, band width and spatial resolution.

Spectral band	Name	Centre wavelength (nm)	Band width (nm)	Original Spatial resolution (m)
B2	Blue	490	65	10
B3	Green	560	35	10
B4	Red	665	30	10
B5	Red-Edge 1	705	15	20
B6	Red-Edge 2	740	15	20
B7	Red-Edge 3	783	20	20
B8	NIR	842	115	10
B11	SWIR1	1610	90	20
B12	SWIR2	2190	180	20

### 8.2.5 Vegetation indices

Nine vegetation indices were evaluated for N prediction in coffee plantations as listed in Table 8.4. The selection of these vegetation indices was based on performance evaluation of Sentinel-2 based vegetation indices in characterising coffee health (Chemura et al., 2016) or Chapter 6). Other vegetation indices were obtained from reported Sentinel-2 performance evaluations in vegetation biochemical assaying principally from Frampton et al. (2013).

Table 8.4: Name, formula and Sentinel-2 bands utilized in selected vegetation indices (VIs) evaluated in the study.

Name	Formula	Bands	Source
Normalized Difference Vegetation Index	$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$	B8, B4	Rouse et al. (1973)
Green Chlorophyll Index	$GCI = \left( \frac{\rho_{NIR}}{\rho_{GREEN}} \right) - 1$	B8, B3	Gitelson et al. (2005)
Renormalized Normalized Difference Vegetation Index	$RNDVI = \frac{\rho_{NIR} - \rho_{RED}}{\sqrt{\rho_{NIR} + \rho_R}}$	B8, B4	Gitelson and Merzlyak (1994b)
Simplified Canopy Chlorophyll Content Index	$SCCCI = \frac{NDVI.RE3}{NDVI}$	B8, B4, B7	Barnes et al. (2000)
Red-edge Chlorophyll Index	$CI_{RE1} = \left( \frac{\rho_{NIR}}{\rho_{RE1}} \right) - 1$	B8, B5	Gitelson et al. (2005)
Sentinel-2 Red-Edge Position	$S2REP = 705 + 35 * \frac{\left( \frac{\rho_{NIR} + \rho_R}{2} \right) - \rho_{RE1}}{\rho_{RE2} - \rho_{RE1}}$	B4, B5, B6, B8	Frampton et al. (2013)
MERIS Terrestrial Chlorophyll Index	$MTCI = \frac{(\rho_{NIR} - \rho_{RE1})}{(\rho_{RE1} - \rho_R)}$	B4, B5, B8	Dash and Curran (2004)
Normalised Difference Infrared Index	$NDII = \frac{(\rho_{NIR} - \rho_{SWIR1})}{(\rho_{NIR} + \rho_{SWIR1})}$	B8, B11	Hardisky et al. (1983)
Inverted Red-Edge Chlorophyll Index	$IRECI = \frac{(\rho_{NIR} - \rho_R)}{(\rho_{RE2} / \rho_{RE1})}$	B4, B5, B6, B8	Frampton et al. (2013)

### 8.2.6 Foliar N Prediction algorithm and its implementation

The Random Forest (RF) algorithm was used for modelling coffee foliar N from different spectral bands and vegetation indices in their optimized and unoptimized combinations. RF is an ensemble machine learning algorithm developed by Breiman (2001) to solve regression problems through a multitude of decision trees. RF employs an iterative bagging (bootstrap aggregation) operation where the number of trees (*n<sub>tree</sub>*) are independently built using a random subset of samples from the training samples. Each node is then split using the best, among a subset of input variables (*m<sub>try</sub>*). Each tree is then independently grown to a maximum size based on a bootstrap

sample of about two-thirds the training dataset. Each node is then split using the best, among a subset of input variables (*mtry*). The ensemble then classifies the data that are not in the trees as out-of-bag (OOB) data, and by averaging the OOB error rates from all trees, the RF algorithm gives an error rate called the OOB classification error for each input variable. This way, the RF algorithm assesses the importance of each input variable to the outcome by comparing how much the OOB error increases when a variable is removed, while all others are left unchanged (Breiman & Cutler, 2007; Adelabu et al., 2013).

In many applications, this algorithm produces one of the best accuracies to date and has important advantages over other techniques in terms of ability to handle highly non-linear data, robustness to noise and tuning simplicity (Rodriguez-Galiano et al., 2012; Lebedev et al., 2014). The default number of trees (*ntree*) of 500 was used while *mtry* is automatically determined as the square root of the total number of variables used (Breiman, 2001). Coffee N modelling and mapping was done in R (R Core Team, 2013), using the R package randomForest to run the RF modelling (Liaw et al., 2009) for both training and prediction. The coffee foliar diagnostic norms for mature field coffee obtained from Logan and Biscoe (1987) were used for describing the predicted N as low N level (<2.5%), sufficient N (2.5-3.5%) and high N (>3.5%).

### 8.2.7 Accuracy assessment

In order to assess the performance of all models in coffee foliar N estimation, *k*-fold cross validation (with 10 folds) was used. The correlation coefficient (*r*) and coefficient of determination (*R*<sup>2</sup>) were used to assess the goodness of fit of the predicted and measured coffee foliar N. In addition, Mean Absolute Error (MAE, Equation 8.2) Root Mean Square Error (RMSE, Equation 8.3), and percent bias (pBias, Equation 8.4) were used to determine the errors of the model in predicting coffee foliar N from both spectral bands and vegetation indices.

$$\text{MAE} = \left( \frac{1}{n} \sum |y_i - \hat{y}_i| \right) \quad [8.2]$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad [8.3]$$

$$\text{pBias} = \left( \frac{\sum (y_i - \hat{y}_i) * 100}{\sum y_i} \right) \quad [8.4]$$



where for all cases  $n$  is the number of data points,  $y_i$  is the measured foliar N content at that data point and  $\hat{y}_i$  is the model predicted coffee foliar N content at that data point (Moriassi et al., 2007; DeJonge et al., 2016).

### 8.3 Results

#### 8.3.1 Correlation between bands and vegetation indices with foliar N

The relationship between Sentinel-2 MSI bands and derived vegetation indices is shown in Table 8.3. There was a general poor relationship between N and most of the Sentinel 2 MSI bands and vegetation indices ( $R^2 < 0.7$  for both bands and vegetation indices). Surprisingly, the NIR band had the strongest relationship to coffee foliar N ( $R^2 = 0.57$ ) compared to red-edge bands. The spectral bands in the visible and the shortwave region of the spectrum had weak relationships with N (Table 8.5). For vegetation indices, the results indicate that IRECI was more related to N followed by RNDVI while MTCI was the least related to coffee foliar N. The performance of vegetation indices showed that the number of spectral bands in the index is less important to the arrangement of the bands in that index as IRECI and S2REP had exactly the same spectral bands but are contrastingly related to N. Of two band ratios, RNDVI was more strongly related to N ( $R^2 = 0.48$ ) but has the same spectral bands as NDVI which was not strongly related to N.

Table 8.5: Relationship between Sentinel-2 MSI variables and coffee foliar N.

Spectral Bands			Vegetation indices		
Name	Band	$R^2$	Name	Bands	$R^2$
B	B2	0.108	$CI_{RE}$	B8, B5	0.281
G	B3	0.184	GCI	B8, B3	0.337
R	B4	0.32	IRECI	B4, B5, B6, B8	0.668
RE1	B5	0.127	MTCI	B4, B5, B8	0.107
RE2	B6	0.496	NDII	B8, B11	0.376
RE3	B7	0.429	NDVI	B8, B4	0.152
NIR	B8	0.572	RNDVI	B8, B4	0.484
SWIR1	B11	0.069	S2REP	B4, B5, B6, B8	0.279
SWIR2	B12	0.242	SCCCI	B8, B4, B7	0.341

### 8.3.2 Foliar N prediction with spectral bands

The results of predicting coffee foliar N with spectral bands only are shown in Table 8.6. Optimization of variables slightly increased performance of modelling coffee foliar N with spectral bands (Figure 8.2). The performance of the RF model in predicting coffee foliar N with all spectral bands is shown in Figure 8.2a. The map showing the N levels predicted from using all spectral bands in the field is shown in Figure 8.3a. Compared to other variables, using all variables resulted in overestimation of coffee foliar N levels, which explains the low performance for this set of variables ( $R^2=0.57$  and  $RMSE=0.32$ ).

Table 8.6: Accuracy of coffee foliar N prediction with all and optimized Sentinel-2 bands.

Accuracy metrics	All bands	Optimum bands
# of variables	9	4
MAE	0.27	0.28
RMSE	0.32	0.32
NRMSE %	67.3	66.7
r	0.75	0.76
$R^2$	0.57	0.58

Using the RF *oob* error, only four spectral bands, which are RE2 (B6), RE3(B7), NIR(B8) and SWIR2(B12), were identified as significant in modelling coffee foliar N (Figure 8.4). It is interesting that the RF model identified two bands in the red-edge region as influencing N distribution and yet these were less correlated to N compared to the NIR band. Also, it was surprising that the RF results put the SWIR2 band in higher importance than the red, red-edge-1 and green bands in N estimation. Using the four optimized variables did not considerably improve the prediction accuracy of N using Sentinel-2 spectral bands (Table 8.6, Figure 8.2b). The map of predicted N from using optimized spectral bands only is shown in Figure 8.3b). This map was similar to that obtained using all Sentinel-2 MSI spectral bands, in terms of areas with low and sufficient N levels.

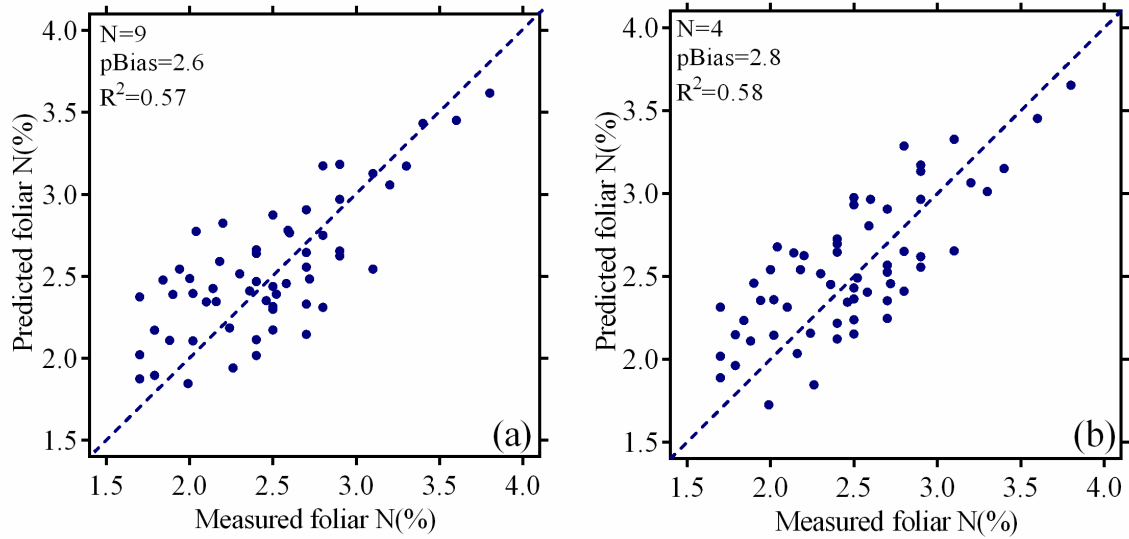


Figure 8.2: 1:1 plot showing the relationship between measured and predicted coffee foliar N using (a) all spectral bands and (b) optimized spectral bands.

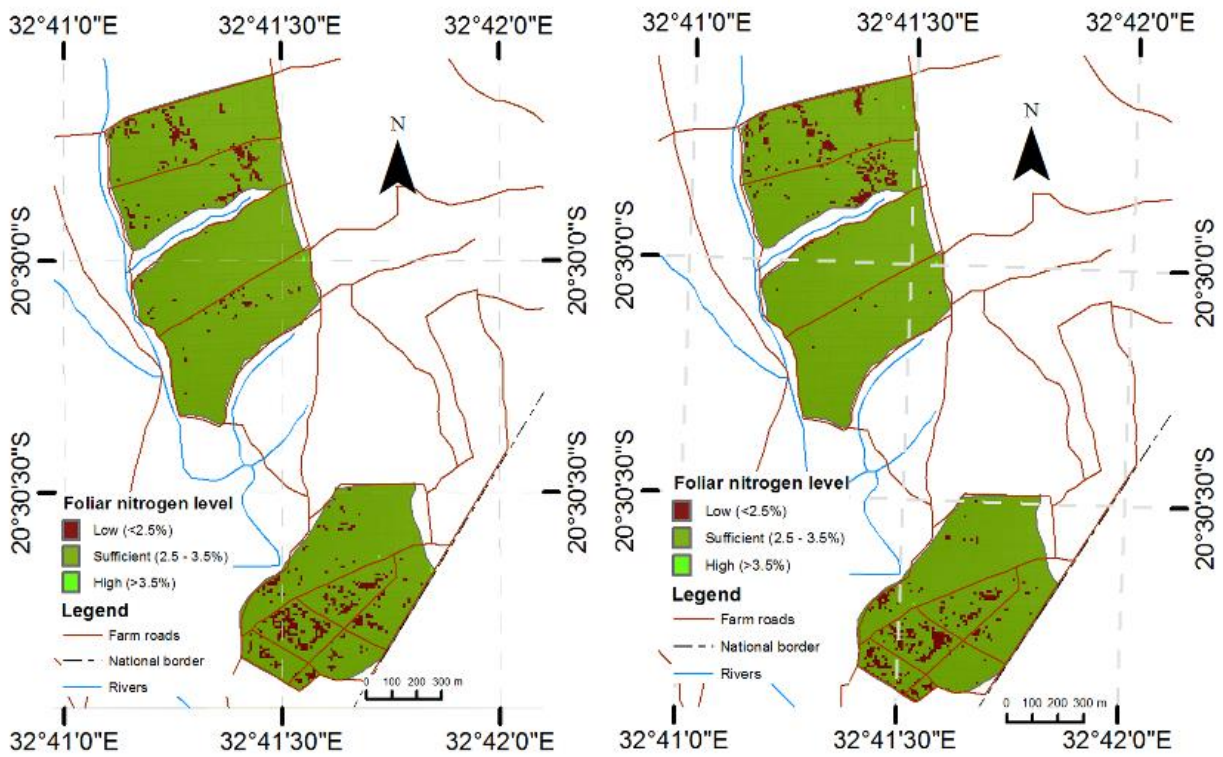


Figure 8.3: Predicted distribution of foliar nitrogen levels in coffee leaves obtained from (a) modelling with all spectral bands and (b) modelling with optimized spectral bands.

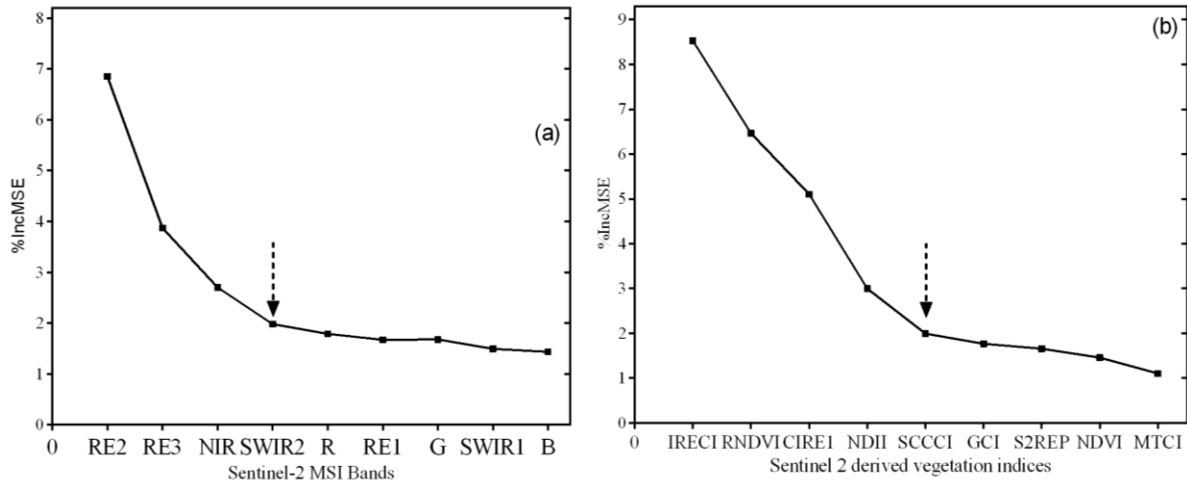


Figure 8.4: Variable optimization for prediction of coffee foliar N using (a) spectral bands and (b) vegetation indices.

### 8.3.3 Foliar N prediction with Sentinel-2 vegetation Indices

Predicting coffee foliar N with vegetation indices considerably improved the prediction accuracy compared to using Sentinel-2 MSI spectral bands only (Table 8.7, Figure 8.5). The RF *oob* error identified five vegetation indices as important in the modelling of coffee foliar N from Sentinel-2 MSI data. All the optimized vegetation indices had the NIR band while three had spectral bands from the red-edge portion of the spectrum.

Table 8.7: Performance of predicting coffee foliar N with all and optimized Sentinel-2 vegetation indices.

Accuracy metrics	All vegetation indices	Optimum vegetation indices
# of variables	9	5
MAE	0.22	0.2
RMSE	0.27	0.25
NRMSE %	55.9	51.8
r	0.84	0.85
R <sup>2</sup>	0.71	0.73

As with spectral bands, optimisation of vegetation indices slightly improved the performance of modelling coffee foliar N. The N levels distribution map produced from using all nine vegetation indices is shown in Figure 8.6. It was interesting that using all vegetation indices had the highest

positive bias (+2.9) indicating overestimation of foliar N. On the contrary, optimized vegetation indices had the least bias of all variables used (+0.4), indicating no systematic shifts in foliar N prediction (Figure 8.5). While the general areas with low N levels were similar to those obtained from using spectral bands, when optimized spectral bands were identified, the areas predicted as having high N content were more than those from other approaches (Figure 8.6). In addition, other areas in the middle block were only identified as having low foliar N by optimized vegetation indices alone.

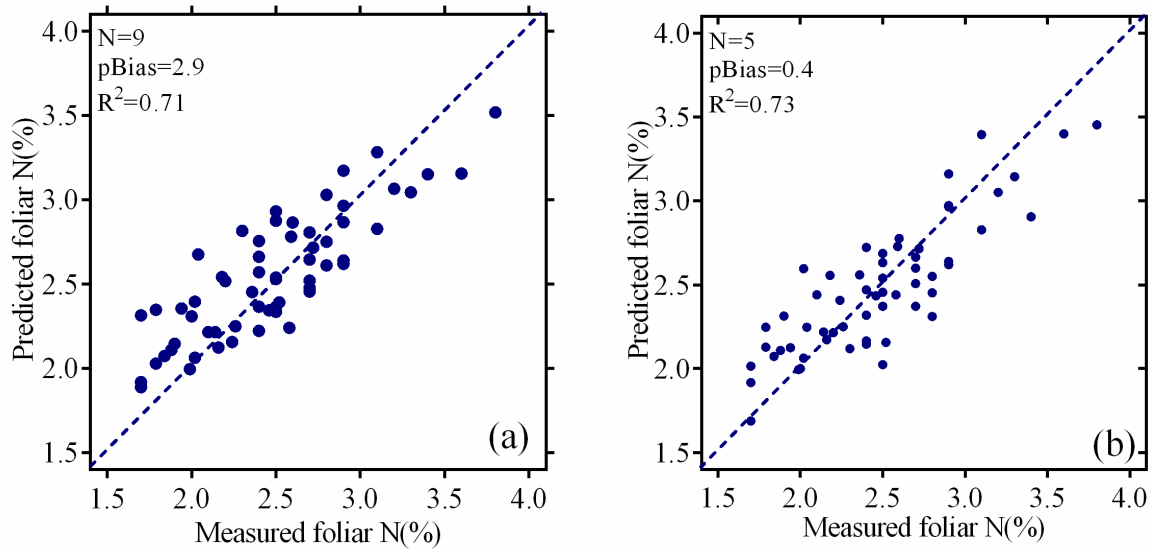


Figure 8.5: 1:1 plot showing the relationship between the measured and the predicted foliar N levels with (a) all nine vegetation indices and (b) five optimized vegetation indices.

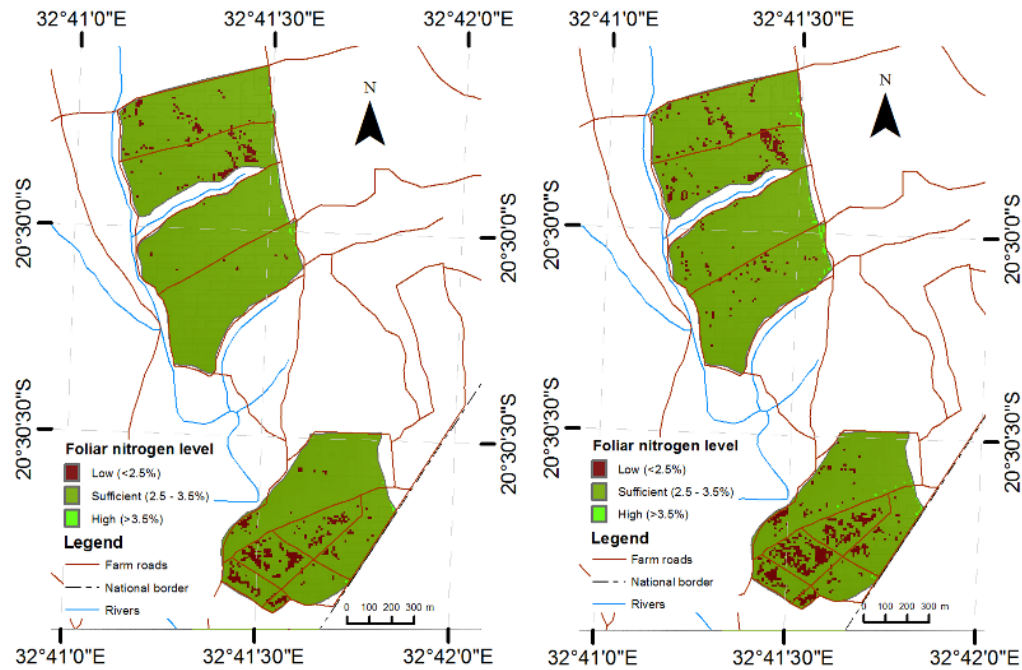


Figure 8.6: Predicted distribution of foliar nitrogen levels in coffee leaves obtained from (a) modelling with all vegetation indices and (b) modelling with optimized vegetation indices.

### 8.3.4 Foliar N prediction with combination of bands and vegetation indices

The optimum spectral bands (N=4) and the optimum vegetation indices (N=5) were stacked together and used to predict foliar N distribution in coffee fields with the RF (N=9). A combination of optimal bands and vegetation indices performed the best as shown by the highest accuracy metrics and the least error metrics (Table 8.8). All N prediction models showed that they underestimated N when values were over 3.0%.

Table 8.8: Performance of predicting coffee foliar N with a combination of optimal bands and vegetation indices.

Accuracy metrics	Optimum bands and vegetation indices
# of variables	9
MAE	0.19
RMSE	0.23
NRMSE %	48.5
r	0.89
R <sup>2</sup>	0.78

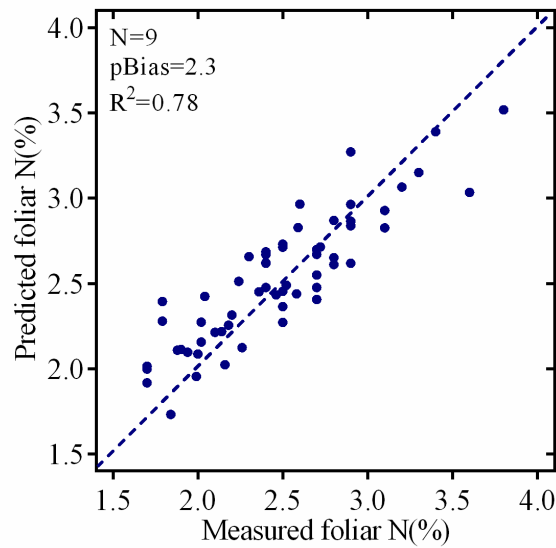


Figure 8.7: 1:1 plot showing the relationship between the measured and the predicted foliar N levels using a combination of optimal bands and vegetation indices.

The distribution of predicted N levels from optimized bands and vegetation indices is shown in Figure 8a. Compared to other N modelling approaches, a larger area was predicted as having low N levels when a combination of optimal bands and vegetation indices were used (Figure 8.8b).

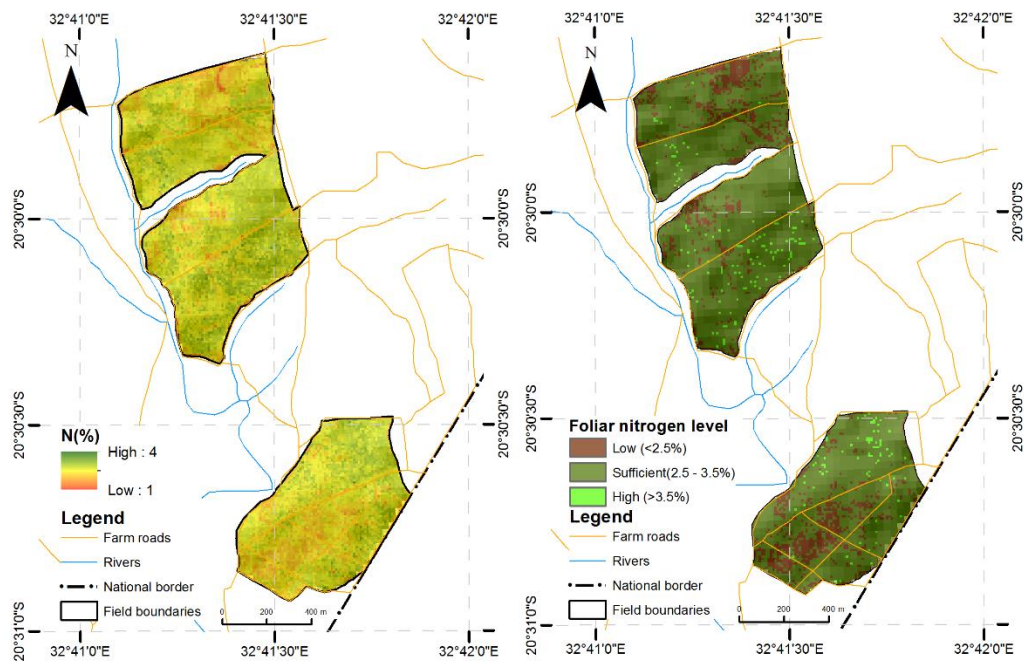


Figure 8.8: Predicted distribution of foliar nitrogen levels from combination of optimal bands and vegetation indices showing (a) variation in foliar N and (b) distribution of N levels.

### 8.3.5 Mapping of low foliar nitrogen areas in coffee plantations

Areas with low, sufficient and high foliar N levels are shown in Table 8.8, with Figure 8.8 comparing positive (high foliar N) and negative (low N) anomalous areas. Using optimized bands had the highest area considered to have low N levels (11% of the study area) while concurrently having the least area with positive anomalies of high foliar N levels (0.14 ha). Contrary, using all spectral bands identified a greater part of the area as having sufficient foliar N levels (Table 8.8). 8.5 ha (2.6%) of coffee area was identified as having low N levels by all datasets. However, considering that vegetation indices and combination of optimized variables produced better accuracy, an intersection of these showed that 15.2 ha (4.7%) of the mature coffee area in the study area have low foliar N levels (<2.5%).

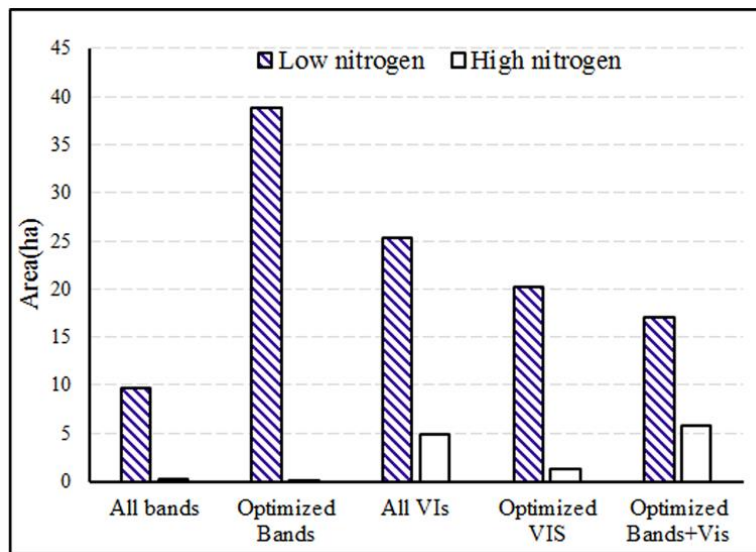


Figure 8.9: Comparison of coffee area with positive (high foliar N) and negative (low N) anomalous foliar N levels from different modelling approaches.



Table 8.9: Sizes (in ha) and % of low, sufficient and high coffee foliar N levels.

Approach	Level	Low nitrogen	Sufficient nitrogen	High nitrogen	Total
All bands	Area(ha)	9.7	342.9	0.2	352.9
	%	2.8	97.2	0.1	100.0
Optimized Bands	Area(ha)	38.8	313.9	0.1	352.9
	%	11.0	89.0	0.0	100.0
All VIs	Area(ha)	25.3	322.6	4.9	352.9
	%	7.2	91.4	1.4	100.0
Optimized VIS	Area(ha)	20.2	331.3	1.3	352.9
	%	5.7	93.9	0.4	100.0
Optimized Bands + Vis	Area(ha)	17.0	330.0	5.9	352.8
	%	4.8	93.5	1.7	100.0

## 8.4 Discussion

The aim of this study was to assess the feasibility for Sentinel-2 spectral bands and vegetation indices to predict foliar N content in coffee plantations with the view of identifying areas that require management attention. The study also sought to determine the most valuable spectral bands and vegetation indices for coffee foliar N prediction to be used for optimisation of the machine-learning algorithm used in prediction. Such determination allows for adoption of remotely sensed data for provision of timely, repeatable, spatially explicit, easier to implement and cost-effective means of managing nutrient dynamics in coffee plantations.

### 8.4.1 Identifying Sentinel-2 bands and vegetation indices related to coffee foliar N

The relationship between Sentinel-2 MSI spectral bands and coffee foliar N, where the red-edge and NIR bands dominated the list of mostly correlated bands was not surprising. This finding was expected as the region's relationship to vegetation N content is already established (Clevers & Kooistra, 2012; Schlemmer et al., 2013b). However, it was anticipated that the red-edge band's relationship to coffee foliar N would be stronger than the NIR bands, as the red-edge portion is known for its value in biochemical assessment. Two possible reasons may explain our finding.

First, the Sentinel-2 NIR band has almost 6-8 times larger band width (115 nm) of the red-edge bands (15 -20 nm), making it more sensitive to vegetation characteristics. Secondly, coffee stands are characterised by soil background due to planting arrangement, as such, the recorded reflectance is dominated by impure vegetation pixels. In this case the NIR is better suited for foliar N determination than the red-edge bands (Frampton et al., 2013).

In this study, the RF variable optimisation ranked the red-edge bands above the NIR in coffee foliar N prediction. These findings are consistent with findings by Clevers and Gitelson (2013) who demonstrate the value of the red-edge bands in characterising specific leaf biochemical properties on maize and grassland, due to their narrower bands that mimic hyperspectral capabilities. Although relatively low, the relationship of the Sentinel-2 SWWIR2 band (B12, 2190 nm) is explained by the reported N absorption features in the SWIR range at 2180 nm due to leaf lignin, cellulose, starch, and proteins (Curran, 1989). This importance was confirmed by the RF variable optimization.

The IRECI, which utilises Sentinel-2 B4, B5, B6 and B8, was the most important variable related to N, as shown by correlation and RF variable ranking. Since this study focused on coffee foliar N at canopy level, this result was expected as this is the only index among the selected vegetation indices that utilises at least two red-edge bands (RE2 and RE1) together with the traditional NIR and R bands. It can therefore be argued that IRECI performs well in canopy biochemical characterisation because it uses the red-edge bands to characterise the red-edge slope while concurrently also making use of the contrasting maximum and minimum reflectance found in the NIR and red bands. A similar strong ( $R^2=0.87$ ) IRECI relationship was reported in chlorophyll estimation (which is directly related to N) by Frampton et al. (2013) at canopy level. However, this relationship diminished when using leaf level estimations ( $R^2=0.24$ ). Whether the strong relationship between IRECI and N is limited to canopy level in coffee, as reported for chlorophyll requires further investigation. Since IRECI is more directly related to chlorophyll estimation, it was concluded that the observed high performance of this vegetation index is more related to the biological associations between nitrogen, chlorophyll, and canopy structure. This is because N is present as the fundamental unit in proteins, nucleic acids, chlorophyll and other organic components in coffee plants (Wang et al., 2016).

It is interesting that the S2REP, which utilises similar bands as IRECI did not perform well in both correlations and variable optimisation. From this, it can be concluded that it is the arrangement of the bands, rather than the bands themselves that are more important in coffee foliar N estimation when using vegetation indices. Thus, the type of transformation adopted by a vegetation index determines how it performs in its analysis of leaf biochemical properties such as N. The MTCI performed poorly for correlation analysis and variable ranking.

#### **8.4.2 Quantification and mapping coffee foliar N distribution**

Our findings demonstrate the feasibility of using Sentinel-2 MSI spectral bands and vegetation indices in determining foliar N in coffee. However, it is apparent that the use of vegetation indices produces better accuracy than use of spectral bands in coffee foliar N estimation. Chemura et al. (2016) reported similar performance from Sentinel-2 derived vegetation indices in discriminating coffee health levels (Chapter 6). They attributed this to vegetation indices ability to better deal with confounding factors such as reflectance saturation, leaf area, roughness and moisture in the leaf and canopy, which reduce the performance of raw spectral bands. However, this could also be attributed to the Sentinel-2's sensor design that leads to reduced band and vegetation indices saturation in hedgerow coffee that accounts for soil or weeds background reflectance. This saturation has been demonstrated in closed canopy forests and in densely planted agricultural crops such as maize (Schlemmer et al., 2013b). A combination of optimal bands and vegetation indices benefits from the strengths of most important variables and thus achieves the best results. This is possible only in machine learning methods such as the RF, which are able to deal with collinearity as the vegetation indices are products of the individual bands.

While the results for modelling coffee foliar N at landscape scale are promising, it is important to understand the potential effects of omission or commission errors to costs of production and the environment. For instance, using optimized Sentinel-2 spectral bands produced a larger area with low N (below 2.5%). Thus, at current applications rates of 270kg/ha of ammonium fertilisers per ha<sup>-1</sup> per year (Logan & Biscoe, 1987), these low areas will require an additional 10.5t of fertilisers, which is four times more than the requirement obtained from using all Sentinel-2 spectral bands (2.5t). While these two results had the least accuracy (all bands and optimized bands), they demonstrate that the results have to be understood in context to avoid increased costs and environmental damage from prediction errors arising from an approach that

seeks to achieve the converse. Nonetheless, the results have demonstrated the possibility of site based nutrient management, specifically for N (and its associated benefits) in coffee using multispectral Sentinel-2 data.

## **8.5 Conclusions**

This study has assessed the feasibility of Sentinel-2 MSI spectral bands and vegetation indices in empirical estimation of coffee foliar N content at landscape level with satellite data. From this study, it was concluded that:

- 1) Coffee foliar N is strongly related to Sentinel-2 MSI red-edge (B6 and B7), NIR (B8) and SWIR2 (B11) bands and to IRECI, RNDVI,  $CI_{RE1}$  and NDII vegetation indices that can be used for its prediction.
- 2) Using vegetation indices produces better accuracy in coffee foliar N modelling compared to using spectral bands. Selecting optimal bands and vegetation indices produces the best results and is recommended for coffee foliar N estimation from Sentinel-2 MSI data.
- 3) It is possible to use the derived N distribution maps to make site specific nutrient management in coffee plantations from using Sentinel-2 MSI data.

## **CHAPTER 9: MODELLING SPATIAL VARIABILITY OF COFFEE CROP CONDITION WITH MULTISPECTRAL REMOTE SENSING DATA: A SYNTHESIS**



Photo credits: A. Chemura (2017)

## 9.1 Introduction

Coffee is an important commodity in international agricultural trade especially for producer countries that depend on it for off-setting balance of payments, local taxes and livelihoods. The major players in the coffee sector are producers who struggle with climatic, edaphic and other biotic factors to produce coffee of sufficient quantity and quality to satisfy a demand of over 2 billion coffee cups consumed daily (Waston & Achinelli, 2008). Biotic and abiotic calamities are increasingly affecting the coffee sector, with these explaining fluctuations in world coffee supplies and market prices (Van der Vossen, 2001). Given the importance of coffee in developing producer economies, consumer countries and ecosystems, there is a need for systematic landscape-scale monitoring of coffee crop condition to increase productivity through reducing impacts of potential stressors. These monitoring mechanisms are required for assisting on farm decisions especially for large farms, for insurance and risk services and national coffee sector planning. Monitoring coffee condition is also important in precision farming in which case agrochemicals are efficiently applied to reduce costs and associated environmental impacts from over-application of fertilizers and pesticides. However, current condition assessment methods in the coffee sector are based on sampling that assumes the selected sites represent the whole field and also on subjective visual assessments that are laborious and conclusive once economic damage has been inflicted.

Developments in remote sensing data acquisition, storage and processing provide an opportunity for spatially explicit and objective vegetation condition assessments. However, there are a number of challenges in applications of remote sensing coffee for crop area mapping and condition assessment that render traditional remote sensing approaches limited in coffee producing areas. These challenges include the fact that coffee is a perennial crop and as such requires long-term monitoring beyond one season. Secondly, at landscape scale, producers plant coffee of different ages, with age of the coffee plants influencing percentage canopy cover and its resultant effect on background effect. There is also a systematic planting arrangement in hedgerows that ensures gaps between rows for efficient farm management operations but also results in background spectral effect (Brunsell et al., 2009). In addition, the LAI of coffee follow biennial cycle that is influenced by fruiting sequence and this influences applications of remote sensing applications (Bernardes et al., 2012). Coffee stressors can also concurrently occur, making it difficult to separate them spectrally or to understand the root cause of the problem.

(Logan & Biscoe, 1987; Kutuywayo et al., 2013). Given the need for crop condition assessment in resource-constrained coffee producer countries and the challenges from the nature of the crop and the production system, the objectives of this work were to:

1. Develop detailed age-specific thematic maps for coffee for heterogeneous agricultural landscapes for use as age-masks in condition assessments.
2. Integrate age in the identifying and mapping of incongruous patches in coffee plantations using multi-temporal multispectral level anomalies.
3. Evaluate the potential of multispectral remote sensing data for quantifying leaf water stress in coffee using only the secondary effects of water absorption.
4. Assess the potential of multispectral remote sensing bands and vegetation indices for discriminating and predicting coffee leaf rust infection levels at leaf level.
5. Determine the effects of spectral settings, spatial resolution and crop canopy cover on ability of multispectral level data to predict leaf chlorophyll content as an indicator of crop condition in coffee plantations.
6. Model the spatial variability in foliar nitrogen in coffee plantations with multispectral level data for agronomic decision making.

## **9.2 Mapping coffee plantations and development of age-masks**

The fact that coffee plants have different age groups presents challenges not only in separating coffee from other land use/cover types with remote sensing data but also in coffee condition assessments. In addition, remote sensing-based age-specific thematic maps for coffee areas are important for coffee farm planning such as the rate of application of fertilisers and pest control chemicals are applied, which are age-dependent. The objective of this study was to capitalize on sensor improvements in Landsat 8 OLI data and machine learning algorithms in developing age-specific thematic maps for coffee producing areas. The overall outcome is that it is possible to develop age-specific thematic maps for coffee in a heterogeneous agricultural landscape using Landsat 8 OLI and the RF algorithm. The ability to achieve reliable classification accuracy for coffee age-groups (young, mature and old) is based on the significance of the differences in spectral reflectance among coffee classes of different ages and between coffee and other land use/cover types that is captured by higher sensor fidelity of the Landsat 8 OLI. For example, Landsat 8 OLI uses numerous elongated sets of detectors for each waveband, which increases

the sensitivity of the sensor to the age-based biophysical characteristics that determine vegetation reflectance (Dube & Mutanga, 2015).

This study confirmed the importance of the NIR in remote sensing coffee plantations. For example, this study observed that the spectral difference between mature coffee and old coffee was only discernible in the NIR band. It was however clear from this study that, splitting coffee into three age classes reduces the classification accuracy (for example by a 4.1% decrease in overall accuracy). However, the reduction in classification accuracy that occur when coffee is split into three classes is small, indicating that it could be worthwhile to trade-off some accuracy to obtain a more functional thematic map with age-specific classes in mapping complex agricultural landscapes. The success of development of age-specific thematic maps for coffee is premised on the improvements in spectral characteristics of the Landsat 8 OLI sensor, coupled by an intelligent and robust non-parametric classifier, the RF. The developed age-specific thematic maps are useful not only in separating coffee from other land use/cover types, but among different age categories which deals with the challenge of age in coffee area mapping and condition assessment. Development of age masks has been a remote sensing challenge for other perennial agricultural crops (McMorrow, 2001; Tan, 2013; Chemura et al., 2015b). It would be interesting to assess if the same approach can work in coffee under shade or in robusta coffee that has more canopy cover than arabica coffee. The produced age-masks were used in subsequent condition assessment studies.

### **9.3 Identification of anomalous patches in coffee plantations**

As a perennial crop, coffee requires a consistent monitoring framework to monitor biotic and abiotic stressors and their effect on inter and intra-annual variations in crop conditions, but these are currently largely unavailable. Vegetation indices such as NDVI and LSWI can be used to monitor crop conditions over a large area and to detect any significant changes over time. When the “normal” or expected index value is known, then any departure from this indicates an anomaly. The objective of this study was to use the age-specific thematic maps from the preceding chapter to identify areas in coffee fields that are anomalous and therefore requiring attention. An approach using deviation from age-specific mean NDVI and LSWI was developed using Landsat 8 OLI data and applied for identifying, quantifying and mapping between and within field crop stand variations. The outcome of this study was that age has a significant



influence on coffee NDVI and LSWI values associated with physiological development, photosynthetic efficiency, LAI and canopy cover of the coffee plants (Logan & Biscoe, 1987).

Landsat 8 OLI derived NDVI performed better than LSWI in identifying of anomalous areas in coffee fields. This is because NDVI associates well with plant biophysical characteristics that are affected by age, condition and other factors (Ke et al., 2015). Age-adjusted NDVI deviations from the mean were even more accurate in showing areas that are growing either below or above their expected growth stages. It was therefore clear that age-adjustment of NDVI anomalies improved performance of mapping incongruous areas in coffee plantations. In addition, multi-temporal analysis clearly showed how some coffee areas moved from being healthy to being incongruous over the months. This provides a spatial, quantitative and temporal idea of crop performance for every Landsat 8 pixel per scene date that can be used for farm level decision making. The presented age-adjusted anomaly detection identifies where in the coffee plantations condition is deteriorating either through biotic or abiotic stress or improving through management intervention. Although this approach does not identify the exact cause of the anomaly that is important in decision making, this is a challenge dealt with in subsequent studies in this project. Identifying these anomalous areas is important for identifying specific areas requiring attention in large plantation and for more targeted application of remote sensing methods to determine causes at particular sites

Three important advancements in remote sensing were presented in this study; (i) age-adjustment is important in anomaly detection in coffee, (ii) scene-based anomaly detection is more applicable in tropical areas where coffee is produced because of problems of clouds and seasonality and (iii) multi-temporal remote sensing is a good basis for understanding crop dynamics in coffee condition assessment, which is possible with accessible multispectral data such as Landsat 8 OLI and/or Sentinel-2 MSI data compared to many studies that are dependent on only a single scene. Although this method is very promising for long-term monitoring of perennial tree crops such as coffee which are in the field for decades, it forms a basis for more specific studies on specific causes of anomalies in coffee plantations.

#### **9.4 Predicting plant water content in coffee plants using multispectral remote sensing**

Having developed an approach for identifying anomalous patches in coffee plantations, it was imperative to identify, with remote sensing, the specific cause of the anomalies. Experimental evaluation of the possibility to quantify plant water content using secondary water absorption bands in the VIS/NIR region was done. Wavebands that are selected by reflectance sensitivity were the best in predicting PWC with relatively high correlations between measured and predicted PWC on test data and lower error metrics. Reflectance sensitivity identified bands mainly in the red-edge and the visible, particularly the blue region of the spectrum, with none in the NIR region. However, reflectance in this region is influenced mostly by leaf structure, chlorophyll, biomass and stand age. Since this is a single species study using same age healthy plants, these factors are not significant. This also underscores the importance of separating coffee from other land use/cover types and age categories (Chapter 3). The ability of wavebands in the visible and NIR range to predict PWC is very attractive for application of multispectral scanners, in coffee condition assessments. This is so because multispectral scanners such as Sentinel-2 can be focused on specific crop fields and super-spectral bands strategically positioned across important regions of the VIS/NIR bands applied in PWC estimation. It is important to note that this study relied on healthy coffee leaves and under field conditions; other factors such as diseases and pests (Mutanga & Ismail, 2010; Oumar & Mutanga, 2014) may influence potential for PWC estimation. Overall, this study provides a basis for application of remote sensing in identifying and quantifying PWC, which alone can explain anomalies in coffee fields or indirectly as PWC is related to other plant conditions such as pest and disease attacks.

#### **9.5 Modelling coffee leaf rust using multispectral remote sensing**

PWC is an abiotic factor contributing to coffee stress and this study extended the application of remote sensing in biotic stress by experimentally evaluating multispectral remote sensing of coffee leaf rust caused by the fungus *H. vastatrix*. This study has for the first time demonstrated the utility of the new generation Sentinel-2 MSI sensor in discriminating disease infections. The most important Sentinel-2 MSI's spectral bands (i.e. B4, B6 and B5) for CLR discrimination and modelling were identified. The finding that the red band (B4) was the most significant band in CLR discrimination is explained by the fact that CLR is different from other diseases and plant stress conditions in that it is not necrotic (Belan et al., 2015) and also result in distinctive asymptomatic rustic pustules on the underside of the leaf. As expected, vegetation indices

produced better results in CLR discrimination, when compared to the use of raw spectral bands, which was expected. In CLR severity modelling, the proportion of leaf diseased area was not significantly correlated to most spectral bands because of the non-linear influence of disease infection on spectral reflectance (Zhang et al., 2012; Mahlein et al., 2013).

Using the RBF-PLS showed a high accuracy in CLR severity modelling with resampled data. This is an achievement considering that much of the reported work on disease modelling is based on vegetation indices or at least a combination of vegetation indices and spectral bands, even with hyperspectral data. This is attributed to the Sentinel-2 sensor settings as Ramoelo et al. (2015) also used only spectral bands to model nitrogen in rangelands and obtained a high accuracy, confirming that the spectral settings of Sentinel-2 MSI spectral settings are good for modelling vegetation condition. Findings of this study showed higher accuracy for severe levels, but from an operation perspective, this is not good as economic damage would have already been inflicted on the crop. This study underpins the application of Sentinel-2 MSI data in coffee condition assessment that can improve management of croplands and stewardship of the environment through reduced unnecessary use of crop protection chemical for disease control. Although the results are positive in indicating potential application as sensors in disease levels modelling, these results are based on field spectra at leaf level. There is need for more field studies to apply the RF and RBF-PLS in modelling the biophysical (e.g. LAI reduced to defoliation) and biochemical (e.g. chlorophyll and foliar nitrogen) parameters of coffee that are affected by the disease.

## **9.6 Multispectral remote sensing of coffee chlorophyll content**

Chl is important in agricultural crop management as it is a reliable indicator of photosynthetic rates and leaf health as majority of the plant stressors affect this variable. A field level application of multispectral remote sensing data in mapping and quantifying the spatial variability of chlorophyll in coffee fields that can be related to the abiotic (Chapter 5) and biotic (Chapter 6) factors was presented. The aim of this study was to evaluate the influence of Sentinel-2 MSI spectral settings and spatial resolution and coffee stand age in empirical landscape scale coffee Chl estimation. As expected, coffee biophysical attributes (height and canopy area) were strongly related to coffee age while biochemical parameters (PWC and Chl) were age invariant, and thus more influenced by site conditions such as fertility, insects or pests or other local factors.

It was evident from this study that total coffee Chl empirical modelling with Sentinel-2 MSI data is more related to image's spatial resolution than spectral characteristics. This finding is important as the Sentinel-2 MSI data comes at 10m, 20m and 60m spatial resolutions, tied to unique spectral settings. This is particularly critical to Sentinel-2 MSI data as its VIS-NIR data comes in different spatial resolutions, a departure from previous sensor designs. Thus, our finding that spatial resolution is more important in influencing total Chl modelling in coffee casts aspersions on spectroscopic simulation studies using Sentinel-2 MSI data (e.g Chapter 6) that did not consider spatial resolution. Studies in natural forests have also reported that leaf biochemical estimations are more accurate at finer resolutions than at coarser resolutions (Lepine et al., 2016). This finding may be specific to coffee as the coffee plantations do not have closed canopy. Although, coffee leaf biochemical parameters did not considerably vary with age, there was higher accuracy in modelling total coffee Chl in mature coffee stands only, compared to the use of both young and mature ages. This result is because the NIR reflectance is more reflective in mature coffee (Chapter 3). The effects of canopy area (which is directly related to age) were also apparent in influencing both prediction patterns and magnitude of errors. It was concluded therefore that landscape scale modelling of coffee leaf biochemical properties be applied only to mature coffee fields (over 4 years). In large scale modelling, use of pre-existing age masks (Chapter 3) is therefore very useful for practical applications. When coffee leaf Chl was estimated for mature fields only, the best model (all bands at 10m resolution) explained 77% of the variation in measured coffee leaf Chl.

### **9.7 Multispectral remote sensing of spatial variability of coffee foliar nitrogen**

Nitrogen is one of the most important determinants of coffee productivity and general coffee condition. The utility of multispectral level remote sensing data in landscape scale mapping of N in coffee was presented. The red-edge and NIR bands dominated the list of mostly N correlated bands as region's relationship to vegetation N content is already known (Clevers & Kooistra, 2012; Schlemmer et al., 2013a). The NIR was more correlated to N than red-edge bands because of the crop and sensor-related reasons. Crop because coffee pixels are never pure vegetation as there is always background effect and this reduces the sensitivity of the red-edge bands. Sensor-related reasons because Sentinel-2 NIR band has almost 6-8 times larger band width (115 nm) than the red-edge bands (15 -20 nm), making it more sensitive to vegetation characteristics.

The IRECI, which utilises Sentinel-2 B4, B5, B6 and B8, was the most important variable related to N, as shown by correlation and RF variable ranking. Of the selected indices, this was the only index among the selected vegetation indices that utilises at least two red-edge bands (RE2 and RE1) together with the traditional NIR and R bands. Therefore, IRECI uses the red-edge bands to characterise the red-edge slope while concurrently also making use of the contrasting maximum and minimum reflectance found in the NIR and red bands respectively at canopy level (Frampton et al., 2013). It was apparent that the use of vegetation indices produces better accuracy than the use of spectral bands in coffee foliar N estimation. This finding concurs with findings in Chapter 6 on CLR discrimination which reported better performance of vegetation indices. The results have demonstrated the possibility to use the derived N distribution maps to make site specific nutrient management in coffee plantations from Sentinel-2 MSI data.

### **9.8 Implications for the coffee sector**

Applying remote sensing in coffee condition assessment is important for maintaining productivity, profitability, crop health and in safeguarding long-term investments that farmers make in their plantations. This study intended to develop landscape scale spatially-explicit crop condition assessment approaches that are scientifically sound, easier to implement and cost effective. Adoption of these approaches in the coffee sector is valuable in improving coffee productivity, as well as provision of environmental goods and services from coffee plantations. This is increasingly important with increasing production challenges presented by climate change and variability. Landscape scale modelling of the spatial variability in coffee condition with remote sensing provide farm managers, insurance companies, national regulatory boards and other players in the coffee value chain with spatially explicit information, reducing the subjectivity, costs, spatial singularity and inconveniences associated with field methods. Identification of specific coffee stressors such as plant water stress are not just required for keeping growth vigour but are a necessary agronomic practices, such as inducing uniform and high level of flowering in a short window of time, without which yield and quality, two important parameters in determining farmers' returns, are significantly reduced (Masarirambi et al., 2009).

Many opportunities for coffee condition assessment are linked to Sentinel-2 MSI data. With the high temporal resolution, it is possible to model coffee condition changes by combining analysis in both time and space over fields at a lower cost, a feat predecessor sensors could not achieve.

There are therefore opportunities for development near-real time online and mobile apps for coffee condition assessment that can be used by field managers, insurance companies, extension officers and other interested stakeholders. This is possible, especially once the twin Sentinel-2b satellite, with a five-day temporal resolution is in orbit. The important variables in coffee condition assessment has been identified and can be implemented from the shelf with machine learning approaches. Therefore, landscape scale modelling of coffee condition with multispectral remote sensing provides timely, repeatable, spatially explicit, easier to implement and cost-effective means of managing coffee plantations. More efficient management improves production and profitability in the coffee sector, while ensuring environmental stewardship from prudent use of water, fertilisers and agrochemicals. Production targets can be achieved without increasing area under production thereby satisfying the often conflicting goals of achieving economic development and environmental sustainability in the coffee sector.

## **9.9 Conclusions**

The overall aim of this study was to develop landscape scale modelling approaches for assessing the spatial variability in crop condition in coffee plantations with multispectral remote sensing data. Based on the findings from the objectives of this study, the following conclusions are drawn:

1. It is possible to develop detailed age-specific thematic maps for coffee for heterogeneous agricultural landscapes for use as age-masks in condition assessments using Landsat 8 OLI data. Although the accuracy of such maps is lower than that of not disaggregating coffee by class, the potential benefit and utility of such an age-specific map of coffee areas is greater than the cost of the reduced accuracy.
2. Age-specific Landsat 8 derived NDVI anomalies can be reliably used in identifying and mapping of incongruous patches in coffee plantations and help inform areas requiring attention and monitoring coffee condition over time.
3. Secondary effects of water absorption available in the VIS/NIR region of the spectrum can be successfully used in estimating plant water content in coffee to explain anomalies that may occur in coffee plantations.
4. Sentinel-2 MSI bands and vegetation indices can be used for discriminating and predicting biotic stress factors such as coffee leaf rust infection levels at leaf level.

5. The spatial variability in Chl and foliar N in coffee plantations can be reliably modelled for mature coffee plantations with Sentinel-2 MSI data for landscape scale coffee condition assessment and agronomic decision making.

### **9.10 Outlook for future research**

Although the findings of this study are conclusive within the scope of the study, they lay a foundation for further research in this area. The following future research directions are therefore recommended based on this study.

1. Much of the approaches and results from this study are based on empirical models relating spectral features and their transformations to coffee condition without providing a mechanistic understanding of the identified relationships. More research is therefore required to understand the underlying leaf and canopy processes in coffee that influence the identified relationships.
2. There is a need to understand how the coffee biophysical and biochemical properties evolve during coffee growth cycle as coffee is a perennial crop that is in the field throughout the year. For example, it is important to understand if the leaf nutrient balance during and between the vegetative, senescent, fruiting and harvesting stages changes and identify how these changes affect multispectral level remote sensing estimation of coffee condition.
3. There is a need for research on the social and technical possibility of developing decision support systems for coffee production based on multispectral remote sensing data. Operational aspects, technology design aspects and uptake factors need to be investigated.
4. While the Sentinel-2 data used in this approach showed promising results, there is need for investigation of other satellite based (such as WorldView 3) and unmanned aerial vehicles (UAV) or drones in modelling the spatial variability of coffee condition for use in coffee management.
5. Most of the coffee fields used had sun-coffee and it is therefore important to understand how shade trees can influence the estimation of coffee condition with multispectral remote sensing data.

## References

- 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(10), 3440-3458.
- Adelabu, S., Mutanga, O., & Adam, E. (2014). Testing the reliability and stability of the internal accuracy assessment of Random Forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, 28, 1-24.
- Adelabu, S., Mutanga, O., Adam, E., & Cho, M. A. (2013). Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image. *International Journal of Applied Remote Sensing*, 7, 0734801-07348013.
- African Development Bank. (2010). Coffee Production in Africa and the Global Market Situation Commodity Market Brief (Vol. 1, pp. 9). Tunis: African Development Bank.
- Ahmed, A., Sharkawy, M. E., & Ramily, S. E. (2012). Analysis of Inter-band Spectral Cross-Correlation Structure of Hyperspectral Data. Paper presented at the WSEAS International Conference. Proceedings. Recent Advances in Computer Engineering Series.
- Alcaraz-Segura, D., Cabello, J., Paruelo, J. M., & Delibes, M. (2009). Use of Descriptors of Ecosystem Functioning for Monitoring a National Park Network: A Remote Sensing Approach. *Environmental Management*, 43, 38-48.
- Alves, H. M. R., Vieira, T. G. C., Lacerda, M. P. C., Bertoldo, M. A., & Andrade, H. (2006). *Characterization of coffee agroecosystems of the state of Minas Gerais in Brazil*. Paper presented at the XXXV ISPRS Congress.
- Alves, M. d. C., da Silva, F. M., Pozza, E., & de Oliveira, M. (2009). Modeling spatial variability and pattern of rust and brown eye spot in coffee agroecosystem. *Journal of Pest Science*, 82, 137-148.
- Asmala, A. (2012). Analysis of maximum likelihood classification on multispectral data. *Applied Mathematical Sciences*, 6(129-132), 6425-6436.
- Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*, 5, 949-981.
- Atzberger, C., Guerif, M., Baret, F., & Werner, W. (2010). Comparative assessment of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture*, 73, 165-173.
- Avelino, J., Romero-Gurdián, A., Cruz-Cuellar, H. F., & Declerck, F. A. (2012). Landscape context and scale differentially impact coffee leaf rust, coffee berry borer, and coffee root-knot nematodes. *Ecological applications*, 22(2), 584-596.
- Avelino, J., Willocquet, L., & Savary, S. (2004). Effects of crop management patterns on coffee rust epidemics. *Plant Pathology*, 53(5), 541-547.
- Baker, P., & Hagggar, J. (2007). *Global warming: Effects on global coffee*. SCAA Conference Handout, Long Beach.
- Balasundram, S. K., Memarian, H., & Khosla, R. (2013). Estimating oil palm yields using vegetation indices derived from Quickbird. *Life Science Journal*, 10(4), 851-860.
- Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *Springer Plus*, 2(660), 1-12.
- Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems *Advances in Land Remote Sensing* (pp. 173-201): Springer.
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 35, 161-173.
- Baret, F., Houle's, V., & Gue´rif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany*, 58( 4), 869-880.



- Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S. (2000, 16-19 July 2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. Paper presented at the 5th International Conference on Precision Agriculture, Bloomington, Madison, WI, USA.
- Bausch, W. C. (1993). Soil background effects on reflectance-based crop coefficients for corn. *Remote Sensing of Environment*, 46(2), 213-222.
- Bedimo, J. A. M., Bieysse, D., Cilas, C., & Nottéghem, J. L. (2008, 14 -19 September 2008). Effect of temperature and rainfall variations on coffee berry disease (*Colletotrichum kahawae*). Paper presented at the 22nd International Conference on Coffee Science, Campinas, Brazil.
- Bedimo, J. A. M., Bieysse, D., Cilas, C., & Nottéghem, J. L. (2007). Spatio-temporal dynamics of Arabica coffee berry disease caused by *Colletotrichum kahawae* on a plot scale. *Plant Disease*, 91, 1229-1236.
- Belan, L. L., Pozza, E. A., de Oliveira Freitas, M. L., Pozza, A. A. A., de Abreu, M. S., & Alves, E. (2015). Nutrients distribution in diseased coffee leaf tissue. *Australasian Plant Pathology*, 44(1), 105-111.
- Bernardes, T., Moreira, M. A., Adami, M., Giarolla, A., & Rudorff, B. F. T. (2012). Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. *Remote Sensing*, 4(9), 2492-2509.
- Bhojaraja, B. E., Shetty, A., Nagaraj, M. K., & Manju, P. (2015). Age-based classification of arecanut crops: a case study of Channagiri, Karnataka, India. *Geocarto International*.
- Bispo, R. C., Lamparell, R. A. C., & Rocha, J. V. (2014). Using fraction images derived from MODIS data for coffee crop mapping. *Journal of Agricultural Engineering*, 34, 102-111.
- Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4), 855-867.
- Bolanos, S. (2007). Using Image Analysis and GIS for coffee Mapping (Master of Science ), McGill University, Montreal.
- Bortolotto, R. P., Bruno, I. P., Reichardt, K., Timm, L. C., Amado, T. J. C., & Ferreira, A. d. O. (2012). Nitrogen fertilizer (15N) leaching in a central pivot fertigated coffee crop. *Revista Ceres*, 59(4), 466-475.
- Brauman, K. A., Daily, G. C., Duarte, T. K., & Mooney, H. A. (2007). The nature and value of ecosystem services: An overview highlighting hydrologic services. *Annual Review of Environment and Resources*, 32, 67-98.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman, L., & Cutler, A. (2007). Random forests-classification description. *Department of Statistics, Berkeley*.
- Brenning, A. (2009). Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sensing of Environment*, 113(1), 239-247.
- Brillante, L., Bois, B., Mathieu, O., & Lévêque, J. (2016). Electrical imaging of soil water availability to grapevine: a benchmark experiment of several machine-learning techniques. *Precision Agriculture*, 17, 1-22.
- Brown, H. B. (2008). Smallholder coffee production in Zimbabwe. Harare: CRS Press.
- Brown, J. K. M., & Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297, 537-541.
- Brunsell, N., Pontes, P., & Lamparelli, R. (2009). Remotely sensed phenology of coffee and its relationship to yield. *GIScience & Remote Sensing*, 46(3), 289-304.
- Campos, R. C., Epiphanyo, J. C. N., & Formaggio, A. R. (2005). *Spectral variability of coffee crop related to scene fractional components estimated through high resolution image*. Paper presented at the Anais XII Simposio Brasileiro de Sensoriamento Remoto, Goiania, Brazil.

- Cao, Q., Miao, Y., Shen, J., Yu, W., Yuan, F., & Cheng. (2015). Improving in-season estimation of rice yield potential and responsiveness to topdressing nitrogen application with Crop Circle active crop canopy sensor. *Precision Agriculture*, 16.
- Carter, G. A. (1991). Primary and secondary effects of water content on the spectral reflectance of leaves. *American Journal of Botany*, 916-924.
- Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, 88(4), 677-684.
- Carter, G. A., & Miller, R. L. (1994). Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment*, 50(3), 295-302.
- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Paper presented at the 23rd international conference on Machine learning.
- Castillejo-González, I. L., Peña-Barragán, J. M., Jurado-Expósito, M., Mesas-Carrascosa, F. J., & López-Granados, F. (2014). Evaluation of pixel- and object-based approaches for mapping wild oat (*Avena sterilis*) weed patches in wheat fields using QuickBird imagery for site-specific management. *European Journal of Agronomy*, 59(0), 57-66.
- Chander, G., Markham, B., & Helder, D. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5), 893-903.
- Chemura, A. (2014). The growth response of coffee (*Coffea arabica* L) plants to organic manure, inorganic fertilizers and integrated soil fertility management under different irrigation water supply levels. *International Journal of Recycling of Organic Waste in Agriculture*, 3(2), 1-9.
- Chemura, A., Kutwayo, D., Chidoko, P., & Mahoya, C. (2015a). Bioclimatic modelling of current and projected climatic suitability of coffee (*Coffea arabica*) production in Zimbabwe. *Regional Environmental Change*, 1-13.
- Chemura, A., Mahoya, C., Chidoko, P., & Kutwayo, D. (2014). Effect of soil moisture deficit stress on biomass accumulation of four coffee (*Coffea arabica*) varieties in Zimbabwe. *ISRN Agronomy*, 2014, 1-10.
- Chemura, A., & Mutanga, O. (2015). *Comparing physiological and spectral response of coffee (Coffea Arabica L.) to induced soil moisture stress for assessing potential for remote sensing*. Paper presented at the 16th WaterNet/WARFSA/GWP-SA Symposium, Balaclava, Pointe Aux Piments, Mauritius.
- 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*, Inpress.
- Chemura, A., Mutanga, O., & Dube, T. (2016). Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precision Agriculture*, 17, 1-23.
- Chemura, A., Mutanga, O., & Dube, T. (2017). Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests. *Physics and Chemistry of the Earth ABC*, In Press.
- Chemura, A., van Duren, I., & van Leeuwen, L. M. (2015b). Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, Ghana. *ISPRS Journal of Photogrammetry and Remote Sensing*, 100, 118-127.
- Chidoko, P., Mahoya, C., Kutwayo, D., & Chemura, A. (2017). A study of genetic diversity of cultivated coffee varieties in Zimbabwe using morpho-agronomic markers. Paper presented at the Zimbabwe International Research Symposium, HICC, Harare.
- Clark, M. L., & Aide, T. M. (2011). Virtual interpretation of earth web-interface tool (VIEW-IT) for collecting land-use/land-cover reference data. *Remote Sensing*, 3(3), 601-620.

- Clevers, J. G., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344-351.
- Clevers, J. G., & Kooistra, L. (2012). Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 574-583.
- Coltri, P. P., Zullo, J., do Valle Goncalves, R. R., Romani, L. A. S., & Pinto, H. S. (2013). Coffee Crop's Biomass and Carbon Stock Estimation With Usage of High Resolution Satellites Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3), 1786-1795.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46.
- Congalton, R. G., & Green, K. (1999). Assessing the accuracy of remotely sensed data: Principles and practices. Boca Raton, FL: CRC/Lewis Press.
- Coops, N., Stanford, M., Old, K., Dudzinski, M., Culvenor, D., & Stone, C. (2003). Assessment of Dothistroma needle blight of Pinus radiata using airborne hyperspectral imagery. *Phytopathology*, 93(12), 1524-1532.
- Coste, J. (1992). Coffee: The plant and the product. New York: Longman.
- Cressey, D. (2013). Coffee rust regains foothold. *Nature*, 493, 587.
- Cristancho, M., Roza, Y., Escobar, C., Rivillas, C., & Gaitán, A. (2012). Outbreak of coffee leaf rust (*Hemileia vastatrix*) in Colombia. *New Disease Reports*, 25(19), 2044-2052.
- Croft, H., Chen, J. M., Zhang, Y., & Simic, A. (2013). Modelling leaf chlorophyll content in broadleaf and needle leaf canopies from ground, CASI, Landsat TM 5 and MERIS reflectance data. *Remote Sensing of Environment*, 133, 128-140.
- Croome, R. (1989). The potential for satellite remote sensing to monitor coffee, tea, cocoa and coconut plantings in Papua New Guinea Working Paper 2. Series: Designing Monitoring Systems for Small Holders Agriculture in Papua New Guinea: Australian National University.
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3), 271-278.
- Dalponte, M., Bruzzone, L., Vescovo, L., & Gianelle, D. (2009). The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 113(11), 2345-2355.
- DaMatta, F. M. (2004). Exploring drought tolerance in coffee: A physiological approach with some insights on plant breeding. *Braz. J. of Plant Phys.*, 16(1), 256-298.
- DaMatta, F. M., Maestri, M., Barros, R. S., & Regazzi, A. J. (1993). Water relations of coffee leaves (*Coffea arabica* and *C. canephora*) in response to drought. *J. of Hort. Sci.*, 68 741-746.
- Dangwal, N., Patel, N., Kumari, M., & Saha, S. (2016). Monitoring of water stress in wheat using multispectral indices derived from Landsat-TM. *Geocarto International*, 31(6), 682-693.
- Darvishzadeh, R., Matkan, A. A., & Ahangar, A. D. (2012). Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4), 1222-1230.
- Dash, J., & Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25, 5403-5413.
- Daszykowski, M., Serneels, S., Kaczmarek, K., Van Espen, P., Croux, C., & Walczak, B. (2007). TOMCAT: A MATLAB toolbox for multivariate calibration techniques. *Chemometrics and intelligent laboratory systems*, 85(2), 269-277.
- de Carvalho, A. M., da Silva, F. M., Sanches, L., de Carvalho, L. G., & e Silva Ferraz, G. A. (2013). Geospatial analysis of ecological vulnerability of coffee agroecosystems in Brazil. *Applied Geomatics*, 5(2), 87-97.
- de Oliveira, J. G., Alves, P. L. d. C. A., & Vitória, A. P. (2009). Alterations in chlorophyll a fluorescence, pigment concentrations and lipid peroxidation to chilling temperature in coffee seedlings. *Environmental and Experimental Botany*, 67(1), 71-76.

- Deepak, K., Hanumantha, B. T., & Sreenath, H. L. (2012). Viability of coffee leaf rust (*Hemileia vastatrix*) urediniospores stored at different temperatures. *Journal of Biotechnology & Biomaterials*, 2(5), 1-3.
- DeJonge, K. C., Mefford, n. S., & Chávez, J. (2016). Assessing corn water stress using spectral reflectance. *International Journal of Remote Sensing*, 37(10), 2294-2312.
- Delalieux, S., Somers, B., Verstraeten, W., Van Aardt, J., Keulemans, W., & Coppin, P. (2009). Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. *International Journal of Remote Sensing*, 30(8), 1887-1912.
- Demšar, U., Harris, P., Brunson, C., Fotheringham, A. S., & McLoone, S. (2013). Principal component analysis on spatial data: an overview. *Annals of the Association of American Geographers*, 103(1), 106-128.
- Devi, G. S., & Kumar, K. A. (2008). Remote sensing and GIS application for land quality assessment for coffee growing areas of Karnataka. *Journal of the Indian Society of Remote Sensing*, 36(1), 89-97.
- Dian, Y., Fang, S., Le, Y., Xu, Y., & Yao, C. (2014). Comparison of the different classifiers in vegetation species discrimination using hyperspectral reflectance data. *Journal of the Indian Society of Remote Sensing*, 42(1), 61-72.
- Dias, P. C., Araujo, W. L., Moraes, G. A. B. K., Barros, R. S., & DaMatta, F. (2007). Morphological and physiological responses of two coffee progenies to soil water availability. *J. of Plant Physiol.*, 164, 1639-1647.
- Dinesh, K. P., Shivanna, P., & Santa Ram, A. (2011). Identification of RAPD (Random Amplified Polymorphic DNA) markers for Ethiopian wild coffee Arabica L genetic resources in the tropics. *Research: Plant Genomics*, 2(11), 1-7.
- Ding, Y., Zhao, K., Zheng, X., & Jiang, T. (2014). Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 30, 139-145.
- Dixon, J., & Garrity, D. (2014). Perennial Crops and Trees. *Perennial Crops for Food Security*, 307.
- Dong, J., Xiao, X., Sheldon, S., Biradar, C., Zhang, G., Duong, N. D., Hazarika, M., Wikantika, K., Takeuchi, W., & Moore III, B. (2014). A 50m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment. *PLoS One*, 9, e85801.
- Dos Santos, J. A., Gosselin, P.-H., Philipp-Foliguet, S., Torres, R. d. S., & Falao, A. X. (2012). Multiscale classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10), 3764-3775.
- 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. H., Gessler, P. E., Smith, A. M. S., & Robberecht, R. (2006). Suitability of existing and novel spectral indices to remotely detect water stress in *Populus* spp. *Forest Ecology and Management*, 229(1 - 3), 170- 182.
- Eitel, J. U. H., 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.
- El-Shikha, D. M., Barnes, E. M., Clarke, T. R., Hunsaker, D. J., Haberland, J. A., Pinter Jr, P., Waller, P. M., & Thompson, T. L. (2008). Remote sensing of cotton nitrogen status using the canopy chlorophyll content index (CCCI). *Transactions of the ASAE (American Society of Agricultural Engineers)*, 51(1), 73.
- Farley, K. A., Jobbagy, E. G., & Jackson, R. B. (2005). Effects of afforestation on water yield: a global synthesis with implications for policy. *Global Change Biology*, 11, 1565-1576.

- Fassnacht, F. E., Li, L., & Fritz, A. (2015). Mapping degraded grassland on the Eastern Tibetan Plateau with multi-temporal Landsat 8 data — where do the severely degraded areas occur? *International Journal of Applied Earth Observation and Geoinformation*, 42, 115-127.
- Flood, J. (2009). Coffee wilt disease. Cambridge: CABI.
- Fodor, I. K. (2002). A survey of dimension reduction techniques: Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201.
- Foody, G. M. (2004). Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, 70(5), 627-633.
- Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83-92.
- Franklin, S. E., Hall, R. J., Smith, L., & Gerylo, G. R. (2003). Discrimination of conifer height, age and crown closure classes using Landsat-5 TM imagery in the Canadian Northwest Territories. *International Journal of Remote Sensing*, 24, 1823-1834.
- Freeman, K. W., Girma, K., Arnall, D. B., Mullen, R. W., Martin, K. L., Teal, R. K., & Raun, W. R. (2007). By-plant prediction of corn forage biomass and nitrogen uptake at various growth stages using remote sensing and plant height. *Agronomy Journal*, 99(2), 530-536.
- Funk, C., & Budde, M. E. (2009). Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, 113, 115-125.
- Furfaro, R., Ganapol, B. D., Johnson, L. F., & Herwitz, S. (2005). Model-based neural network algorithm for coffee ripeness prediction using Helios UAV aerial images. Paper presented at the SPIE.
- Gay, C., Estrada, F., Conde, C., Eakin, H., & Villers, L. (2006). Potential impacts of climate change on agriculture: A case study of coffee production in Veracruz, Mexico. *Climate Change*, 79, 259-288.
- Ge, Y., Thompspon, J. A., & Sui, R. (2011). Remote sensing of soil properties in precision agriculture: a review. *Frontiers in Earth Science*, 5(3), 229-238.
- Ghini, R., Bettiol, W., & Hamada, E. (2011). Diseases in tropical and plantation crops as affected by climate change: current knowledge and perspectives. *Plant Pathology*, 60, 122-132.
- Ghiyamat, A., Shafri, H. Z. M., Mahdiraji, G. A., Ashurov, R., Shariff, A. R. M., & Mansour, S. (2015). Airborne hyperspectral discrimination of tree species with different ages using discrete wavelet transform. *International Journal of Remote Sensing*, 36(1), 318-342.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2004). Random forest classification of multisource remote sensing and geographic data. Paper presented at the Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58, 289–298.
- Gitelson, A. A., & Merzlyak, M. N. (1994a). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, 22(3), 247-252.
- Gitelson, A. A., & Merzlyak, M. N. (1994b). Quantitative estimation of chlorophyll using reflectance spectra. *Journal of Photochemistry and Photobiology B: Biology*, 22, 247-252.
- Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32, L08403.
- Glenn, E. P., Huete, A. R., Nagler, P. L., & Nelson, S. G. (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors*, 8(4), 2136-2160.
- Gomez, C., Mangeas, M., Petit, M., Corbane, C., Hamon, P., Hamon, S., De Kochko, A., Le Pierres, D., Poncet, V., & Despinoy, M. (2010). Use of high-resolution satellite imagery in an integrated

- model to predict the distribution of shade coffee tree hybrid zones. *Remote Sensing of Environment*, 114, 2731-2744.
- Gordon, L., Steffen, W., Jonsson, B., Folke, C., Falkenmark, M., & Johannessen, A. (2005). Human modification of global water vapor flows from the land surface. *PNAS*, 102(21), 7612-7617.
- Gouveia, M. M. C., Ribeiro, A., Várzea, V. M. P., & Rodrigues, C. J. (2005). Genetic diversity in *Hemileia vastatrix* based on RAPD markers. *Mycologia*, 97(2), 396-404.
- Gutierrez, P. A., Lopez-Granados, F., Jurado-Exposito, J. M. P. M., & Hervas-Martinez, C. (2008). Logistic regression product-unit neural networks for mapping *Ridolfia segetum* infestations in sunflower crop using multitemporal remote sensed data. *Computers and Electronics in Agriculture*, 64, 293-306.
- Haddad, F., Maffia, L. A., Mizubuti, E. S., & Teixeira, H. (2009). Biological control of coffee rust by antagonistic bacteria under field conditions in Brazil. *Biological Control*, 49(2), 114-119.
- Hagar, J., & Schepp, K. (2011). Coffee and Climate Change Impacts of on four pilot countries (pp. 19): University of Greenwich.
- Haggar, J., & Schepp, K. (2011). Coffee and Climate Change Impacts of on four pilot countries (pp. 19): University of Greenwich.
- Hailu, B. T., Maeda, E. E., Pellikka, P., & Pfeifer, M. (2015). Identifying potential areas of understorey coffee in Ethiopia's highlands using predictive modelling. *International Journal of Remote Sensing*, 36(11), 2898-2919.
- Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*(122), 66-74.
- Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of *Spartina alterniflora* canopies. *Photogrammetric Engineering and Remote Sensing*, 49, 77 - 83.
- Hatfield, J. L., Gitelson, A. A., Schepers, J. S., & Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, Special Edition, 117-131.
- Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., & Bonfil, D. J. (2011). LAI assessment of wheat and potato crops by VEN $\mu$ S and Sentinel-2 bands. *Remote Sensing of Environment*, 115, 2141-2151.
- Herwitz, S. R., Johnson, L., Arvesen, J., Higgins, R., Leung, J., & Dunagan, n. E. (2002). *Precision agriculture as a commercial application for solar powered unmanned aerial vehicles*. Paper presented at the 1st AIAA UAV Conference, Portsmouth, VA.
- Herwitz, S. R., Johnson, L. F., Dunagan, S. E., Higgins, R. G., Sullivan, D. V., Zheng, J., Lobitz, B. M., Leung, J. G., Gallmeyer, B. A., Aoyagi, M., Slye, R. E., & Brass, J. A. (2004). Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support. *Computers and Electronics in Agriculture*, 44, 49-61.
- Hess, T., Stephens, W., Weatherhead, K., Knox, J., & Kay, M. (1998). Management of irrigation for tea and coffee: International Centre for Plantation Studies, Cranfield University.
- Hillocks, R. J., Phiri, N. A., & Overfield, D. (1999). Coffee pest and disease management options for smallholders in Malawi. *Crop Protection*, 18(3), 199-206.
- Houborg, R., McCabe, M. F., Cescatti, A., & Gitelson, A. A. (2015). Leaf chlorophyll constraint on model simulated gross primary productivity in agricultural systems. *International Journal of Applied Earth Observation and Geoinformation*, 43, 160-176.
- Huang, W., Lamb, D., 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(4-5), 187-197.
- Hunt, E. R., Daughtry, C., Eitel, J. U., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy journal*, 103(4), 1090-1099.
- ICO. (2009). *Climate change and coffee*. Paper presented at the International Coffee Council, London.
- ICO. (2015). Sustainability of the coffee sector in Africa. London: International Coffee Organisation.
- ITT. (2008). ENVI 4.7: ITT Visual Solutions.

- Jackson, D., Skillman, J., & Vandermeer, J. (2012). Indirect biological control of the coffee leaf rust, *Hemileia vastatrix*, by the entomogenous fungus *Lecanicillium lecanii* in a complex coffee agroecosystem. *Biological Control*, 61(1), 89-97.
- Jayathilaka, P. M. S., Soni, P., Perret, S., Jayasuriya, H. P. W., & Salokhe, V. (2012). Spatial assessment of climate change effects on crop suitability for major plantation crops in Sri Lanka. *Regional Environmental Change*, 12(1), 55-68.
- Jeger, M. J., & Pautasso, M. (2008). Plant disease and global change—the importance of long-term data sets. *New Phytologist*, 177(1), 8-11.
- Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., & Li, B. (2014). Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto International*, 29(8), 941-951.
- Jia, R., Mao, Z., & Chang, Y. (2010). A nonlinear robust partial least squares method with application. Paper presented at the 2010 Chinese Control and Decision Conference.
- Jiang, J., Hu, R., Han, Z., Wang, Z., & Chen, J. (2013). Two-step superresolution approach for surveillance face image through radial basis function-partial least squares regression and locality-induced sparse representation. *Journal of Electronic Imaging*, 22(4), 041120-041120.
- Johl, R., Henke, S., & Schweikart, J. (2014). Acquiring geodata for coffee mapping using remote sensing data based on a pilot study in the Mbinga district Tanzania. *Zbl. Geol. Paläont. Teil*, 1, 211–225.
- Johnson, L., Herwitz, S., Lobitz, B., & Dunagan, S. (2004). Feasibility of monitoring coffee field ripeness with airborne multispectral imagery. *Applied Engineering in Agriculture*, 20(6), 845.
- Junior, J. Z., Coltri, P. P., do Valle Gonçalves, R. R., & Romani, L. A. S. (2015). Multi-resolution in remote sensing for agricultural monitoring: A review. *Revista Brasileira de Cartografia*, 3(66/7).
- Kahn, P. C., Molnar, T., Zhang, G. G., & Funk, C. R. (2011). Investing in perennial crops to sustainably feed the world. *Issues in Science and Technology*, 27(4), 75.
- Kalacska, M., Lalonde, M., & Moore, T. R. (2015). Estimation of foliar chlorophyll and nitrogen content in an ombrotrophic bog from hyperspectral data: Scaling from leaf to image. *Remote Sensing of Environment*, 169, 270-279.
- Kamiran, N., & Sarker, M. L. R. (2014). *Exploring the potential of high resolution remote sensing data for mapping vegetation and the age groups of oil palm plantation* Paper presented at the 8th International Symposium of the Digital Earth (ISDE8).
- Kanke, Y., Tuban, B., Dalen, M., & Harrell, D. (2016). Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precision Agriculture*, 17.
- Ke, Y., Im, J., Lee, J., Gong, H., & Ryu, Y. (2015). Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. *Remote Sensing of Environment*, 164, 298-213.
- Kilambo, D. L., Reuben, S. O. W. M., & Mamiro, D. P. (2013). Responses of compact coffee clones against coffee berry and coffee leaf rust diseases in Tanzania. *Journal of Plant Studies*, 2(2), 81-94.
- Knyazikhin, Y., Schull, M. A., Stenberg, P., Mottus, M., Rautiainen, M., Yang, Y., Marshak, A., Latorre Carmona, P., Kaufmann, R. K., & Lewis, P. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Science*, 110, E185-E192.
- Ko, J., Maas, S. J., Mauget, S., Piccinni, G., & Wanjura, D. (2006). Modeling water-stressed cotton growth using within-season remote sensing data. *Agronomy Journal*, 98, 1600-1609.
- Kruse, J. K., Christians, N. E., & Chaplin, M. H. (2006). Remote sensing of nitrogen stress in creeping bentgrass. *Agronomy Journal*, 98, 1640-1645.
- Kumar, L. (2007). High-spectral resolution data for determining leaf water content in Eucalyptus species: leaf level experiments. *Geocarto International*, 22(1), 3-16.
- Kuri, F., Murwira, A., Murwira, K. S., & Masocha, M. (2014). Predicting maize yield in Zimbabwe using dry dekads derived from remotely sensed Vegetation Condition Index. *International Journal of Applied Earth Observation and Geoinformation*, 33, 39-46.

- Kutywayo, D. (2002a). The coffee white stem borer problem in Zimbabwe. Paper presented at the Launching function and workshop of the ICO-CFC white stem borer project, Karnataka, India.
- Kutywayo, D. (2002b). *Observations into the life history of the coffee white stem borer (Monochamus leuconotus)* Paper presented at the 19th Colloquium of the Association for Science and Information on Coffee (ASIC), Rome, Italy.
- Kutywayo, D., Chemura, A., Kusena, W., Chidoko, P., & Mahoya, C. (2013). The Impact of Climate Change on the Potential Distribution of Agricultural Pests: The Case of the Coffee White Stem Borer (*Monochamus leuconotus* P.) in Zimbabwe. *Plos One*, 8(8), e73432.
- Kutywayo, D., Chingwara, V., Mahoya, C., Chemura, A., & Masaka, J. (2010). The effect of different levels of irrigation water and nitrogen fertilizer on vegetative growth components and yield of coffee in Zimbabwe. *MSU J. Sci. & Tech*, 2(2), 45-54.
- Laderach, P., Jarvis, A., & Ramirez, J. (2006). The impact of climate change in coffee-growing regions: The case of 10 municipalities in Nicaragua ADAPCC: CafeDirect/GTZ.
- Lagerblad, L. (2010). Assessment of environmental flow requirements in Buzi River basin, Mozambique. (PhD), Umeå University, Umeå.
- Landmann, T., & Dubovyk, O. (2014). Spatial analysis of human-induced vegetation productivity decline over eastern Africa using a decade (2001–2011) of medium resolution MODIS time-series data. *International Journal of Applied Earth Observation and Geoinformation*, 33, 76-82.
- Lanfredi, M., Coppola, R., Simoniello, T., Coluzzi, R., Imbrenda, V., & Macchiato, M. (2015). Early identification of land degradation hotspots in complex Bio-Geographic Regions. *Remote Sensing*, 7(6), 8154-8179.
- Lashermes, P., & Anthony, F. (2007). Coffee In C. Kole (Ed.), *Technical Crops: Genome mapping and molecular breeding in plants* (pp. 109 -118). Heidelberg, Berlin: Springer Berlin.
- Lawlor, D. W. (1995). Photosynthesis, productivity and environment. *Journal of Experimental Botany*, 46, 1449-1461.
- Lebedev, A., Westman, E., Van Westen, G., Kramberger, M., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., & Tsolaki, M. (2014). Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, 6, 115-125.
- Lee, Y., Yang, C., Chang, K., & Shen, Y. (2008). A simple spectral index using reflectance of 735 nm to assess nitrogen status of rice canopy. *Agronomy Journal*, 100, 202-212.
- Lepine, L. C., Ollinger, S. V., Ouimette, A. P., & Martin, M. E. (2016). Examining spectral reflectance features related to foliar nitrogen in forests: Implications for broad-scale nitrogen mapping. *Remote Sensing of Environment*, 173, 174-176.
- Li, H., Lee, W., Wang, K., Ehsani, R., & Yang, C. (2014). Extended spectral angle mapping (ESAM)' for citrus greening disease detection using airborne hyperspectral imaging. *Precision Agriculture*, 15(2), 162-183.
- Li, M., Jungho, L. M., & Beier, C. (2013). Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. *GIScience & Remote Sensing*, 50(4), 361-384.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Liaw, A., Wiener, M., Breiman, L., & Cutler, A. (2009). Package "randomForest".
- Lillesand, T. M., & Kiefer, R. W. (2008). *Remote Sensing and Image Interpretation* (5th ed.). New York: Wiley & Sons.
- Lin, B. B. (2011). Resilience in agriculture through crop diversification: adaptive management for environmental change. *BioScience*, 61(3), 183-193.
- Liu, C., Frazier, P., & Kumar, L. (2007). Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107, 606-616.
- Liu, L., Zhao, C., Huang, W., & Wang, J. (2003). Estimating winter wheat plant water content using red edge width. Paper presented at the Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International.



- Logan, W. J. C., & Biscoe, J. (1987). *Coffee Handbook*. Harare: Zimbabwe Coffee Growers' Association.
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28, 823-870.
- Mahlein, A.-K., Oerke, E.-C., Steiner, U., & Dehne, H.-W. (2012a). Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, 133(1), 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.
- Mahlein, A.-K., Steiner, U., Hillnhütter, C., Dehne, H.-W., & Oerke, E.-C. (2012b). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant methods*, 8(1), 3.
- Malenovský, Z., Mishra, K. B., Zemek, F., Rascher, U., & Nedbal, L. (2009). Scientific and technical challenges in remote sensing of plant canopy reflectance and fluorescence. *Journal of Experimental Botany*, Special Edition, 156-169.
- Marcon, M., Mariano, K., Braga, R. A., Paglis, C. M., Scalco, M. S., & Horgan, G. W. (2011). Estimation of total leaf area in perennial plants using image analysis. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 15(1), 96-101.
- Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., & Goulart, L. R. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, 35(1), 1-25.
- Masarirambi, M. T., Chingwara, V., & Shongwe, V. D. (2009). The effect of irrigation on synchronization of coffee (*Coffea arabica* L.) flowering and berry ripening at Chipinge, Zimbabwe. *Physics and Chemistry of the Earth*, 34(13-16), 786-789.
- Maxwell, S. K., Nuckols, J. R., Ward, M. H., & Hoffer, R. M. (2004). An automated approach to mapping corn from Landsat imagery. *Computers and Electronics in Agriculture*, 43(1), 43-54.
- McKight, P. E., & Najab, J. (2010). Kruskal-Wallis Test. *Corsini Encyclopedia of Psychology*.
- McMorrow, J. (2001). Linear regression modelling for the estimation of oil palm age from Landsat TM. *International Journal of Remote Sensing*, 22(12), 2243-2264.
- Moharana, S., & Dutta, S. (2016). Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 17-29.
- Moreira, M. A., Adami, M., & Rudorff, B. F. T. (2004). Spectral and temporal behavior analysis of coffee crop in Landsat images. *Pesq. agropec. bras., Brasília*, 39(3), 223-231.
- Moreira, M. A., Rurdorf, B. F. T., Barros, M. A., De Faria, V. G. C., & Adami, M. (2010). Geotechnologies to map coffee fields in the states of Minas and Sao Paulo. *Eng. Agríc.*, 30(6), 1123-1135.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.
- 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(3), 173-188.
- Mtenga, D. J., Reuben, S. O. W. M., Varzea, V. M. P., Pereira, P., Kilambo, D. L., & Teri, J. M. (2008, 14-19 September 2008). Variation in resistance to coffee berry disease (*Colletotrichum kahawae*) among germplasm progenitors in Tanzania. Paper presented at the 22nd International Conference on Coffee Science, Campinas, Brazil.
- Mukashema, A., Veldkamp, A., & Vrieling, A. (2014). Automated high resolution mapping of coffee in Rwanda using an expert Bayesian network. *International Journal of Applied Earth Observation and Geoinformation*, 33, 331-340.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358-371.

- Muller-Wilm, U., Louis, J., Richter, R., Gascon, F., & Niezette, M. (2013). Sentinel-2 Level 2A prototype processor: Architecture, algorithms and first results: ESA Special Publication.
- Muller, R. A., Berry, D., Avelina, J., & Bieysse, D. (2004). Coffee diseases. In J. N. Wintgens (Ed.), *Coffee: growing, processing & sustainable production: A guidebook for growers, processors, traders, and researchers* (pp. 491-545). Weinheim: Wiley-VCH
- Murphy, T. S., Phiri, N. A., Sreedharan, K., Kutwayo, D., & Chanika, C. (2008). Integrated stem borer management in smallholder coffee farms in India, Malawi and Zimbabwe: Final Technical Report (pp. 220): CFC/CABI/ICO/CRS/CB.
- Musoli, P. C., Kangire, A., Leroy, T., Nabaggala, A., Nakendo, S., Olal, S., Ochugo, J., Kabole, C., Pande, J., Cilas, C., Charrier, A., Bieysse, D., Ogwang, J. A., & Kyetere, D. T. (2008, 14 -19 September 2008). Towards a variety resistant to coffee wilt disease (CWD): a case for robusta coffee (*Coffea canephora*) in Uganda. Paper presented at the 22nd International Conference on Coffee Science, Campinas, Brazil.
- Mutanga, O., Adam, E., & Cho, M. A. (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.
- Mutanga, O., & Ismail, R. (2010). Variation in foliar water content and hyperspectral reflectance of *Pinus patula* trees infested by *Sirex noctilio*. *Southern Forests: A Journal of Forest Science*, 72(1), 1-7.
- Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19), 3999-4014.
- Mutanga, O., & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), 34-42.
- N'Doume, C. N. P., Lachenaud, A., Hard, H. N. V., & Flori, A. (2000). Feasibility study for the statistical inventory mapping of coffee and cocoa plantations in Ivory Coast using satellite remote sensing. *Bulletin - Société française de photogrammétrie et de télédétection*, 157, 3-10.
- Nair, K. P. P. (2010). *The Agronomy and Economy of Important Tree Crops of the Developing World*: Elsevier.
- Nelson, S. C. (2008). Cercospora leaf spot and berry blotch of coffee: University of Hawai'i at Manoa, College of Tropical Agriculture and Human Resources, Cooperative Extension Service.
- Netto, A. T., Campostrini, E., de Oliveira, J. G., & Bressan-Smith, R. E. (2005). Photosynthetic pigments, nitrogen, chlorophyll a fluorescence and SPAD-502 readings in coffee leaves. *Scientia Horticulturae*, 104(2), 199-209.
- Nicolin, S. (2011). Implementing environmental water requirements in Buzi River basin, Mozambique: An impact analysis based on the Water Resource Yield Model. (MSc), Uppsala University, Sweden.
- Nutter Jr., F. W., Tylka, G. L., Guan, J., Moreira, A. J. D., Marett, C. C., Rosburg, T. R., Basart, J. P., & Chong, C. S. (2002). Use of remote sensing to detect soybean cyst nematode-induced plant stress. *Journal of Nematology*, 34(3), 222-231.
- O'Loughlin, E., & Nambiar, E. K. S. (2001). Plantations, farm forestry and water A discussion paper (Vol. RIRDC Publication No. 01/137, pp. 37). Canberra: RIRDC/LWA/FWPRDC: Joint Venture Agroforestry Program.
- Okalebo, R. J., Gathua, K. W., & Woome, P. L. (2002). Laboratory methods for soil and plant analysis: A working manual (2nd ed.). Nairobi: TSBF-CIAT & SACRED Africa.
- Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, 189, 375-394.
- Ollinger, S. V., Reich, P. B., Frolking, S., Lepine, L. C., Hollinger, D. Y., & Richardson, A. D. (2013). Nitrogen cycling, forest canopy reflectance, and emergent properties of ecosystems. *Proceedings of the National Academy of Science*, 110, E2437-E2437.
- Ollinger, S. V., Richardson, A. D., Martin, M. E., Hollinger, D. Y., Frolking, S. E., Reich, P. B., Plourde, L. C., Katul, G. G., Munger, J. W., & Oren, R. (2008). Canopy nitrogen, carbon assimilation, and

- albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences*, 105, 19336-19341.
- Omont, H., Nicolas, D., & Russell, D. (2006). The future of perennial tree crops: what role for agroforestry. *World Agroforestry into the Future*. World Agroforestry Centre, Nairobi, Kenya.
- Orr, M. J. (1996). Introduction to radial basis function networks: Technical Report, Center for Cognitive Science, University of Edinburgh.
- Ortega-Huerta, M. A., Komar, O., Price, K. P., & Ventura, H. J. (2012). Mapping coffee plantations with Landsat imagery: an example from El Salvador. *International Journal of Remote Sensing*, 33(1), 220-242.
- Oumar, Z., & Mutanga, O. (2014). Predicting water stress induced by *Thaumastocoris peregrinus* infestations in plantation forests using field spectroscopy and neural networks. *Journal of Spatial Science*, 59(1), 79-90.
- Ozdemir, I. (2008). Estimating stem volume by tree crown area and tree shadow area extracted from pan-sharpened Quickbird imagery in open Crimean juniper forests. *International Journal of Remote Sensing*, 29(19), 5643-5655.
- Pahlevan, N., & Schott, J. R. (2013). Leveraging EO-1 to evaluate capability of new generation of Landsat sensors for coastal/inland water studies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2), 360-374.
- Peña, M., & Brenning, A. (2015). Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. *Remote Sensing of Environment*, 171, 234-244.
- Peng, Y., Gitelson, A. A., Keydan, G., Rundquist, D. C., & Moses, W. (2011). Remote estimation of gross primary production in maize and support for a new paradigm based on total crop chlorophyll content. *Remote Sensing of Environment*, 115(4), 978-989.
- Peñuelas, J., Gamon, J., Fredeen, A., Merino, J., & Field, C. (1994). Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sensing of Environment*, 48(2), 135-146.
- Petropoulos, G. P., Koutsoyias, C. C., & Keramitsoglou, I. (2012). Land cover mapping with emphasis to burnt area delineation using co-orbital ALI and Landsat TM imagery. *International Journal of Applied Earth Observation and Geoinformation*, 18, 344-355.
- Phiri, N., & Baker, P. (2009). Coffee Wilt Disease in Africa: A synthesis of the work of the regional coffee wilt programme 2000–2007 (pp. 258). Nairobi: CABI.
- Pompelli, M. F., Martins, S. C. V., Antunes, W. C., Chaves, A. R. M., & DaMatta, F. M. (2010). Photosynthesis and photoprotection in coffee leaves is affected by nitrogen and light availabilities in winter conditions. *Journal of Plant Physiology*, 167, 1052-1060.
- Pontius, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32, 4407-4429.
- Pontius, R. G., Shusas, E., & Menzie, M. (2004). Detecting important categorical land changes while accounting for persistence. *Agriculture, Ecosystems & Environment*, 10(2-3), 251-268.
- Prabhakar, M., Prasad, Y. G., Desai, S., Thirupathi, M., Gopika, K., Rao, G. R., & Venkateswarlu, B. (2013). Hyperspectral remote sensing of yellow mosaic severity and associated pigment losses in *Vigna mungo* using multinomial logistic regression models. *Crop Protection*, 45, 132-140.
- Price, T., Gross, R., Wey, J., & Osborne, C. (1993). A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (*Hemileia vastatrix*). *Australian Journal of Experimental Agriculture*, 33, 97–101.
- Prosser, I. P., & Walker, G. R. (2009). A review of plantations as a water intercepting land use in South Australia: CSIRO: Water for a Healthy Country National Research Flagship.
- R Core Team. (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rabe, A., Jakimow, B., Held, M., van der Linden, S., & Hostert, P. (2014). EnMAP-Box, Version 2.0. software available at [www.enmap.org](http://www.enmap.org).

- Rahn, E., Läderach, P., Baca, M., Cressy, C., Schroth, G., Malin, D., van Rikxoort, H., & Shriver, J. (2013). Climate change adaptation, mitigation and livelihood benefits in coffee production: where are the synergies? *Mitigation and Adaptation Strategy for Global Change*, 1-19.
- Ramirez, G. M. R., Jurandir, Z. J., Assad, E. D., & Pinto, H. S. (2006). Comparison between Ikonos-II and Landsat/ETM+ satellites data in the study of coffee areas. *Pesq. agropec. bras.*, 41(4), 661-666.
- Ramoelo, A., Cho, M., Mathieu, R., & Skidmore, A. K. (2015). Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of Applied Remote Sensing*, 9(4), 09409611-094096112.
- Reis, A. R., Favarin, J. L., Malavolta, E., Junior, J. L., & Moraes, M. F. (2009). Photosynthesis, chlorophylls, and SPAD readings in coffee leaves in relation to nitrogen supply. *Communications in soil science and plant analysis*, 40(9-10), 1512-1528.
- Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, 5, 1704-1733.
- Reschke, J., & Hüttig, C. (2014). Continuous field mapping of Mediterranean wetlands using sub-pixel spectral signatures and multi-temporal Landsat data. *International Journal of Applied Earth Observation and Geoinformation*, 28 220-229.
- Ribeyre, F., & Avelina, J. (2011). Impact of field pests and diseases on coffee quality. In T. Oberthur, P. Laderach, H. A. J. Pohlen & J. H. Cook (Eds.), *Speciality Coffee: Managing Quality* (pp. 151-176). New-York: SCAA.
- Riedell, W., Osborne, S., Hesler, L., Langham, M., & Robert, P. (2003). Remote sensing of barley yellow dwarf and wheat streak mosaic disease in winter wheat canopies. Paper presented at the Proceedings of the 6th International Conference on Precision Agriculture and Other Precision Resources Management, Minneapolis, MN, USA, 14-17 July, 2002.
- Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Rosipal, R. (2010). Nonlinear partial least squares: An overview. *Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques*, 169-189.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. Paper presented at the Third Earth Resources Technology Satellite-1 Symposium, Greenbelt, Maryland, USA.
- Roy, D. P., Wulder, M., Loveland, T., Woodcock, C., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., & Kennedy, R. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172.
- Rulinda, C. M., Dilo, A., Bijker, W., & Stein, A. (2012). Characterising and quantifying vegetative drought in East Africa using fuzzy modelling and NDVI data. *Journal of Arid Environments*, 78, 169-178.
- 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(1), 91-99.
- Rutherford, M. A., & Phiri, N. (2006). *Pests and Diseases of Coffee in Eastern Africa: A Technical and Advisory Manual*. Ascot: CAB International.
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), 2507-2517.
- Safaa, A., & Pontius, R. G. (2012). Intensity analysis to unify measurements of size and stationarity of land changes by interval, category, and transition. *Landscape and Urban Planning*, 106, 103-114.
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, N. R., Pavek, M. J., & Glover, J. (2015). Low-

- altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, 70, 112-123.
- 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.
- Savitzky, A., & Golay, M. (1964). Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry*, 36(8), 1627-1638.
- Scharf, P. C., Schmidt, J. P., Kitchen, N. R., Sudduth, K. A., Hong, S. Y., Lory, J. A., & Davis, J. G. (2002). Remote sensing for nitrogen management. *Journal of Soil and Water Conservation*, 57(6), 518-524.
- Schlemmer, M., Gitelson, A., Schepers, J., Ferguson, R., Peng, Y., Shanahan, J., & Rundquist, D. (2013a). Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation*, 25, 47-54.
- Schlemmer, M., Gitelson, A. A., Schepers, J. S., Ferguson, R. B., Peng, Y., Shanahan, J., & Rundquist, D. (2013b). Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation*, 25, 47-54.
- Schroth, G., Laderach, P., Dempewolf, J., Philpott, S., Hagggar, J., Eakin, H., Castillejos, T., Garcia Moreno, J., Soto-Pinto, L., Hernandez, R., Eitzinger, A., & Ramirez-Villegas, J. (2009). Towards a climate change adaptation strategy for coffee communities and ecosystems in Sierra Madre de Chiappas, Mexico. *Mitigation and Adaptation to Global Change*, 14, 605-625.
- Sembiiring, H., Lees, H., Raun, W., Johnson, G., Solie, J., Stone, M., DeLeon, M., Lukina, E., Cossey, D., & LaRuffa, J. (2000). Effect of growth stage and variety on spectral radiance in winter wheat. *Journal of plant nutrition*, 23(1), 141-149.
- Shoeman, P. S., van Hamburg, H., & Pasques, B. P. (1998). The morphology and phenology of the coffee white stem borer, *Monochamus leuconotus* (Pascoe) (Coleoptera: Cerambycidae), a pest of Arabica coffee. *African Entomology*, 6(1), 83-89.
- Sibanda, M., Mutanga, O., & Rouget, M. (2016). Comparing the spectral settings of the new generation broad and narrow band sensors in estimating biomass of native grasses grown under different management practices. *GIScience & Remote Sensing*, 53(5), 614-633.
- Silva, S. d. A., Queiroz, D. M. d., Pinto, F. d. A. C., & Santos, N. T. (2014). Coffee quality and its relationship with Brix degree and colorimetric information of coffee cherries. *Precision Agriculture*, 15, 543-554.
- Southworth, J., Nagendra, H., & Tucker, C. (2002). Fragmentation of a landscape: Incorporating landscape metrics into satellite analyses of land-cover change. *Landscape Research*, 27(3), 253-269.
- Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, 50(9), 1250-1265.
- Stibig, H. J., Belward, A., Roy, P., Rosalina-Wasrin, U., Agrawal, S., Joshi, P., Beuchle, R., Fritz, S., Mubareka, S., & Giri, C. (2007). A land-cover map for South and Southeast Asia derived from SPOT-VEGETATION data. *Journal of Biogeography*, 34(4), 625-637.
- Stone, C., Chisholm, L., & Coops, N. (2001). Spectral reflectance characteristics of eucalypt foliage damaged by insects. *Australian Journal of Botany*, 49(6), 687-698.
- Suratman, M. N., Bull, G. Q., Leckie, D. G., Lemay, V. M., Marshall, P. L., & Mispan, M. R. (2004). Prediction models for estimating the area, volume, and age of rubber (*Hevea brasiliensis*) plantations in Malaysia using Landsat TM data. *International Forestry Review*, 6(1), 1-12.
- Tan, K. P. (2013). Use of the UK-DMC2 for identifying age of oil palm trees in Southern Peninsular Malaysia. *International Journal of Remote Sensing*, 34:20(20), 7424-7446.
- Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171-179.

- Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004a). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91(3), 354-376.
- Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., Van der Meer, B., & Enclona, E. (2004b). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing*, 25(23), 5447-5472.
- Torbick, N., & Salas, W. (2014). Mapping agricultural wetlands in the Sacramento Valley, USA with satellite remote sensing. *Wetlands Ecology & Management*, 23(1), 79-94.
- Trabaquini, K., Miglioranza, É., França, V. d., & Pereira Neto, O. C. (2010). The use of geotechnology to characterize coffee crops in Londrina-PR, for altimetry, slope and soil type. *Engenharia Agrícola*, 30(6), 1136-1147.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150.
- Tully, K. L., Lawrence, D., & Scanlon, T. M. (2012). More trees less loss: Nitrogen leaching losses decrease with increasing biomass in coffee agroforests. *Agriculture, ecosystems & environment*, 161, 137-144.
- Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, 113, S67-S77.
- Van der Vossen, H. A. M. (2001). Agronomy I: Coffee Breeding Practices. In R. J. Clarke & O. G. Vitzthum (Eds.), *COFFEE: Recent Developments World Agricultural Series* (pp. 184-201). London: Blackwell Science.
- Van der Vossen, H. A. M., & Walyaro, D. J. (2009). Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in arabica coffee (*Coffea arabica* L.). *Euphytica*, 165, 105-101 111.
- Vancutsem, C., Pekel, J. F., Evrard, C., Malaisse, F., & Defourny, P. (2009). Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series. *International Journal of Applied Earth Observation and Geoinformation*, 11, 62-76.
- Vandermeer, J., Jackson, D., & Perfecto, I. (2014). Qualitative dynamics of the coffee rust epidemic: educating intuition with theoretical ecology. *BioScience*, 124, 034.
- Venteris, E. R., Tagestad, J., Downs, J., & Murray, C. (2015). Detection of anomalous crop condition and soil variability mapping using a 26 year Landsat record and the Palmer crop moisture index. *International Journal of Applied Earth Observation and Geoinformation*, 39, 160-170.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, 118, 127-139.
- Vieira, T. G. C., Alves, H. M. R., Lacerda, M. P. C., Veiga, R. D., & Epiphanyo, J. C. N. (2006). Crop parameters and spectral response of coffee (*Coffea arabica* L.) areas within the state of Minas Gerais, Brazil. *Coffee Science*, 1(2), 111-118.
- Vieira, T. G. C., Alves, H. M. R., Lacerda, M. P. C., Veiga, R. D., & Epiphanyo, J. C. N. (2012). *Crop parameters for the evaluation of the spectral response of coffee (Coffea arabica) areas within the state of Minas Gerais, Brazil*. Paper presented at the XXVII ISPRS Symposium
- Vincini, M., Amaducci, S., & Frazzi, E. (2014). Empirical estimation of leaf chlorophyll density in winter wheat canopies using Sentinel-2 spectral resolution. *IEEE Transactions of GeoScience & Remote Sensing*, 52(6), 3220-3235.
- Vincini, M., Calegari, F., & Casa, R. (2015). Sensitivity of leaf chlorophyll empirical estimators obtained at Sentinel-2 spectral resolution for different canopy structures. *Precision Agriculture*, 16, 1-19.
- Walawender, J. P., Hajto, M. J., & Iwaniuk, P. (2012). A new ArcGIS toolset for automated mapping of land surface temperature with the use of LANDSAT satellite data. Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.

- Waller, J. M., Bigger, M., & Hillocks, R. J. (2007). Coffee pests, diseases and their management: CABI.
- Wang, C., Fritschi, F. B., Stacey, G., & Yang, Z. W. (2011). Phenology-based assessment of perennial energy crops in North American tallgrass prairie. *Annals of the Association of American Geographers*, 101(4), 742-751.
- Wang, F., Huang, J., Xu, J., & Wang, X. (2008). Wavebands selection for rice information extraction based on spectral bands inter-correlation. *Guang pu xue yu guang pu fen xi= Guang pu*, 28(5), 1098-1101.
- Wang, Q., Tenhunen, J., Dinh, N. Q., Reichstein, M., Vesala, T., & Keronen, P. (2004). Similarities in ground- and satellite-based NDVI time series and their relationship to physiological activity of a scots pine forest in Finland. *Remote Sensing of Environment*, 93(1-2), 225-237.
- Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A. K., Jones, S., Suarez, L., Woodgate, W., Heiden, U., Heurich, M., & Hearne, J. (2016). Vegetation indices for mapping canopy foliar nitrogen in a mixed temperate forest. *Remote Sensing*, 8, 491.
- Waston, K., & Achinelli, M. (2008). Context and contingency: the coffee crisis for conventional small-scale coffee farmers in Brazil. *The Geographical Journal*, 174, 223-234.
- Wrigley, G. (1988). Coffee. London: UK: Longman Scientific and Technical.
- Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148(8), 1230-1241.
- Xiao, X., Boles, S., Liu, J., Zhuang, D., Frohling, S., Li, C., Salas, W., & Moore, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, 95, 480 - 492.
- Xiao, X., Braswell, B., Zhang, Q., Boles, S., Frohling, S., & Moore, B. (2003). Sensitivity of vegetation indices to atmospheric aerosols: Continental-scale observations in Northern Asia. *Remote Sensing of Environment*, 84, 385-392.
- Yan, X., Du, W., & Qian, F. (2004). Development of a kinetic model for industrial oxidation of p-xylene by RBF-PLS and CCA. *AIChE journal*, 50(6), 1169-1176.
- Zarco-Tejada, P. J., Ustin, S. L., & Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agronomy Journal*, 97, 641-653.
- Zeru, A., Assefa, F., Adugna, G., & Hindorf, H. (2008, 14 -19 September 2008). Variation of colletotrichum kahawae isolates from diseased cherries of montane rainforest coffee in Ethiopia. Paper presented at the 22nd International Conference on Coffee Science, Campinas, Brazil.
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, 13, 693-712.
- 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.
- Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., & Koti, S. (2005). Selection of optimum reflectance ratios for estimating leaf nitrogen and chlorophyll concentrations of field-grown cotton. *Agronomy Journal*, 97, 89-98.
- Zhou, Y., Xiao, X., Qina, Y., Donga, J., Zhang, G., Kou, W., Jin, C., Wanga, J., & Li, X. (2016). Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. *International Journal of Applied Earth Observation and Geoinformation*, 46 1-12.
- Zhu, Z., Wand, S., & Woodcock, C. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159, 269-277.
- Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83-94.