Crop type mapping in a highly heterogeneous agriculture landscape. A case of Marble Hall using multi-temporal Landsat 8 and Sentinel 2 imageries

By

Mazarire Theresa Taona (1316540)

Supervisor: Dr Elhadi Adam

Co-supervisor: Dr Nyamugama Adolph

A research report submitted in partial fulfillment of the requirements for the degree in Master of Science in GIS and Remote Sensing

Faculty of Science

University of the Witwatersrand

2019

Johannesburg, South Africa

ABSTRACT

Detailed and accurate studies designed to map crop types play a strategic role in giving information about proper management strategies for precision agriculture, crop monitoring, crop yield estimation, and food security management. The availability of moderate to high resolution remote sensing imageries such as Sentinel-2 in the last few years with less or no cost has enormously increased their use for research and other applications. Together with the latest Landsat mission, Landsat-8, these sensors provide reliable, inexpensive and timely data coupled with a wide range of spatial, temporal and spectral characteristics suitable for mapping the dynamics in agriculture landscape worth exploring. Endeavors to improve crop type maps accuracy have seen the progression of machine learning algorithms to further advance the image classification techniques. Regardless of these efforts, crop classifications generated from these techniques are still deemed as inadequate for agriculture applications, due to the slight inconsistency between the derived classified maps and the information on the ground. In this regard, this research pursued two main objectives: firstly, to compare and explore the utility of Sentinel-2 and Landsat-8 imageries for mapping crop types, secondly, to test the two machine learning algorithms- Random Forest (RF) and Support Vector Machine (SVM) performance and efficiency in classifying different crop types and discriminating them from the co-existing land use classes in a heterogeneous agriculture landscape in Marble Hall, Limpopo, South Africa.

Multi-temporal images from both sensors (from June to September 2017) were processed in order to map the crop types and co-existing land use classes. Fourteen classes in total were classified using both RF and SVM. The optimum time based on the study was July-August as this is the peak time for crop development. The utility and contribution of different bands in each sensor for classification were evaluated using RF mean decrease Gini for variable importance. Vegetation red edge, SWIR, and blue bands contributed the most in classifying Sentinel-2 data whilst the nearinfrared, panchromatic and aerosol bands were mostly utilized in Landsat-8 data. To measure the accuracy of the generated thematic maps, accuracy assessment was undertaken with respective independent validation data sets. The best performance was obtained from August 2017 Sentinel2 data classification achieving Overall Accuracy (OA) of 95.4% and a kappa value of 0.94%. The lowest accuracy was obtained from September 2017 Landsat-8 data classification which obtained an overall accuracy of 91.19 % and 87 % with a kappa coefficient of 0.9 and 0.85 using SVM and RF respectively. The present study found that RF and SVM

classifiers performed similarly though SVM outperformed RF by 1.5%. Overall, it was concluded that Sentinel-2 data performs better than Landsat-8 data in discriminating crop type classes and co-existing land use classes in a heterogeneous landscape because of its high spectral and spatial resolution. SVM was the most efficient and best classifier due to its capability to process a few training samples using limited parameters. Consequently, this research revealed the ability of medium resolution multispectral data in crop type classification using the propagated image classification techniques. Conclusively, the need to employ more advanced image classification techniques and utilize diffusion of radar and optical data for crop type mapping on complex landscapes remain paramount.

Dedication

To my beloved grandmother Gogo Chingombe for her unending love, encouragement, and belief in my hard work and success long before I could even imagine it. You were taken from the world before you could witness my success. You are truly missed.

To my dearest parents Charles and Bridgette Mazarire, Mr. and Mrs. Chingombe, thank you for your love, motivation, and support

To my uncle and sister, the Makores, for your patience, unfailing support and continuous encouragement throughout my studies. Many thanks. I could not have finished my studies without you. Mr. Makore, you are my hero

Acknowledgments

All the honor and glory belong to the Lord Almighty for his grace and blessings all throughout my studies.

A special appreciation goes to my supervisors, Dr. Adam Elhadi and Dr. Nyamugama for your guidance, mentoring and profound insight throughout my research. To the University of Witwatersrand, Department of Geography, Archaeology and Environmental studies lecturers, staff- Dr. Merlo, Ms. Chabalala to mention just a few, I thank you for patiently guiding, supporting, mentoring and providing me with knowledge, skills, and expertise throughout the course of my MSc studies. Thank you for enriching me with expertise that enabled me to become an independent scientist.

I would like to give special gratitude to the Agriculture Research Council for providing me with resources and support to process my data and finish my dissertation.

I appreciate the valuable contribution from my colleagues from the Agriculture Research Council, the geo-informatics team-Siphokazi Gcayi, Reneilwe Maake, Kgaogelo Mogano and Wonga Masiza for your assistance with my data analysis and support throughout my research. To my school colleagues, I am grateful for your encouragement, teaching, and continuous support throughout the sleepless nights when I felt like giving up.

I deeply acknowledge my family and friends for their continued understanding and believing in me. To my sisters Anna and Mai Burondi, I highly appreciate your unconditional love, the scolding, prayers, and constructive advice when I felt like it was all too much to bear. To Mutsai Makore, I thank you for your unfailing support, financial assistance and understanding throughout the course of my studies. The journey was not easy, but you were there always. God Bless.

Declaration 1-Plagiarism

I, **Mazarire Theresa Taona** (Student number: 1316540), hereby declare that:

- The "Crop type mapping in a highly heterogeneous agriculture landscape. A case of Marble Hall using multi-temporal Landsat 8 and Sentinel 2 imageries'' is my own work.
- All the sources quoted in this research have been re-rewritten and acknowledged in the reference list.
- This research report has never been submitted at any other university for examination or any degree.
- This research report does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the research report and in the References section.

Signature: TMazarire Date: 29-05-2019

Mazarire Theresa Taona

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Ha=hectares

L-8= Landsat-8 OLI

MSI= Multispectral Instrument

OLI= Operational Land Imager

RF= Random Forest

S2= Sentinel-2 MSI

SVM= Support Vector Machine

SWIR=Short-wave Infrared

CHAPTER ONE

Introduction

1.1 General Introduction

The world faces a major food security challenge although food production has been rising globally. With the estimated global human population growth reaching between 7.4 and 10.6 billion by 2050, the pressure on food security is an ongoing concern facing the world (Unies, 1999). There are various initiatives for improving food security on a global level namely the Food and Agricultural Organization, the World Food Programme and the World Bank to mention just a few (Löw, 2013). Although the rest of the world has made substantial progress towards alleviating food insecurity, Africa, especially Sub-Saharan Africa remains behind (Blein *et al.*, 2013).

Agriculture being the primary source of food supply to humankind plays a crucial role in emancipating food insecurity in Southern Africa. It remains the backbone of South Africa's economy and it has grown by 22% in the last 5 years while its contribution to the GDP was estimated to be R90 458 million at nominal prices thus 2.2 % in 2017 (Greyling, 2012, (DAFF, 2018DAFF, 2018). The agricultural sector in the country has not been able to meet the demand for the main food items consumed domestically since 2000 (Greyling, 2012, Cai *et al.*, 2017). The ability to expand the agricultural cropland is limited hence enhanced agricultural production is a key factor to the agricultural community. Critically, the ability to meet this grave world concern is by enhancing agriculture production without exerting pressure on the natural environment. To meet the unending future food demands, there is a need for effective agriculture management and planning to understand the crops dynamic and distribution. This involves using accurate, reliable and inclusive agriculture intelligence technology to better manage the agriculture landscape (Löw, 2013).

Remote sensing technology has emerged as an effective and important source of land use information giving an opportunity to develop and produce land cover classifications for both large and small areas of different landscape characteristics. It has been successfully used for crop mapping (Vyas *et al.*, 2005, Conrad *et al.*, 2010, Asgarian *et al.*, 2016, Vuolo *et al.*, 2018). Medium to high-resolution multi-spectral data such as MODIS, RapidEye, Landsat, Sentinel, and SPOT have been used for crop mapping because of their favorable resolutions (Immitzer *et al.*, 2016). However, agriculture landscapes have a complex and highly dynamic land use pattern which is challenging to accurately classify using medium spatial resolution data (Asgarian *et al.*, 2016). The landscape is characterized by fragmented, small parcel sized fields and different crop types within a pixel as well as highly heterogeneous crop cover. In addition, there is dirverse spatial and temporal variations between each crop type within moderately short time intervals (Immitzer *et al.*, 2016). This dynamic character of the agricultural landscape presents a challenge to agriculture land use mapping. The advent of unprecedented satellite-based imaging technology provides multi-temporal data which possess regular revisiting intervals and weather independent acquisition capabilities has made the mapping of crop types dynamics manageable (Li *et al.*, 2014). At the regional scale, high spatial resolution sensors such as SPOT and ASTER have been used to accurately differentiate between agriculture crops in such a landscape (Conrad *et al.*, 2010, Inglada *et al.*, 2015). However, this high-resolution data is inaccessible due to its high cost, unavailability and low temporal resolution. The introduction of frequent temporal revisit sensors such as Sentinel-2 and Landsat 8 give new possibilities to identify and capture the dynamics in highly heterogeneous agriculture landscape in a developing economy such as South Africa.

Crop type mapping was traditionally carried out using field-based survey. However, census and the ground-based survey are not feasible in a large scale agriculture landscape as it is laborious, expensive, prone to errors and time consuming hence the use of timely, less expensive and faster methods of crop mapping has become a necessity (Ouzemou *et al.*, 2018). The remote sensing imagery is important in enhancing broad-scale agronomic management and within field monitoring such as precision agriculture. Precision agriculture being a crop management technique that is created to optimize agriculture production provides the possibility to give effective and feasible solutions at the right time (Zhang *et al*., 2002). The technology provides valuable data used to identify crop types and their corresponding location, the nutrient status of each crop and the field characteristics of large acreages necessary for precision agriculture technique (Asgarian *et al.*, 2016). The implausible advances provide new possibilities for classification of agricultural crop types. This technological advancement gives crop information which enables effective spatiotemporal monitoring of crops as well as managing irrigation water resources for crop health thus enhancing adaption of agriculture for climatic changes (Ouzemou *et al.*, 2018). This will additionally assist in identifying the crop health status, expected crop yield and assessment of actual crop distribution and extent of damage thus saving time and resources.

The existing low-efficiency classification algorithms in categorizing and discriminating crop types has raised the idea of developing other classifiers such as Random Forest and Neural Networks which utilize the crop type spectral behavior and reflectance values (Odenweller and Johnson, 1984). These classification approaches require field-based identification of the crop types to enhance extraction of spectral values from satellite images using remote sensing software. The corresponding spectral-temporal data of the crop types represents a distinctive spectral characteristic of the crop thus enabling accurate crop type mapping in a heterogeneous landscape (Asgarian *et al.* 2016).

1.2 Problem statement

In South Africa, the use of remote sensing technology in agriculture is particularly challenging due to a wide range of crop species and commonly small field sizes causing a high spatial heterogeneity in the agriculture landscape (Cai *et al.*, 2017). Marble hall area in Groblersdal, Limpopo has a diverse large-scale agriculture production unit under pivot irrigation which has successfully grown maize, citrus fruits, potatoes, wheat, and green beans to mention just a few. The agriculture potential of this area is high considering the well-drained rich soils as well as well-established irrigation infrastructure from the Loskop dam under their effective irrigation scheme (Waals, 2013). In addition, Marble hall area is considered as the citrus and table grape producer exporting 58% of its produce to the local market and 42% to other countries, primarily the Middle East and Europe. Due to the study area’s significance and massive input to the country’s agriculture economy, there is a need to accurately map the agriculture landscape in the area. To properly and accurately use satellite technology for mapping crop types in a highly complex agriculture area like Marble hall, there is a need for a frequent revisit, which is expensive.

Different crop types have high variability in phenological stages within fields such as early sprouting, establishment, and maturation (Asgarian *et al.*, 2016), which might lead to the same spectral signature at some point in their development. This might pose a challenge to accurately detect and discriminate between crop types and the co-existing land use. This challenge roots the idea to use multi-temporal data to identify the different crop types as the crop characteristics change as they grow and mature. In the absence of distinct crop border, crop type identification becomes a challenge as there is need to employ expensive high spectral resolution to identify salient features to distinguish one crop from another (Veloso *et al.*, 2017). When neighboring crops and the co-existing land use classes are of the same species and when the crop's temporal signatures are the same, the minute separation becomes saturated at the Landsat scale (Graesser and Ramankutty, 2017). This presents the need to test the temporal data suitable to identify the optimum time to discriminate and categorize the different classes accurately. The task remains daunting when mapping heterogeneous agricultural landscapes where fields are often smaller than individual pixels thus presenting a mixed pixel. There is a need to explore more on the use of the free recently launched satellite, the Sentinel-2 MSI and Landsat-8 OLI to classify such agricultural landscape.

1.3 The aim of the study

- To investigate the performance of Sentinel-2 MSI and Landsat-8 OLI data for mapping crop types and the co-existing land use in a highly heterogeneous agriculture landscape in Marble hall area of Limpopo, South Africa.

1.3.1 Objectives

The specific objectives of this study are:

i. To compare the use of Landsat 8 OLI and Sentinel 2 data in mapping crop types using selected machine learning algorithms. ii. To determine the optimum time to map different crop types using the multi-temporal Sentinel-2 MSI and Landsat 8 OLI data.

iii. To examine the performance of Support Vector Machine and Random Forest machine learning algorithms in classifying heterogeneous agricultural landscape based on Sentinel-2 MSI and Landsat-8 OLI imageries.

CHAPTER TWO

Literature review

2.1 Remote sensing-based agriculture monitoring

Progressively, agriculture has benefited from the several services offered by satellite technology. Recent studies have revealed that remotely sensed images are the solution to food insecurity especially in developing countries and that significant gaps and uncertainties remain unclear to investment decisions and policy-making (Ozdogan *et al.*, 2010). The National Development Plan (NDP) in South Africa sets to stimulate the country's economic growth in multiple sectors including agriculture, hence mapping of croplands and crop types is of greater importance (Cai *et al.*, 2017). To successfully implement the above strategy, agriculture should have accurate crop production statistics data and therefore, SiQ developed the Producer Independent Crop Estimate System in 2006. The Agriculture Census and Baseline Mapping project under the Department of Agriculture and Rural development managed to map cropland and crop types using SPOT 5 satellite imagery which is however costly to obtain.

Although agriculture is a relatively small sector in South Africa, it accounts for approximately 2.2 % of national Gross Domestic Product (GDP) and 7% of formal employment (Greyling, 2012, Cai *et al.*, 2017), thus more information about this sector is of crucial importance. Most studies in South Africa however, have only looked more into the mapping of irrigated agriculture land particularly delineating agriculture areas, neglecting to map and distinguish the crop types within those irrigated lands (Cai *et al.*, 2017).

2.2 Use of multi-temporal and multispectral data for crop mapping

The conventional methods of crop types mapping such as field surveying have become laboriously expensive bringing in the necessity to employ efficient, affordable and synoptic methods such as remote sensing (Ouzemou *et al.,* 2018). The remote sensing technology has been recognized to provide inexpensive and accurate means of collecting agriculture land information required for mapping crop types. In the last few years, mankind has witnessed an extraordinary improvement in space technologies providing exceptional data for small scale land use monitoring. The advancement of the orbit sensors has provided refined spectral and spatial settings associated with shorter temporal resolution (Sothe *et al.*, 2017). These high-resolution earth observation satellites are becoming more freely available at local and national scales for several applications including agriculture. Retrieving crop type details, location information, and their spatial coverage at a specific time provides a robust decision support system valuable to the agriculture community (Ouzemou *et al.*, 2018).

Multispectral and multi-temporal remote sensing data has been widely used in the past few decades for crop identification and crop mapping (Vyas *et al.*, 2005, Conrad *et al.*, 2010, Veloso *et al.*, 2017, Ouzemou *et al.*, 2018) since they are believed to offer high quality and affordable source of data that allows land use dynamics to be detected and assessed at a much finer spatial scale, especially agricultural crops (El Hajj *et al.*, 2009). Multi-temporal data is used to cater for the temporal changes within the agriculture landscape. Vyas *et al.* (2005), found the multi-temporal data analysis approach using multi-date IRS LISS 111 data to be accurate in discriminating and mapping various crops whereas single date data is only good for homogeneous area classification. Immitzer *et al.* (2016), concurs with this notion as their study also recommended using multi-temporal data instead of single date data for further increasing classification results of crop type mapping. Furthermore, Vuolo *et al.* (2018), used Sentinel-2 MSI data to evaluate the impact of multi-temporal S2 data on crop type classification as recommended by Immitzer *et al.* (2016) and Vyas *et al.* (2005). The study confirmed that multi-date data increases the chances of finding the optimal temporal window in which the highest accuracy can be achieved and the study obtained high overall accuracies of over 85%.

Multispectral sensors use a parallel sensor array which is able to detect radiation over a small number of broad wavelength bands. The difference in spectral reflectance and crop features such as the shape and texture of the vegetation provide important parameters that should be considered in mapping highly heterogeneous landscapes (Beeresh *et al.*, 2014). The ability of separating or discriminating one crop from another accurately depends on the unique spectral signature that each crop or land use type reflects on the electro-magnetic spectrum. (Schmedtmann and Campagnolo, 2015). However, there is limitation associated with mapping crop types and discriminating them from the co-existing land uses such as the parcel to-parcel variability of the plant reflectance of the same class and spectral similarities between the co-existing land use and crop types (Schmedtmann, 2014). Accordingly, it is important to differentiate between the vegetation spectral characteristics hence multispectral data has become the obvious choice for crop mapping. This is due to its high radiometric and spectral resolution that allows acute vegetation properties to be identified. Multi-spectral data with a medium spatial resolution of 30 m or greater has been used for crop mapping (Conrad *et al.*, 2010, Beeresh *et al.*, 2014, Li *et al.*, 2014, Ouzemou *et al.*, 2018), although some scholars report that 30 m resolution data is not adequate for mapping highly heterogeneous intensive agriculture landscape as it could lead to numerous mixed parcel, merging of borders of agriculture field and low precision (Conrad *et al.*, 2010, Ouzemou *et al.*, 2018).

Several studies compared the performance of Sentinel-2 and Landsat-8 in land cover mapping (Sibanda *et al.*, 2016, Sothe *et al.*, 2017, Forkuor *et al.*, 2018). Forkuor *et al.* (2018), compared the synergistic use of the 2 sensors in mapping LULC in West Africa with the aim of exploring the potential of Sentinel-2 MSI data in improving the accuracy. The study established that the addition of the three red-edge bands in Sentinel-2 MSI assisted in obtaining an overall accuracy of 4% better than Landsat-8. This result agrees with the findings of Sothe *et al.* (2017), who also compared the performance of Sentinel-2 and Landsat-8 for discriminating different successive forest stages in Brazil and found higher accuracy in Sentinel-2 than in Landsat 8.

Essentially, understanding the strengths and weaknesses of different sensors is critical when choosing remotely sensed imagery for land use classification of crop types. The characteristics of the study area, user's need, scale, data availability and the analyst experience assist in deciding the choice of remotely sensed data to be used (Lu and Weng, 2007). The availability of free

Sentinel data within the Copernicus program (Drusch *et al.*, 2012) and Landsat 8 data (Irons *et al.*, 2012) has made crop type mapping possible. Consequently, the present study used this data because of its availability as well as the appropriate resolution both sensors offer which are suitable for crop mapping.

2.2.1 Crop type mapping using Landsat 8 OLI data

US Geological Survey archive provides access to vast Earth Resource Observation and Science (EROS) data that can be used for land use mapping including crop type mapping (Drusch *et al.*, 2012). The launch of a follow-up mission after Landsat 7 known as the Landsat Data Continuity Mission(LDCM) which is commonly referred to as Landsat 8 (Irons *et al.*, 2012), saw an increase in the frequency of freely available observations with an improved duty cycle larger than previous Landsat satellites (Zhu and Woodcock, 2014). The LDCM observatory orbits the earth every 99 minutes working in a 716 km near circular, near-polar and sun-synchronous cycle over a 16-day temporal coverage (Reuter *et al.*, 2011). The satellite follows a second Worldwide Reference System (WRS-2) stipulated path which is defined as a coordinate reference system which is used to catalog image data acquired from the previous Landsat missions of Landsat 4, Landsat 5 and Landsat 2 (Irons *et al.*, 2012).

The Landsat-8 satellite, launched in 2013, carries two imaging instruments, the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) which coincidently collect multispectral digital images in the reflective and thermal atmospheric windows (Pahlevan and Schott, 2013). The OLI has nine spectral bands that have a spatial resolution of 30m including the high resolution 15m panchromatic band as illustrated in (Table 2.1). Having an advanced spectral range for some bands which enhances spectral responses across the band channels (Jia *et al.*, 2014), OLI is expected to be suitable for discriminating the acute spectral characteristics in crop types. In addition, it has two spectral bands that include cirrus bands and coastal aerosol which is centred at the chlorophyll absorption peak wavelength together with an exceptional radiometric fidelity of 12 bits (Pahlevan and Schott, 2013), and a longer integration time resulting in higher signal to noise ratios as compared to previous Landsat missions (Irons *et al.*, 2012). Asgarian *et al.* (2016), employed Landsat-8 data to classify heterogeneous agriculture landscape in Iran and managed to accurately map the crop types with more than 90% accuracies for all the experiments. Accordingly, the Landsat 8 OLI sensor has become the obvious choice for this study as it is regarded as a gold standard among remote sensing satellites having reasonable spectral and spatial resolution as well as exceptional image orthorectification and radiometric correction (Leslie *et al.*, 2017).

2.2.2 Crop type mapping using Sentinel 2 MSI data

The Global Monitoring for Environment and Security, a joint initiative of the European Space Agency successfully launched the Sentinel-2 Multi-spectral Instrument in 2015. The optical sensor was designed as a dependable multispectral earth observation system that will ensure that data coverage over global terrestrial surfaces is accomplished by Landsat and SPOT satellite series (Sothe *et al.*, 2017). The Sentinel-2 MSI offers extraordinary capabilities which include a global temporal revisit of five days with two satellites in orbit, system global coverage of land surface and an advanced wide swath of 290 km. The Sentinel-2 satellites are twin satellites (Sentinel-2A and Sentinel-2B) which operate concurrently in a sun-synchronous orbit at an altitude of 786km (Drusch *et al.*, 2012).

The sensor offers data of high and medium spatial resolution of 10, 20, and 60m over 13 spectral bands (Table 2.1) in the visible, near infrared and short wave infra-red part of the electromagnetic spectrum (Malenovský *et al.*, 2012). The four bands which have 10m resolution (blue, green, red and near-infrared) enable compatibility with SPOT 4 and 5 (Arnau, 2016), and six bands at 20m resolution assist in the Level 2 parameters for image pre-processing (Drusch *et al.*, 2012). The remaining bands at 60m are for atmospheric corrections (aerosols retrieval) and cloud screening (cirrus detection) (Drusch *et al.*, 2012), hence they enable spatial variability of the atmospheric geophysical parameters to be sufficiently captured (Fernandez *et al.*, 2013). The the improved characteristics characteristics of the multi-spectral sensor meet the requirements for crop mapping as it has shorter revisit periods to cater for the temporal dynamics of crop growth together with medium spatial resolution over wide spectral channels enabling the discrimination of acute vegetation spectral signatures.

Considering that the satellite has been recently launched, studies to investigate its potential for crop type mapping are still emerging. Immitzer *et al.* (2016), demonstrated the potential of Sentinel-2 data for crop types and tree species mapping and concluded that high spectral, spatial and global coverage is the best in giving high classification accuracies of 76% considering six crops. Due to this background, the present study used the Sentinel-2 data to establish the performance of the data in mapping heterogeneous agriculture landscape provided by the outstanding characteristics of this sensor.

Table 2.1: Summary of the multispectral data characteristics.

Sentinel-2 MSI				**Landsat-8 OLI**			
Band Name	Wavelength(nm)	Description	Spatial Resolution (m)	**Band Name**	Wavelength(nm)	Description	Spatial Resolution (m)
Band 1	443	Aerosol	60	**Band 1**	0.43 - 0.45	Coastal	30
Band 2	490	Blue	10	**Band 2**	0.45 - 0.51	Blue	30
Band 3	560	Green	10	**Band 3**	0.53 - 0.59	Green	30
Band 4	665	Red	10	**Band 4**	0.63 - 0.67	Red	30
Band 5	705	RedEdge 1	20	**Band 5**	0.85 - 0.88	NIR	30
Band 6	740	RedEdge 2	20	**Band 6**	1.57 - 1.65	SWIR 1	30
Band 7	783	RedEdge 3	20	**Band 7**	2.11 - 2.29	SWIR 2	30
Band 8	842	NIR	10	**Band 8**	0.50 - 0.68	Pan	15
Band 8A	865	RedEdge 4	20	**Band 9**	1.36 - 1.38	Cirrus	30
Band 9	945	Water vapour	60				
Band 10	1375	SWIR-Cirrus	60				
Band 11	1610	SWIR1	20				
Band 12	2190	SWIR2	20				

Source: (Drusch *et al.*, 2012, Irons *et al.*, 2012)

2.3 Use of machine learning algorithms for crop type mapping

The quality of satellite images have improved thus conventional statistical method becomes insufficient for accurately interpreting highly complex landscapes**.** Besides selecting the right remote sensing data appropriate for the study, the right choice of classification method is also equally important for a successful crop type mapping (Sothe *et al.*, 2017). Many studies have been using the machine learning nonparametric methods land cover classification and crop mapping since their introduction. Remote sensing data classification is a complex process which requires determination of suitable classification method, image pre-processing, and a sufficient number of training samples, feature extraction, post-classification processing and accuracy assessment (Lu and Weng, 2007). The most commonly used machine learning algorithms in heterogeneous land cover classification include Decision trees, Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM) amongst others. The nonparametric classifiers ignore the assumption of a normal distribution for the dataset and statistical parameters are not required when separating image classes hence making it suitable for the use of spectral data in the classification process (Lu and Weng, 2007). The superiority of nonparametric algorithms have been extensively discussed in recent studies (Yang *et al.*, 2011, Adam *et al.*, 2014, Inglada *et al.*, 2015, Sothe *et al.*, 2017, Ouzemou *et al.*, 2018)

Nitze *et al.* (2012), comparatively evaluated ANN, RF, MLC, and SVM for crop classification using Rapid-Eye images. Although ANN produced good results, the findings reported that it has complex architecture optimization and low calculation robustness thus giving low accuracies compared to RF and SVM. Foody and Arora (1997), verified this findings as their research postulates that the variation in the dataset dimensionality and the features of the training and testing data sets diminish the overall accuracy of the classification of the ANN classifier. The modern non-parametric machine learning algorithms such as RF and SVM have been reported to overcome the shortcomings of the traditional algorithms like Maximum Likelihood because they have an ability to synthesize classification functions using either discrete or continuous datasets (Sothe *et al.*, 2017), as well as being insensitive to noise thus they are not constrained by parametric distribution assumptions (Löw *et al.*, 2013).

Adam *et al.* (2014), concurs with these finding as their study found that RF and SVM algorithms are the most effective classification approaches. The SVM classifier has emerged in some studies as the best classification method for mapping crop types based on multi-temporal and multispectral data outperforming RF by 2% (Nitze *et al.*, 2012), using a Gaussian Kennel density function (Burges, 1998, Pal and Mather, 2005, Inglada *et al.*, 2015, Kumar *et al.*, 2017). SVM was established as the most competitive algorithm in classifying high dimensional datasets (Huang *et al.*, 2002, Foody and Mathur, 2004). Other studies reported mixed performances of RF and SVM whilst other authors to report that RF performs marginally better than SVM in heterogeneous landscape classification (Adam *et al.*, 2014).

Mapping of different crop types has been successfully done recently using the advanced remote sensing technology. The free availability of the new multispectral Sentinel-2 image together with the Landsat-8 OLI imagery provides opportunity for successful crop mapping at national level which was not possible in the past. The development of machine learning algorithms offers an interesting prospect to further investigate how the classifiers perform in a heterogeneous agriculture landscape. Most studies in South Africa however, have looked more into mapping irrigated agriculture land cover particularly delineating agriculture land cover, neglecting to map and distinguishing the crop types within those irrigated lands.

It is therefore of crucial importance to evaluate the performance of the reportedly most effective algorithms in classifying crop types as previous studies have mostly used these classifiers to map land cover, forests and rangeland (Sibanda *et al.*, 2016, Sothe *et al.*, 2017, Forkuor *et al.*, 2018). The spectral and spatial capabilities of the latest open-source dataset were evaluated to ascertain their contribution in improving the discrimination and detection of different land use types and crop types in a heterogeneous agriculture landscape in South Africa. Finally, the analysis was performed to determine the optimum time to classify the crop types in Mpumalanga, South Africa.

CHAPTER THREE

Methodology

3.1 Study area and materials

3.1.1 Study area

Marble Hall is a cross border area formerly situated in Mpumalanga province but now in the south of Limpopo province of South Africa (Figure 3.1). Located at 24 $^{\circ}$58´ South and 29°17´ East the area is 26 kilometers from north-west Groblersdal and 96 kilometers south-south-east of Mokopane. Being the seat area of the Ephraim Mogale Local Municipality which is a local municipality that constitutes Sekhukhune District Municipality in the province of Limpopo, it comprises of a land area of approximately 1793 square kilometers of which 80 % is used for agricultural purposes (Maponya, 2012). The area was discovered in 1942 as a marble town owing to marble quarrying at the time which contributed to its development.

Ephraim Mogale municipality is located in platinum-rich Sekhukhune district which goes through the Bushveld complex north-eastern side. It is a massive geological structure that comprises of the world's biggest reserves of platinum group metals (Eriksson and Reczko, 1995), dominated by coarse-grained granite, sandstones with limited occurrences of other rocks. The area has a flat terrain with a slight slope down towards the east with an attitude of between

890 in the east and 910 in the west (Waals, 2013). Marble Hall climate consists of an average winter temperature of 23 ^{0}C and an average summer temperature of 29, 5 ^{0}C. The average annual precipitation is about 634 millimeters (Maponya and Mpandeli, 2012).

The area is an agricultural hub of the region considered as the citrus producer which features a pivot irrigation farming system in the Loskop dam floodplains which enables all season planting of crops. The agriculture system owes its success to the well-drained rich soils and the area is known for farming crops such as cotton, maize, potatoes, wheat and citrus which plays an important role in South Africa’s economy providing for the great demand countrywide and internationally (Cai *et al*., 2017).

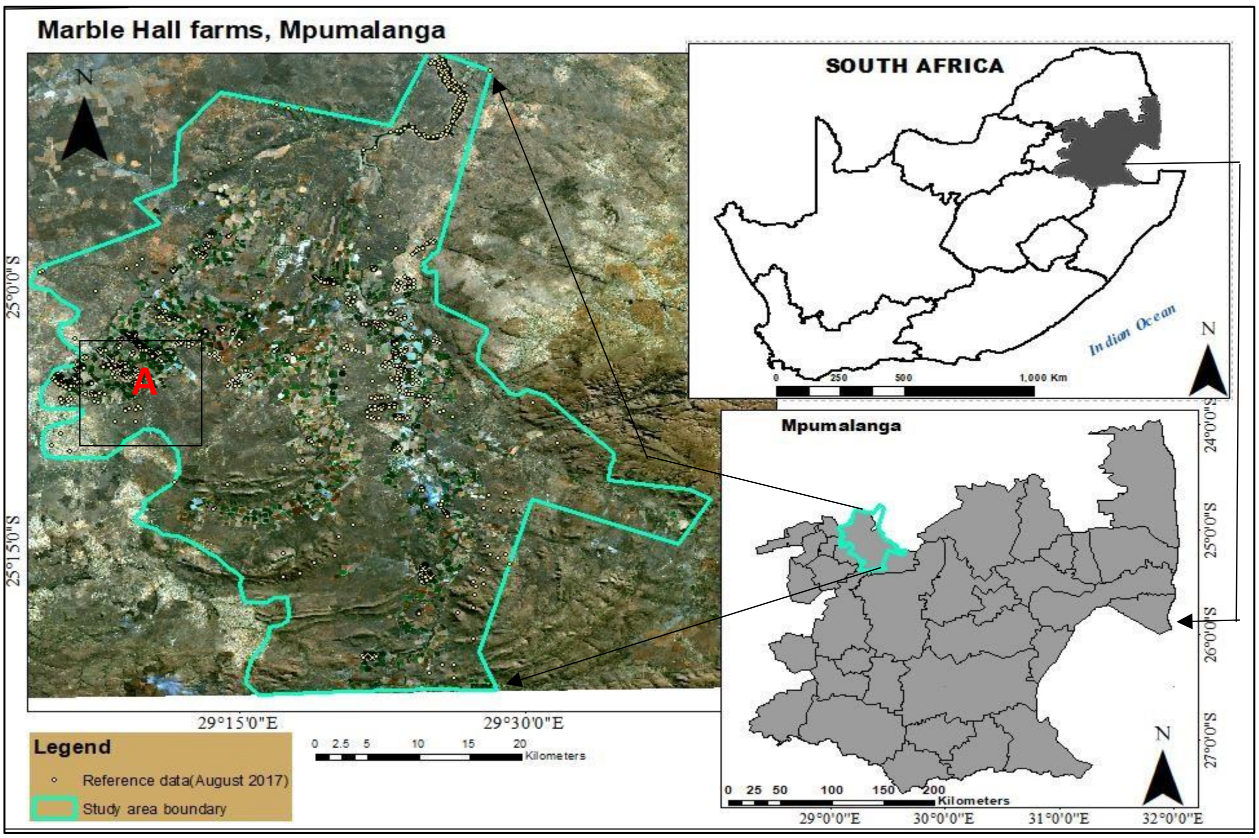


Figure 3.1: The location of the study area (Marble Hall)

Figure 3.2: August 2017 Sentinel-2 natural color composite image of the zoomed location for the study area shown as A in Figure 1.

3.1.2 Data acquisition and pre-processing

3.1.2.1. Satellite data acquisition

The multispectral images used in this study were from Sentinel-2 (S-2) Multi-Spectral sensor and Landsat 8 (L-8) Operational Land Imager. Four multi-temporal L-8 and S-2 images, acquired in the same months of 2017, were used for this study (Table 3.2). The acquisition dates correspond to the sprouting and or harvesting time of the winter crops in order to make discrimination between the crops easier and to differentiate between the different crops. The S-2 and L-8 images were downloaded free of charge from the USGS' Earth Explorer portal (https://earthexplorer.ugs.gov/).

Sentinel-2 imagery was used which has two identical satellites (Sentinel2A and 2B) with a combined constellation revisit of 5 days and 13 total bands that allows for land monitoring(ESA, 2016). Ten out of thirteen S-2 bands with a spatial resolution of 10m and 20m were used in this study (Table 3.1).

Also, Landsat 8 operational land imager imageries were used. The study utilized all seven spectral bands (Table 3.1).

Table 3.1: Spatial and spectral resolutions of L-8 OLI and S-2 MSI data used.

Sentinel-2 MSI				**Landsat-8 OLI**			
Band Name	Wavelength (nm)	Description	Spatial Resolution (m)	**Band Name**	Wavelength(nm)	Description	Spatial Resolution (m)
Band 2	490	Blue	10				
Band 3	560	Green	10	**Band 2**	0.45 - 0.51	Blue	30
Band 4	665	Red	10	**Band 3**	0.53 - 0.59	Green	30
Band 5	705	RedEdge 1	20	**Band 4**	0.63 - 0.67	Red	30
Band 6	740	RedEdge 2	20	**Band 5**	0.85 - 0.88	NIR	30
Band 7	783	RedEdge 3	20	**Band 6**	1.57 - 1.65	SWIR 1	30
Band 8	842	NIR	10	**Band 7**	2.11 - 2.29	SWIR 2	30
Band 8A	865	RedEdge 4	20	**Band 8**	0.50 - 0.68	Pan	15
Band 11	1610	SWIR1	20				
Band 12	2190	SWIR2	20				

Table 3.2: Acquisition dates of the L-8 OLI and S-2 MSI scenes used.

Satellite	Scene	Acquisition date	Satellite	Tile	Acquisition date
Landsat-8 OLI	170/77	17/06/2017	Sentinel-2 MSI	T35JQN	21/06/2017
Landsat-8 OLI	170/77	19/07/207	Sentinel-2 MSI	T35JQN	21/07/2017
Landsat-8 OLI	170/77	20/08/2017	Sentinel-2 MSI	T35JQN	20/08/2017
Landsat-8 OLI	170/ 77	21/0 9/2017	Sentinel-2 MSI	T35J QN	19/09/2 017

3.1.2.2. Image pre-processing

The Sentinel-2 images were georeferenced and geo-coded using UTM zone 35S. The images were atmospherically corrected in order to get true reflectance values of the features of the image. This was done using the Level 2A prototype processor (Sen2Cor) which is a plugin in Snap software package. SenCore is a prototype processor developed by ESA for Sentinel-2A product processing. Level 2A processing is applied to the Top-Of-Atmosphere Level 1C ortho-image reflectance products to produce ortho-image Bottom-Of-Atmosphere (BOA) corrected 13 JPEG-2000 images which are associated with the 13 Sentinel-2 spectral bands. The generated output images are resampled and generated with an equal spatial resolution for all bands and for this study, 10m resolution was selected. The model is equipped with its own python distribution based on *Anaconda* which uses the additional third-party software *pyTables* and *GDAL* (Muller-Wilm *et al*., 2013).

All S-2 bands with a spatial resolution of 20m (Band 5, 6, 7, 8A, 11 and 12) were resampled to 10m using the nearest-neighbor interpolation algorithm to enable comparison with 10m bands (Band 2, 3, 4 and 8) without reducing the high spatial resolution of these bands. The corrected image was exported to BEAM-DIMAP to enable further analysis in ArcGIS. The corrected 10 bands (Table 3.1) were loaded into ArcMap 10.4 and stacked chronologically into one image which was then converted to .tiff format. This operation was independently applied to all four S-2 MSI images. The group points were overlaid on each of the images and spectral reflectance values that correspond to each GPS point were extracted. These points were exported to .csv format and further analysis was done using R programming language.

The Landsat-8 OLI data were obtained as standard L1t files which are geo-registered and orthorectified to Universal Transverse Mercator 35S projection based on the World Geodetic Datum (WGS 84) coordinates system. The digital number (DN) values of the multispectral bands were converted to Top-of-Atmosphere (TOA) spectral radiance then to at-sensor surface reflectance values using the rescaling coefficients which are provided in the image metadata file. Fast LineOf-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm in Environment for Visualizing Images version 5.4 software was used to enable the retrieval of accurate reflectance spectra for further analysis. This operation was independently applied to all four L-8 OLI images with the corrected bands. Ground points were overlaid on the images and spectral reflectance values that correspond to each GPS point were extracted. These points were then exported to .csv format and further analysis was done using an R programming language.

3.1.3 Field data collection

The ground control points were collected through a field survey which was conducted on the 20th of August 2017. The sample points were randomly collected within the study area (Figure 3.1). A total of 14 crop types/LULC classes were identified during the field survey. A Global positioning system receiver (GPS) was used to record the longitude and latitude coordinates of the different crop types and land use classes in the study area. Sample points for the consecutive months were manually digitized by visual interpretation of SPOT 6/7 and S-2 MSI imagery, eleven classes in June, thirteen in July and fourteen classes in August and September. The different number in crop types was mainly attributed to different crop planting and harvest dates. The ground control points were split into 70% training and 30% validation for subsequent classification as shown in Table 3.3.

Table 3.3: Training and validation data used for classification.

Crop type/LULC class	**Code**	**Training (70%)**	**Test (30%)**	**Total GCP**
Bare rock/soil	BR	42	19	61
Bushland	BL	56	25	81
Cabbage	C	56	24	80
Cleared field	CF	105	46	151
Green beans	GB	113	48	161

Greenhouses	GH	87	37	124
Lemon tree	LE	73	33	106
Lucerne	LS	70	30	100
Maize	MZ	88	37	125
Orange tree	O	52	21	73
Potatoes	PT	86	37	123
Settlement	ST	24	10	34
Water body	WB	56	25	81
Wheat	WT	77	33	110

3.2 Image classification

Supervised machine learning algorithms were used for the study because prior knowledge about the study area was known. Support Vector Machine and Random Forest algorithms were used to classify crop types and LULC around the study area. These two classifiers were selected due to their high performance in classifying crop types and heterogeneous landscapes (Nitze *et al.*, 2012).

3.2.1 Support Vector Machine classification

Support Vector machine is supervised classifier non-parametric technique which is trained to find an optimal classification hyperplane by grouping classes based on the statistical learning theory hence no assumption on the underlying data distribution (Mountrakis *et al.*, 2011). SVM was developed in 1992 by Vladimir Vapnik and it quickly gained popularity in the pattern recognition communities due to its theoretical and computational merits as it overcomes difficulties such as noisy data, high dimensionality and non-Gaussian distribution of data (Soman *et al.*, 2009). The hyperplane in SVM was developed using the training data and the ability of the hyperplane was validated using the independent testing data set (Table 4). Th**e** classifier is a distribution-free algorithm which functions by maximizing the distance between data points of each land-cover class to the best optimal separating linear hyperplane (Petropoulos *et al.*, 2011). The optimum hyperplane used to measure the margins is termed as support vectors(Kumar *et al.*, 2017). The SVM algorithm identifies the optimal decision boundary between classes thus minimizing misclassifications of the crop types (Mountrakis *et al.*, 2011). The principal advantage of SVM is that it could successfully work with a small number of training samples. The penalty parameter(C) is the introduced if misclassification in unavoidable and it controls the trade-off between margin a misclassification error (Kumar *et al.*, 2017). Where classes are not linearly separable, a kernel function is used allowing points to scatter thus a linear hyperplane can be fitted (Foody and Mathur, 2004). The kernel functions were used to transform non-linear problems into linear problems in order to map the original input space into a new feature space with higher dimensions (Oommen *et al.*, 2008). For this study, radial basis function was used due to its popularity in remote sensing studies as it outperforms the other kernels (Pal and Mather, 2005, Inglada *et al.*, 2015, Kumar *et al.*, 2017).

The equation for the Radial Basis Function is as follows (Kumar *et al.*, 2017):

$$K\colon (x, xj) = (-y\|(xi, xy)\|\ \ 2\ ), y > 0$$

Where *y* is the gamma which controls the width of the Gaussian kernel function.

The accuracy of the SVM algorithm is determined by algorithm parameterization which requires the regularization of parameter *C*-cost and definition of the kernel parameter *y*- gamma (Oommen *et al.*, 2008).

In this study, the two parameters, *C* and *Y* were obtained by testing ranges of parameters using a grid search with internal performance estimation. 10-fold cross validation was utilized which overcomes the overfitting problems (Huang *et al.*, 2002). Pairs of y and c were tested hence parameters with the best performance were automatically used for training the SVM model. The gamma and cost parameter values were taken at 0.33 and 100 respectively. The optimized SVM parameter values were then used to classify the images. S-2 MSI and L-8 OLI bands (Table 3.1) were used to define the space feature of SVM. The SVM model was run in R statistical software using caret and e1071 package (Hornik *et al.*, 2006). This procedure was done for each of the L-8 OLI and S-2 MSI images.

3.2.2 Random Forest classification

Random Forests are an ensemble of tree predictors that depend on the value of independent random vector sampled for all the trees in the forest (Breiman, 2001). Each tree contributes with a single vote of the most frequent class from the input data. The ensemble classification constructs a set of classifiers and then it classifies new data points by taking a vote of their predictions (RodriguezGaliano *et al.*, 2012). The algorithm is combined with many ensemble regression and classification trees to build binary classification trees using several bootstrap samples with other trees drawn from the original dataset. It is superior to many tree-based algorithms since it is not sensitive to noise, highly accurate, robust to outliers as well as the calculation of different internal diagnostic indicators like out of bag (OOB) and variable importance thus not subject to overfitting (Breiman, 2001). RF constructs many decision trees using the training dataset (70%) which is a random, bootstrapped subset hence keeping the remaining dataset (30%) for the internal error testing (Nitze *et al.*, 2015).

Two parameters were defined to initiate the algorithm, the number of trees used in the forest (*ntree*) and a number of random variables used in each tree (*mtry*) (Breiman, 2001). For improved accuracy, the parameters (*mtry* and *ntree*) that provide sufficiently low correlation with adequate predictive power were optimized using a grid search approach. The optimum results were obtained by setting the number of variables (*m*) equal to the square root of the number of overall variables (***M***) as suggested by Breiman (2002). The algorithm makes use of the Classification and Regression Trees (CART) to generate trees (Breiman, 2001), hence the tree node is split according to a criterion. For this research, the GINI index criterion which measures the impurity of a given element with respect to the rest of the classes (Rodriguez-Galiano *et al.*, 2012), is used to perform the split.

Gini Index can be written as the equation below (Ok *et al.*, 2012):

$$\sum \sum_{j \neq i} \left(\frac{f(Ci,T)}{|T|} \right) \left(\frac{f(Ci,T)}{|T|} \right)$$

Where T is the 70% training set

C_i is the class that a randomly selected pixel belongs to

f(Ci, T)

$\frac{\ }{|T|}$ is the probability that the selected case belongs to class C_i

An increase in the GINI index indicates an increase in class heterogeneity whilst a decrease in the index indicates an increase in class homogeneity (Rodriguez-Galiano *et al.*, 2012). Variable importance was calculated using the mean decrease gini which provides a measure of the contribution of each variable in training the data into defined LULC classes during model building process (Forkuor *et al.*, 2018). The relative importance of different variables was assessed during the classification process. This was useful for knowing how each predictive variable influences the classification model in order to enable the selection of the best variables as stipulated by Pal and Mather (2005). The random forest package in the R statistical package was used to implement the algorithm.

3.3. Accuracy assessment and map comparison

The accuracy assessment was done on all images from both sensors using the 30% validation data. This was done for all the images from both sensors in order to evaluate the ability of each sensor in discriminating the crop types from the LULC classes. For comparison purposes, confusion matrices were constructed for each thematic map to compute the overall, producers' and user's accuracies (Congalton and Green, 2008). The kappa coefficient was calculated which provide a measure of the difference between the actual agreement between reference data and the algorithm used to perform classification versus the probability that the reference data and the algorithm would agree.

CHAPTER FOUR

Results

4.1 Optimization of the algorithms

4.1.1 Optimization of SVM parameters

The SVM parameters were optimized for classification using the SVM-Kernel radial function. The SVM classification model was optimized to select the best parameters to use in training the classification algorithm of nine grass species. 10-fold cross-validation was used which yielded a gamma (Y) of 1, the lowest error was produced from the combination of *g*amma (γ) value of 0.1 and cost (C) value of 100 for both Landsat-8 and Sentinel-2 experiments (Figure 4.1).

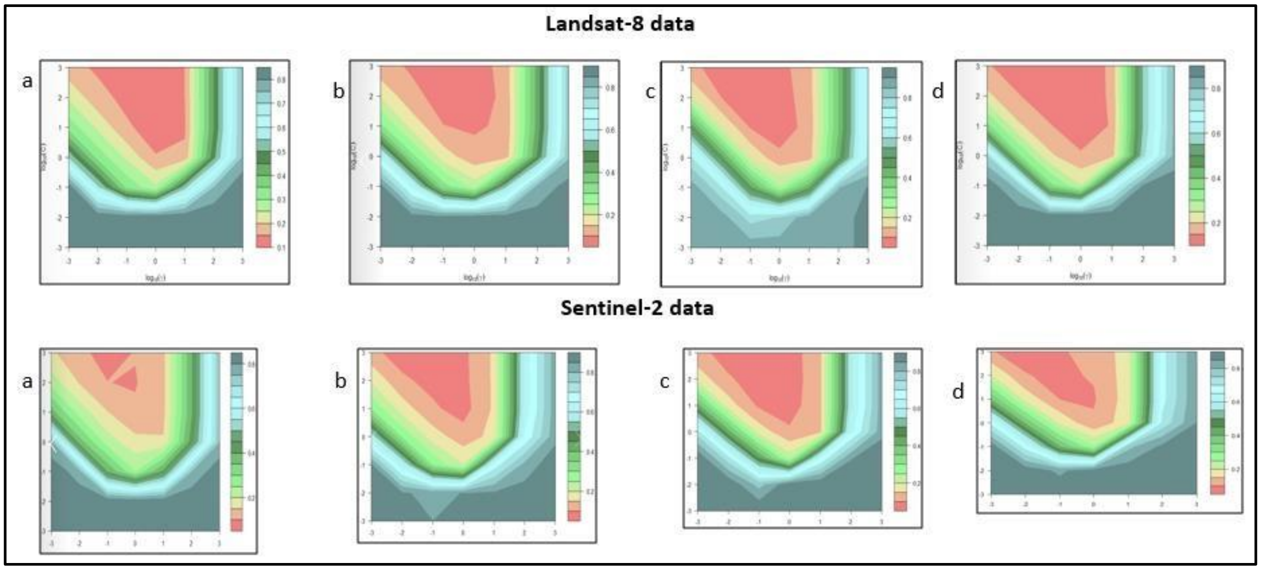


Figure 4.1: Parameter optimization for the SVM using 10-fold search radius for L-8 image for (a) June (b) July (c) August (d) September, and S-2 data for (a) June (b) July (c) August (d) September

4.1.2 Optimization of RF parameters

There are two parameters that significantly affect the performance of the RF algorithm, mtry and the ntree. The RF input parameters (m*try* and n*tree*) were optimized using the grid search procedure and a 10 fold cross validation (CV) method and the results are as shown in Table 4.1 for all the images. The RF optimization graph for all images is as shown in Figure 4.2. Consequently, m*try* value of 5 combined with n*tree* value of 450 produced the lowest OOB error of 9 % based on August Sentinel-2 data (Figure 4.2 c). The chosen *mtry* and *ntree* from the grid search window for both S2 and L8 0LI for all the months. The results illustrated that the highest OOB error was obtained in September for L8 OLI image (14%). The findings indicate that the highest OOB error corresponds to the image with the lowest OA whilst the lowest OOB error correspond to the image with the highest OA using RF classification (Table 4.1).

Table 4.1: RF selected input parameters and the resulting OOB and OA percentage.	**June**		**July**		**August**		**September**	
Sensor	L-8 OLI	S-2 MSI	L-8 OLI	S-2 MSI	L-8 OLI	S-2 MSI	L-8 OLI	S-2 MSI

Mtry	6	5	4	3	2	2	2	6
Ntree	250	800	450	150	350	900	350	100
OOB error rate (%)	12	12	9.5	11.5	0.90	11.0	14	13.6

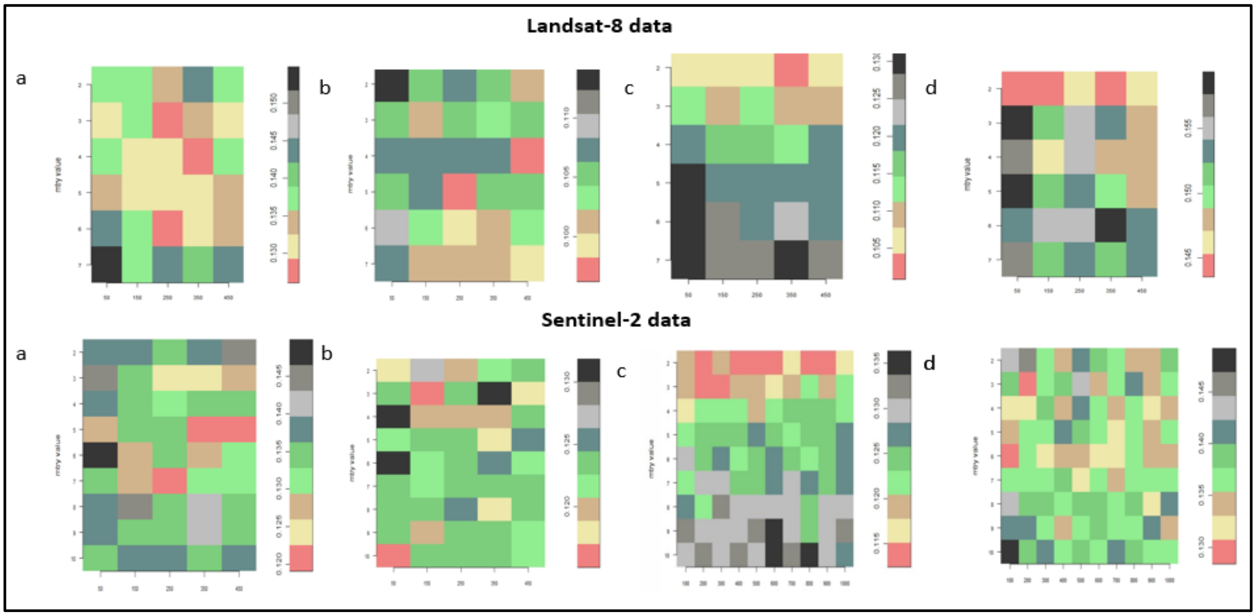


Figure 4.2: Results of RF optimization using cross validation method and training data sets based on L-8 data for (a) June (b) July (c) August (d) September, and S-2 data for (a) June (b) July (c) August (d) September.

4.2 Performance of RF and SVM in mapping crop types and co-existing land use using multi-temporal Sentinel-2 and Landsat-8 data

The thematic maps were generated using RF and SVM classification algorithms. To assess the performance of the classifiers, the area classified by the two algorithms based on L8 and S2 data were computed and tabulated to enable comparison. The study area is mostly covered by bushland, cleared field, wheat, green beans, orange trees and bare rock/soil. Potatoes, green beans, settlements, water body and cabbage cover a small percentage of the area as shown in Table 4.2 (b) and 4.3 (b). Further explanations of the tables are below. The two classification methods used in this study generally showed a reasonable accurate visual depiction of the heterogeneous agricultural land use. fix as: The two classification methods used in this study generally showed a reasonable

Table 4.2 (a): Area (hectares) covered by the crop type/ land use classes as classified by RF and SVM classifiers based on Landsat-8 data

(a)	JUNE		JULY		AUGUST		SEPTEMBER	
Classes	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**
Bare Rock/soil	7585.17	7615.91	7576.24	8423.64	4123.91	6643.6	8391.97	7738.84
Bushland	5719.14	5553.89	5431.73	4345.03	6564.85	4559.64	5697.03	6620.36
Cabbage			376.81	90.62	77.5	76.61	41.83	20.54
Cleared Field	5024.25	4752.33	5354.73	5435.35	5419.82	6285.79	4322.48	3475.66
Green Bean			764.54	483.18	214.29	228.37	317.72	253.54
Green Houses	501.68	484.51	476.22	460.09	611.64	487.12	558.21	770.4
Lemon	192.4	386.05	139.84	412.04	215.49	503.91	386.62	533.22
Lucerne	391.76	430.44	126.29	213.32	86.61	96.34	124.54	69.98
Maize					2144.55	697.43	70.77	50.19
Orange	766.61	686.15	653.37	637.46	1255.33	555.35	652.16	810.17
Potatoes	40.19	11.94	204.12	184.89	357.05	381.97	268.08	216.74
Settlement	629	1022.7	305.49	727.01	330.61	811.47	473.05	619.74
Water Body	207.68	207.37	209.32	249.06	192.66	235.3	284.56	380.33
Wheat	932.77	839.36	371.46	328.47	396.29	427.7	401.67	430.98
Total area (hectares)	21990.6	21990.65	21990.6	21990.6	21990.6	21990.6	21990.6	21990.6

As shown in Table 4.2a, the area was dominated by cleared field (5024.25 ha), bare rock/soil (7585 ha) and bushland (5719 ha) in June. The was introduction of new crop type in July such as cabbage(3768 ha) which is favorable in winter and the emergence of greenhouses increased to cater for winter vegetables planting. Waterbody remained the least dominant class due to rain offseason in the area. There was a decrease in area covered by wheat from 932 ha in June to 371 ha in July. This might be because the wheat was reaching harvest time as winter was ending. The area covered by each of the land classes was nearly the same in all months studied except for bare soil in August where the area classified by RF (7576.24 ha) was significantly different from the area classified by SVM (8423.64 ha). This may be because SVM managed to classify the bare rock/soil accurately thus it has a high accuracy compared to RF.

Table 4.2 (b): Percentage of the area covered by the crop type/ land use classes as classified by RF and SVM classifiers based on Landsat-8 data

(b)	JUNE		JULY		AUGUST		SEPTEMBER	
Classes	RF	SVM	RF	SVM	RF	SVM	RF	SVM
Bare Rock/soil	35	35	34	38	19	30	38	35
Bushland	26	25	25	20	30	21	26	30
Cabbage	0	0	2	0	0	0	0	0
Cleared Field	23	22	24	25	25	29	20	16
Green Bean	0	0	3	2	1	1	1	1
Green Houses	2	2	2	2	3	2	3	4
Lemon	1	2	1	2	1	2	2	2
Lucerne	2	2	1	1	0	0	1	0
Maize	0	0	0	0	10	3	0	0
Orange	3	3	3	3	6	3	3	4
Potatoes	0	0	1	1	2	2	1	1
Settlement	3	5	1	3	2	4	2	3
Water Body	1	1	1	1	1	1	1	2
Wheat	4	4	2	1	2	2	2	2
Total area (%)	100	100	100	100	100	100	100	100

The Table 4.2 (b) shows the percentage of the area classified by the 2 algorithms using Landsat-8 data. A comparative account of the class area that the algorithm manages to classify is given in the Table 4.2 (b) above. The figures show the percentage change in the classes over time. For instance using RF algorithm, percentage area covered by wheat decreased from 4 % in June to 2 % in September. Potatoes, green beans, settlements, water body and cabbage cover a small percentage of the area. Also, greenhouses, lemon trees, and water body occupy less than 5 % of the study area. This may be attributed to the seasonality as winter season planting was just starting for some crops hence more cleared fields and bare soil. The maize crop was planted in August as the summer season is starting as visibly shown in the map. The area covered by orange and lemon trees starts to increase as the citrus trees have fully developed and sprouting. Also, green beans, cabbage, and maize are accurately discriminately and identified in the August month.

Wheat remained constant throughout the four months showing that the Marble hall farms plant wheat crop throughout the winter and summer season due to availability of effective irrigation system.

Table 4.3 (a): Area (hectares) covered by the crop type/land use classes as classified by RF and SVM classifiers based on Sentinel 2 data

(a)	JUNE		JULY		AUGUST		SEPTEMBER	
Classes	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**
Bare Rock/soil	6289.8	13504	10285	15146	7366	13204	5915	10275
Bushland	11335	15968	13229	12313	15465	14089	20549	17007
Cabbage	0	0	679	168	54	94	69	54
Cleared Field	12185	12220	17037	13205	13755	9872	12637	11802
Green Bean	0	0	1630	889	687	618	715	375
Green Houses	1593.7	1623.8	1583	1201	2121	1925	1419	1659
Lemon	504.08	644.21	267	511	525	904	560	986
Lucern	1743.7	971.59	272	270	238	203	273	202
Maize	0	0	0	0	4648	2730	1098	184
Orange	1121.7	1620.8	1332	1745	1550	1365	1435	2343
Potatoes	216.2	59.02	352	502	679	1476	377	300
Settlement	818.28	1052.6	1306	2088	936	1551	2106	1928
Water Body	474.22	605.42	612	591	470	511	976	986
Wheat	4900.9	3894	888	842	978	931	947	979
Total area (hectares)	**49472**	**49472**	**49472**	**49472**	**49472**	**49472**	**49472**	**49472**

Potatoes occupied the least area of 216.19 ha. This might be attributed to the seasonality as less areas have started planting the potato crops. Water body occupies only 208 ha as there is absence of rain in the month of June. Wheat (3894 ha) occupied more area in June compared to all the four months (Table 4.3). This might be because it is more favourable in winter due to low temperatures thus more wheat was planted this month. There was no significant difference in the area classified by the two classifiers using S2 data for all the classes except for the lemon trees in July which showed substantial difference between RF (267 ha) and SVM (511 ha). Settlements occupied 818.28 ha of the area showing that the area was mainly an agriculture area. Bushland and cleared field were dominant from

August. This might be attributed to the fact that the fields are cleared at this time to make way for summer planting. Lucern (203 ha) and water body (511 ha) occupied the least area due to the dry season (Table 4.3a). The area covered classified in water body was the largest in September (976 ha) compared to other months

Table 4.3 (b): Percentage of the area covered by the crop type/land use classes as classified by RF and SVM classifiers based on Sentinel 2 data

(b)	JUNE		JULY		AUGUST		SEPTEMBER	
Classes	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**
Bare Rock/soil	21	22	21	31	15	27	12	21
Bushland	23	32	27	25	31	28	42	35
Cabbage	0	0	1	0	0	0	0	0
Cleared Field	25	25	34	27	28	20	26	24
Green Bean	0	0	3	2	1	1	2	1
Green Houses	5	3	3	2	4	4	3	3
Lemon	2	1	1	1	1	2	1	2
Lucerne	6	2	1	1	0	0	1	0
Maize	0	0	0	0	9	6	2	0
Orange	2	3	3	4	3	3	3	5
Potatoes	4	0	1	1	1	3	1	1
Settlement	2	2	3	4	2	3	4	4
Water Body	1	1	1	1	1	1	2	2
Wheat	10	8	2	2	2	2	2	2
Total area (%)	100	100	100	100	100	100	100	100

The table 4.3 (b) shows the percentage of the area classified by the 2 algorithms using Sentinel 2 data. It gives a comparative account of the class area that the algorithm manages to classify. The classifier shows in August that 31 % of the total area is covered by bare rock/soil followed by cleared field which covers 27 %. This might be because this is the time when fields are cleared to prepare for summer crop season. As shown above, the 2 algorithms perform almost the same for all the classes except for bushland and bare soil where there are significant difference in the area covered within the same month. RF shows that in August, 15 % of the area is covered by bare rock/ soil whilst SVM shows a 27% coverage. The area occupied by greenhouses increased from the previous month. This might be due to the increase in temperature hence need to protect vegetables from the harsh environment.

4.2.1 RF classification of multi-temporal Landsat-8 OLI data.

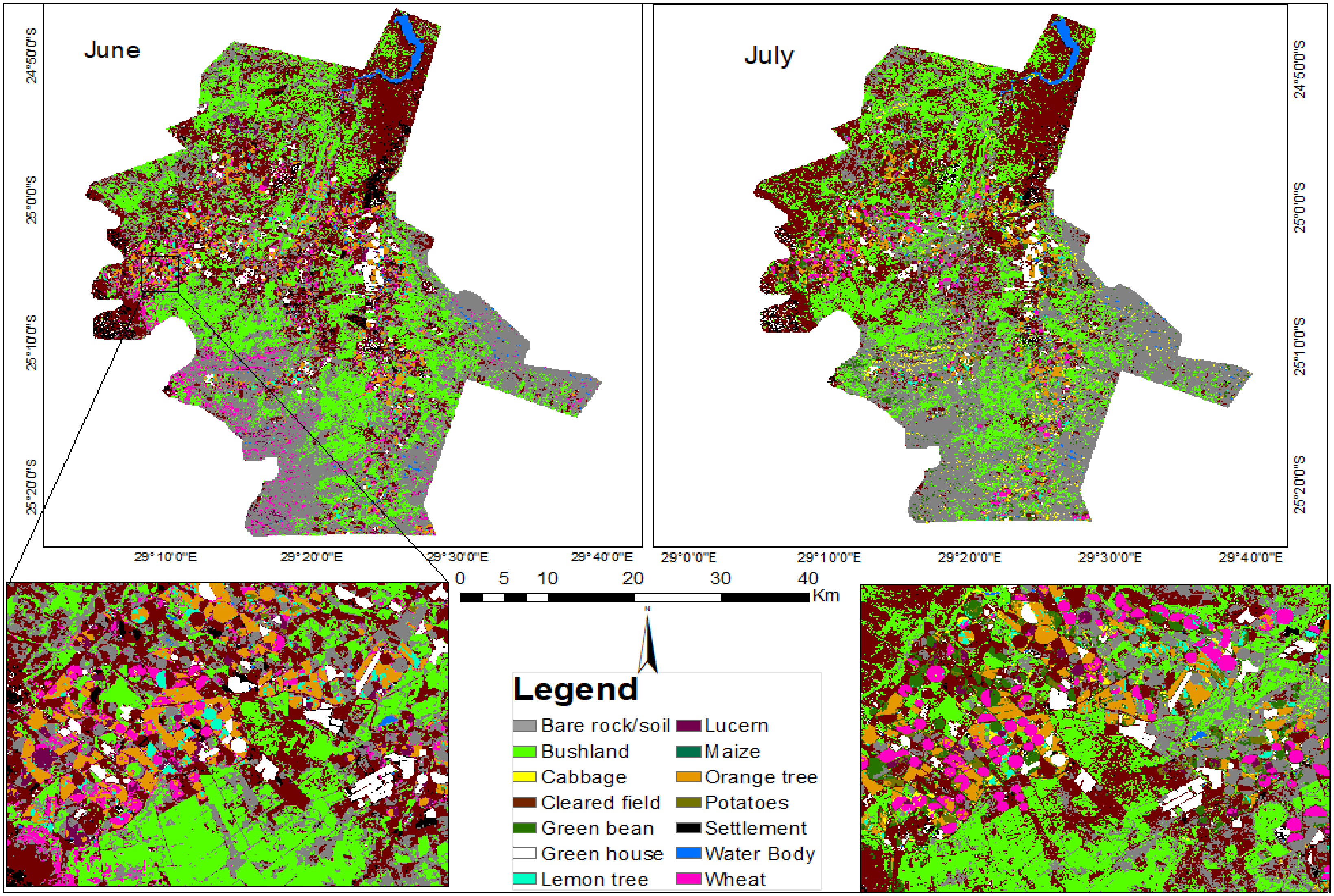


Figure 4.3: June and July classification results

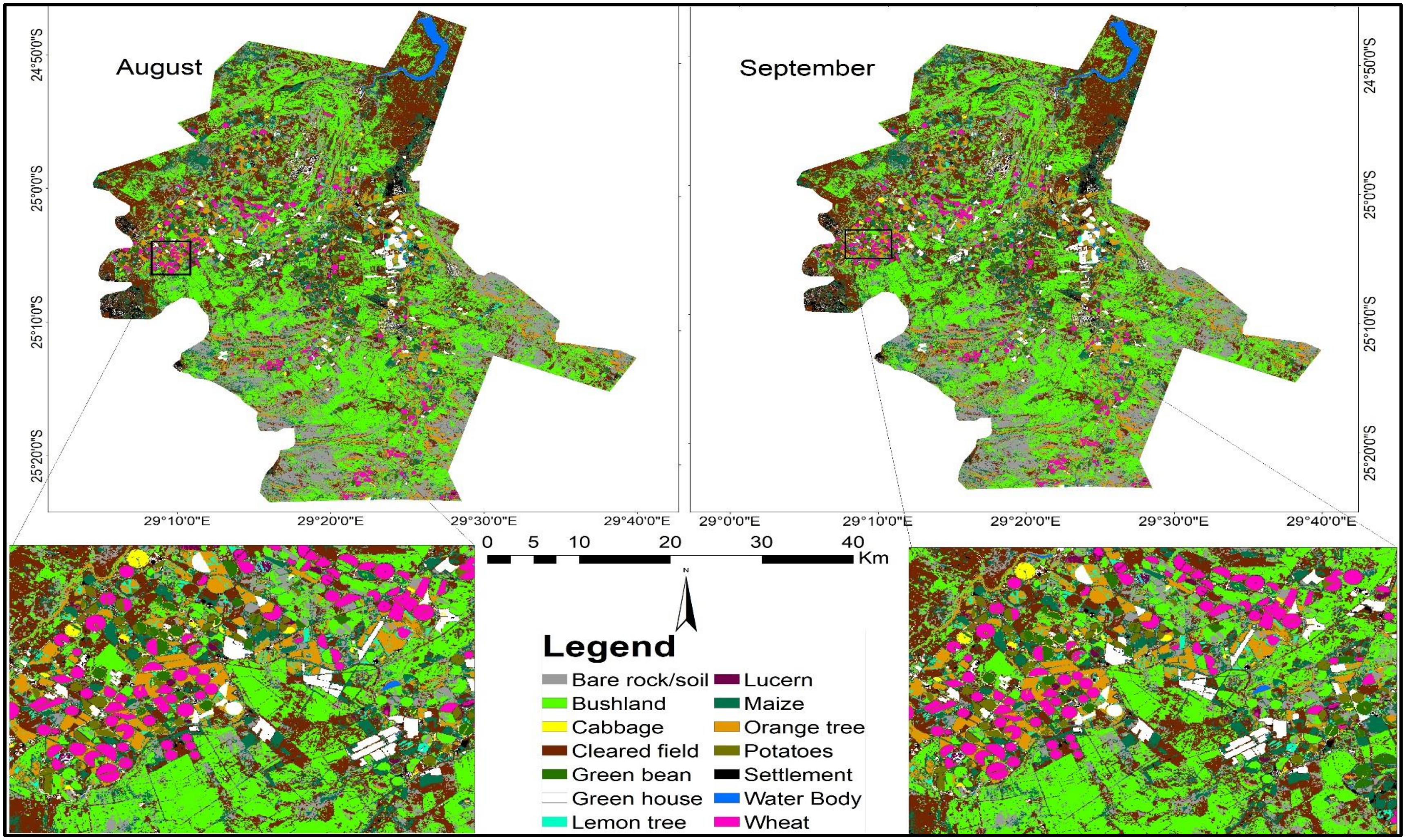


Figure 4.4: August and September classification results

4.2.2 SVM classification of multi-temporal Landsat-8 OLI data.

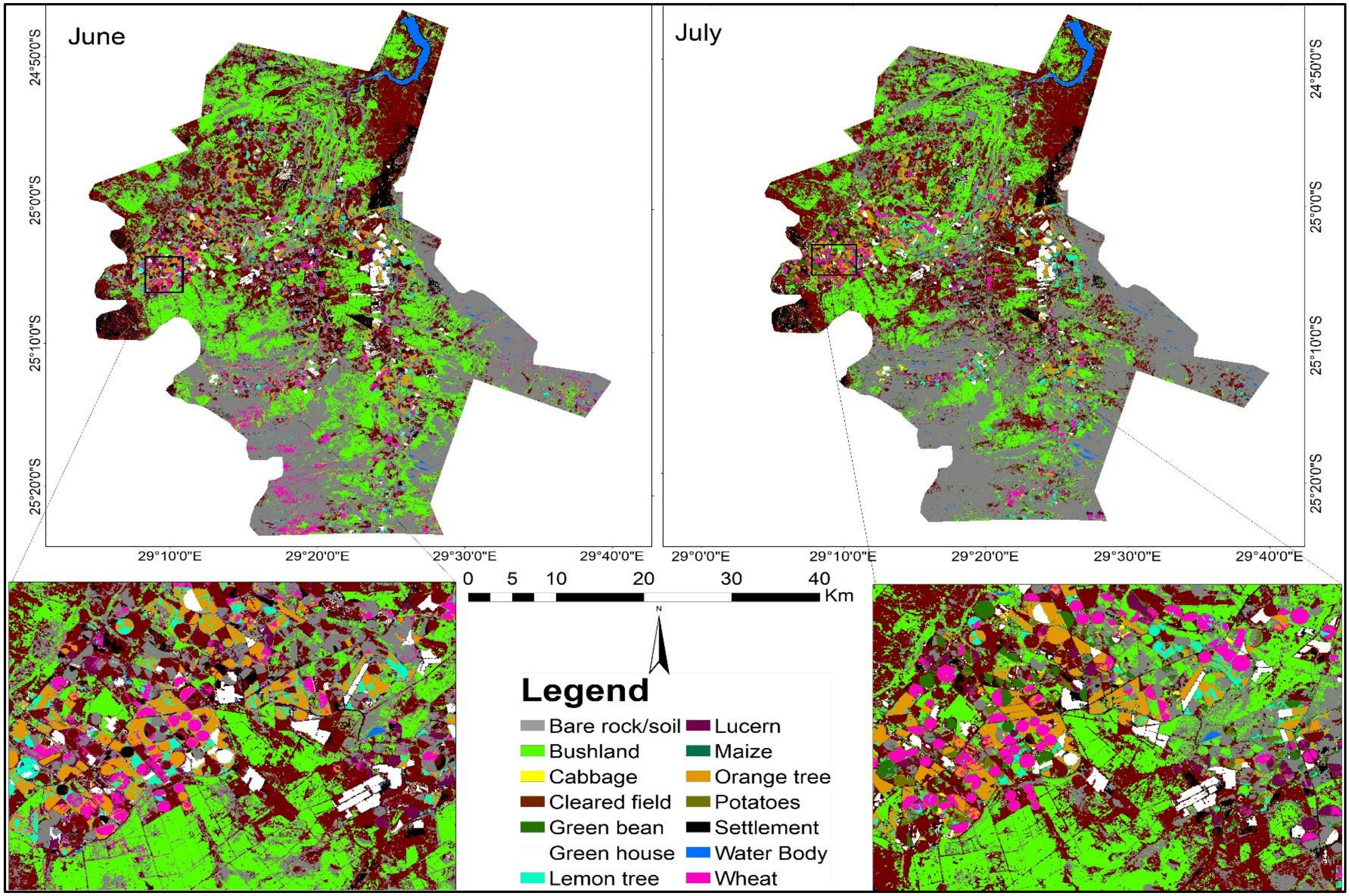


Figure 4.5: June and July classification results

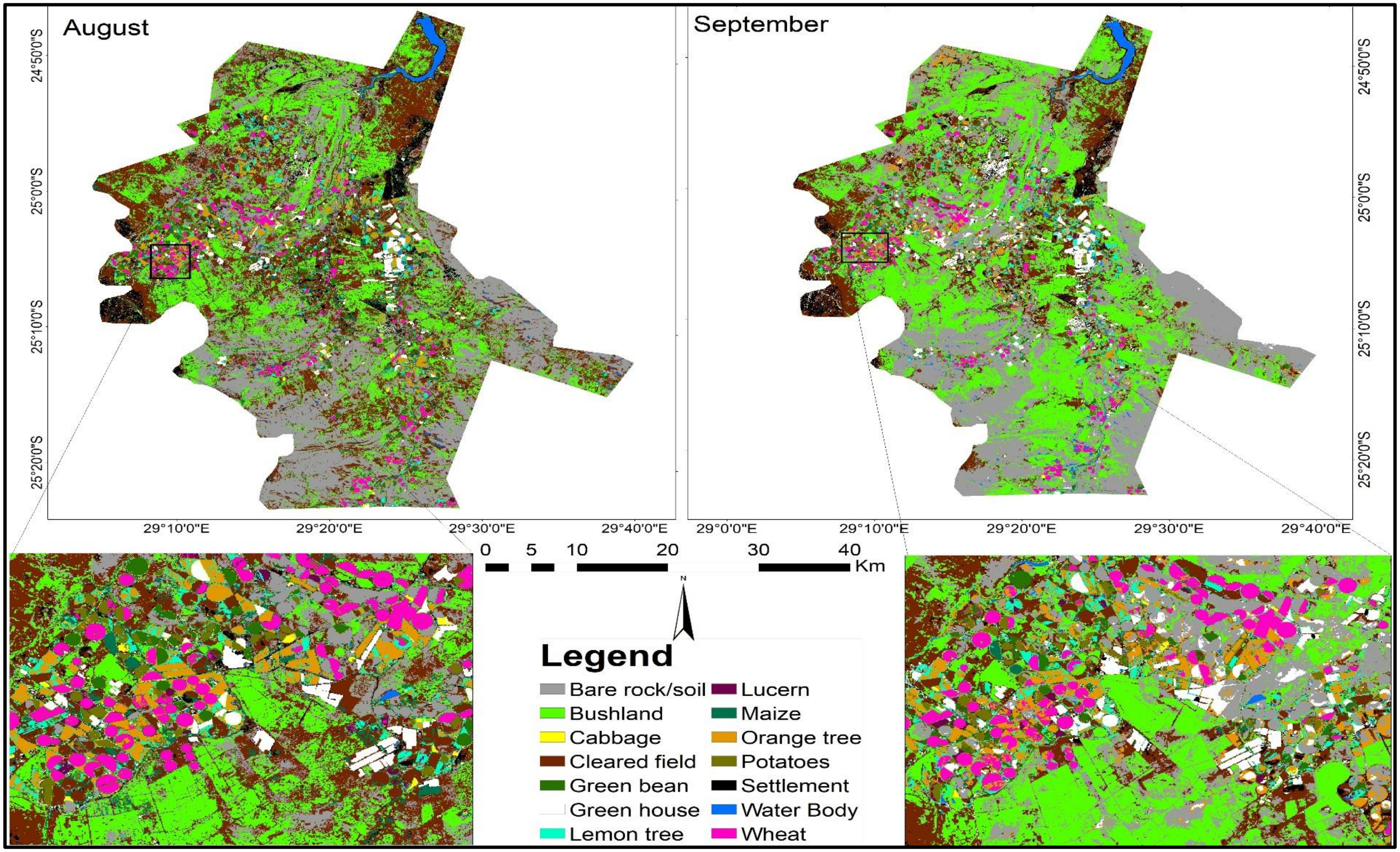


Figure 4.6: August and September classification results

4.2.3 RF classification of multi-temporal Sentinel-2 MSI data.

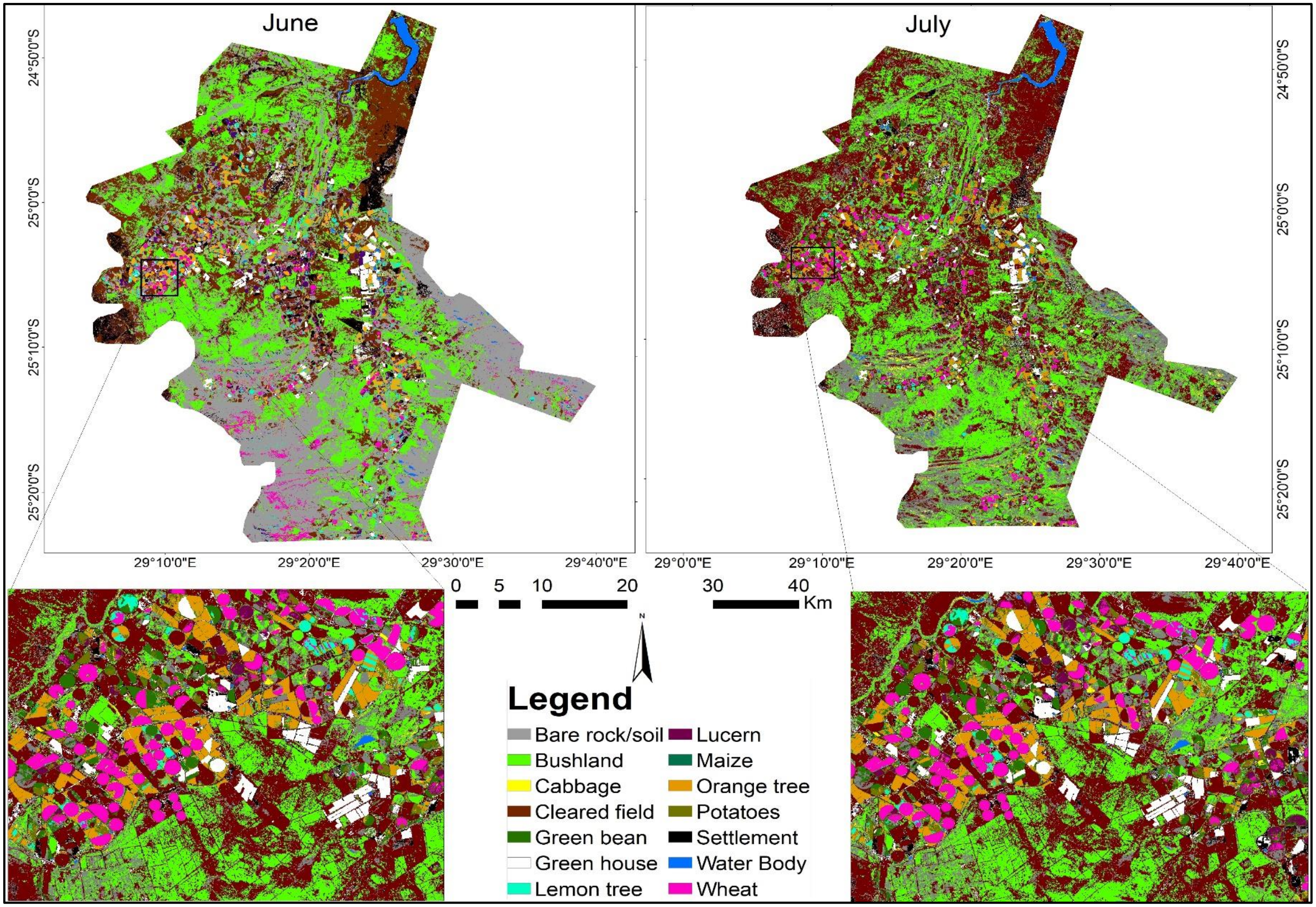


Figure 4.7: June and July classification results

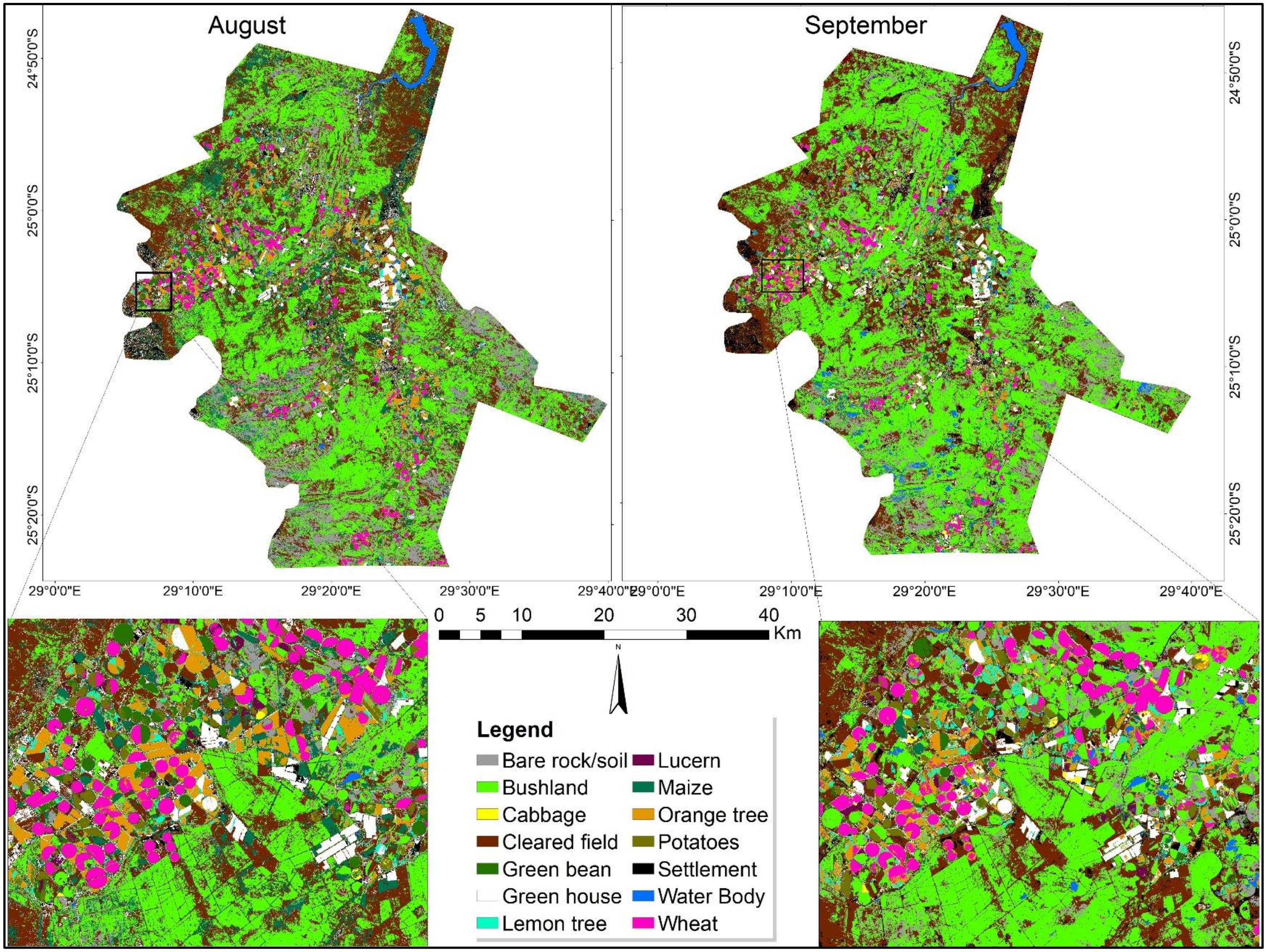


Figure 4.8: August and September classification results

4.2.4 SVM classification of multi-temporal Sentinel-2 MSI data

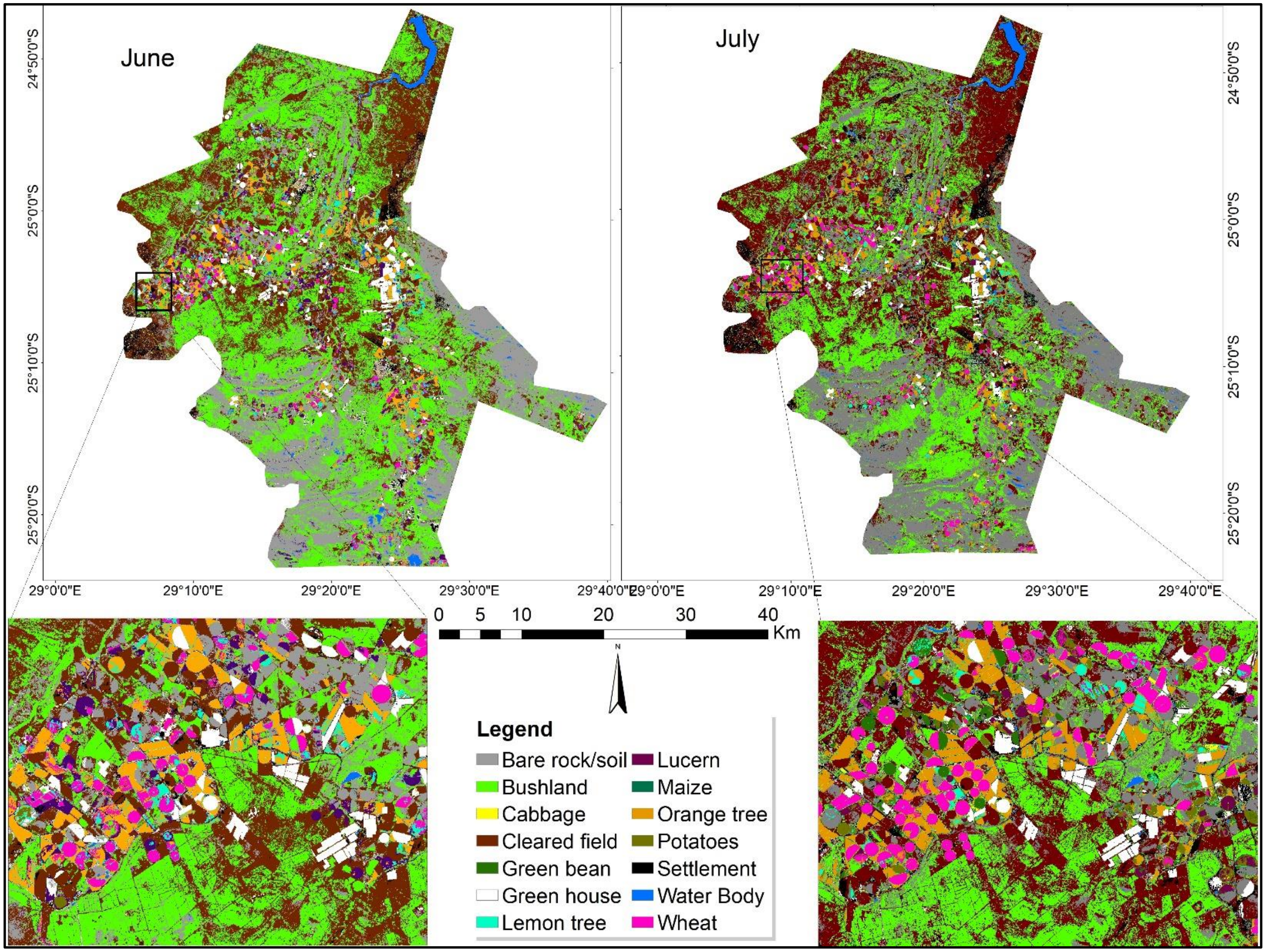


Figure 4.9: June and July classification results

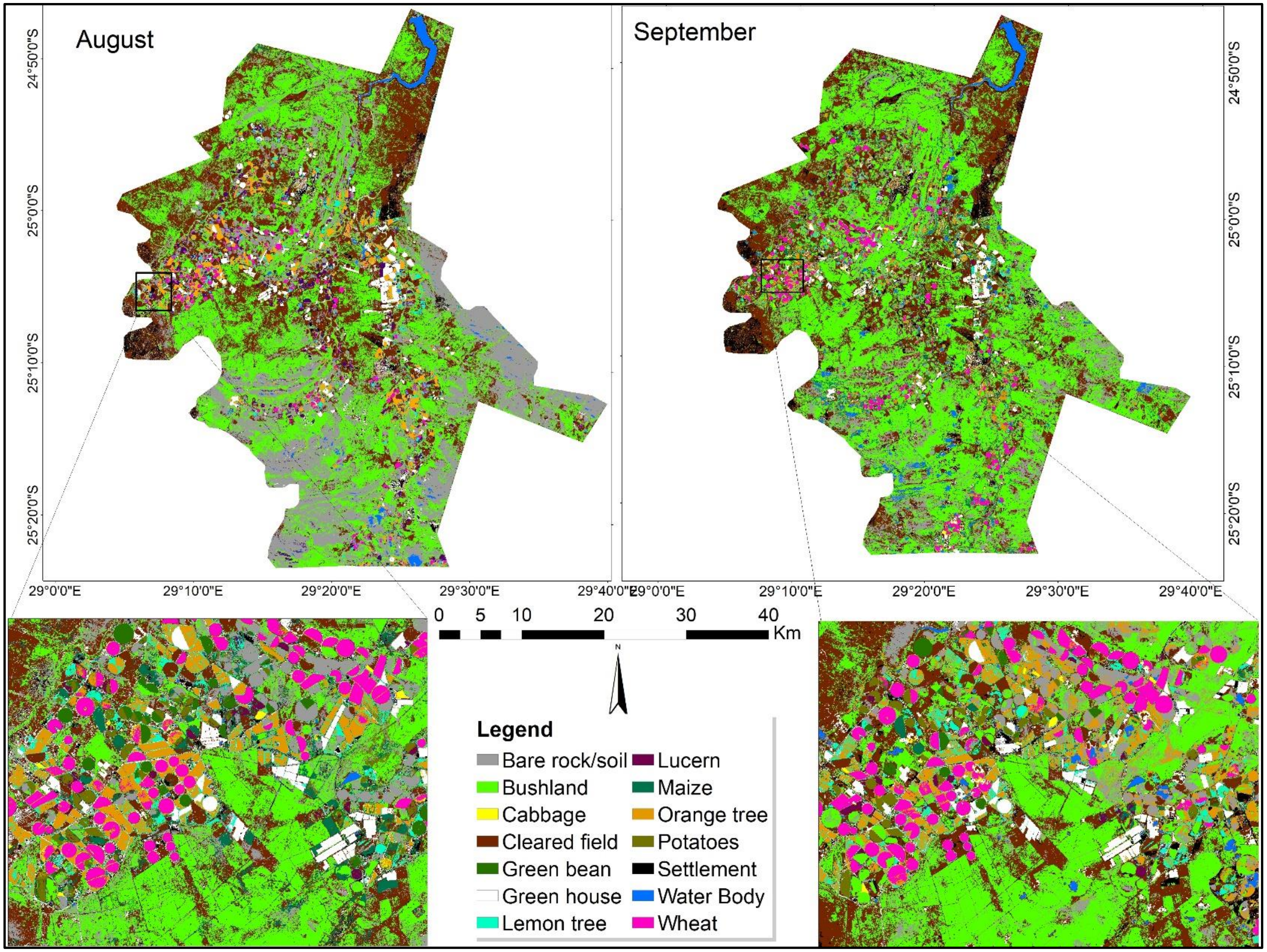


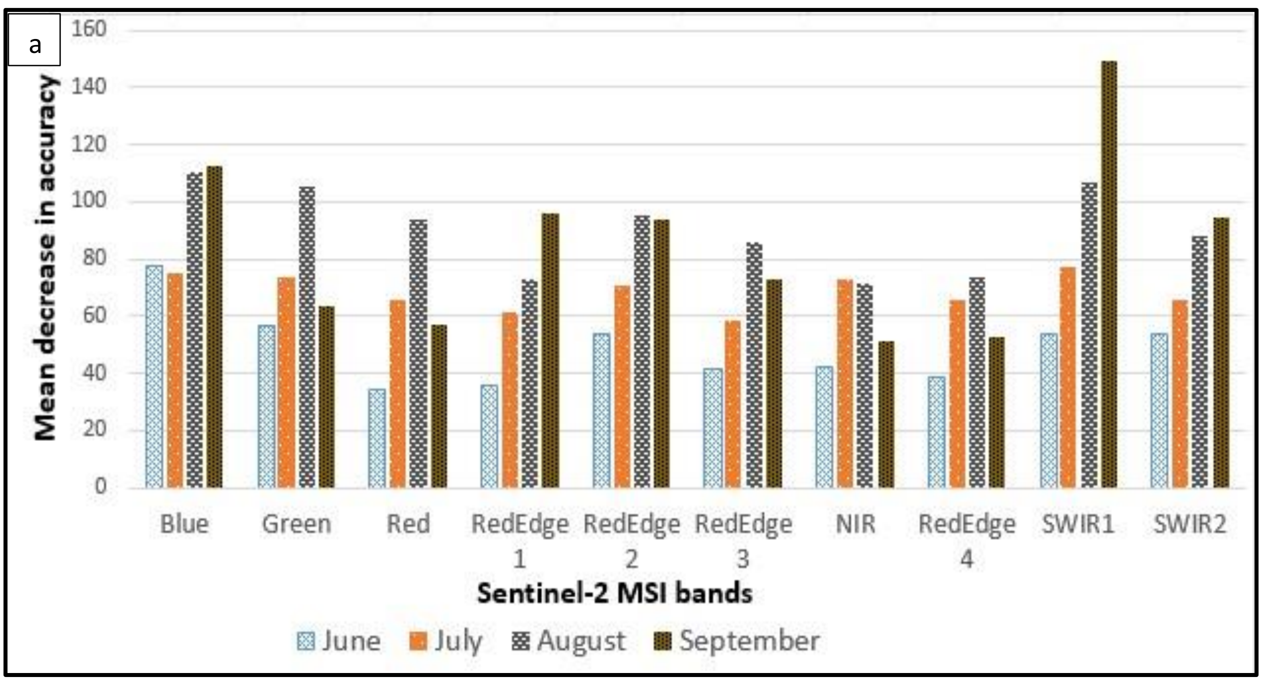
Figure 4.10: August and September classification results

4.3 Landsat-8 and Sentinel-2 bands importance and contribution in mapping the crops types

4.3.1 Variable importance for both sensors

RF was used to provide a variable importance measurement to indicate the role of each band for both sensors in the classification process. As part of the RF classification process, the RF algorithm assesses every variable's contribution to the agriculture landscape mapping by means of a specific measure (Figure 4.19). The variables' contribution in this study was determined by measuring the mean decrease in accuracy during the OOB error calculation (Table 4.1). This indicated the band importance for both Sentinel 2 (n=10) and Landsat 8 0LI (n=7) data in the classification. The variables with a large mean decrease in accuracy are more important for classifying the data. In classifying Sentinel-2 data, the most important bands were allocated in blue, SWIR1, SWIR2 and all the vegetation red edge bands (Figure 4.19a).

The measure clearly illustrates that NIR and panchromatic bands in Landsat 8-OLI (Figure 4.19b), are the most valuable bands for discrimination different crop types form other LULC classes while the blue is the least valuable.



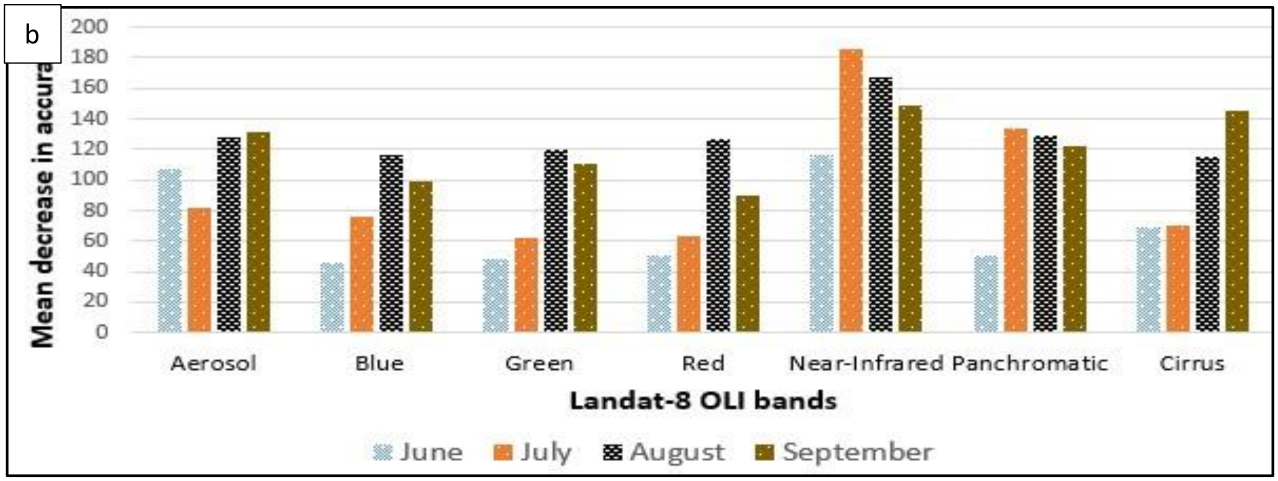
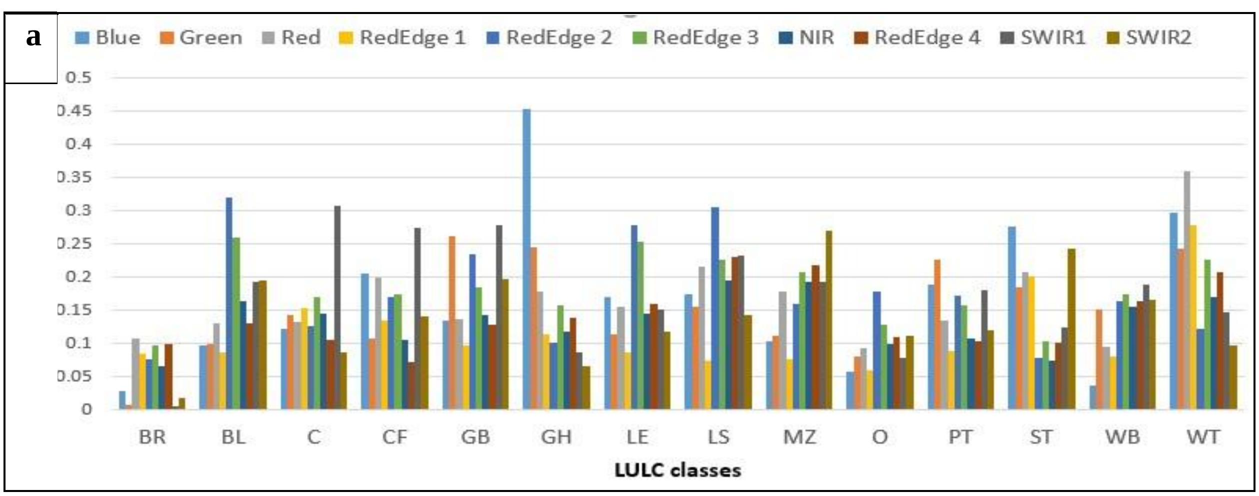


Figure 4.19: Band importance for S-2 MSI (a) and L-8 OLI (b): highest mean decrease in accuracy indicates the most important band.

The utility of each band was examined for each land cover class for both sensors for all the images (Figure 4.20). The high mean decrease accuracy indicates the most important band in that particular land use class. For Sentinel-2 MSI, the red edge bands were popularly important in classifying bushland, lemon, lucerne, potatoes and orange trees while the SWIR bands are the most contributing in classifying cabbage, bushland, cleared field, maize and settlement. The blue band was the least contributing for delineating maize, orange and water body whilst it was mostly contributing to mapping greenhouse and settlement. Further observations can be made from Figure 4.20b for Landsat-8 OLI band importance. The red band is the most contributing band for mapping wheat and had a moderate contribution for the rest of the crop types followed by the SWIR which was popular for cabbage, green beans, and water body. The least important band shown is panchromatic.



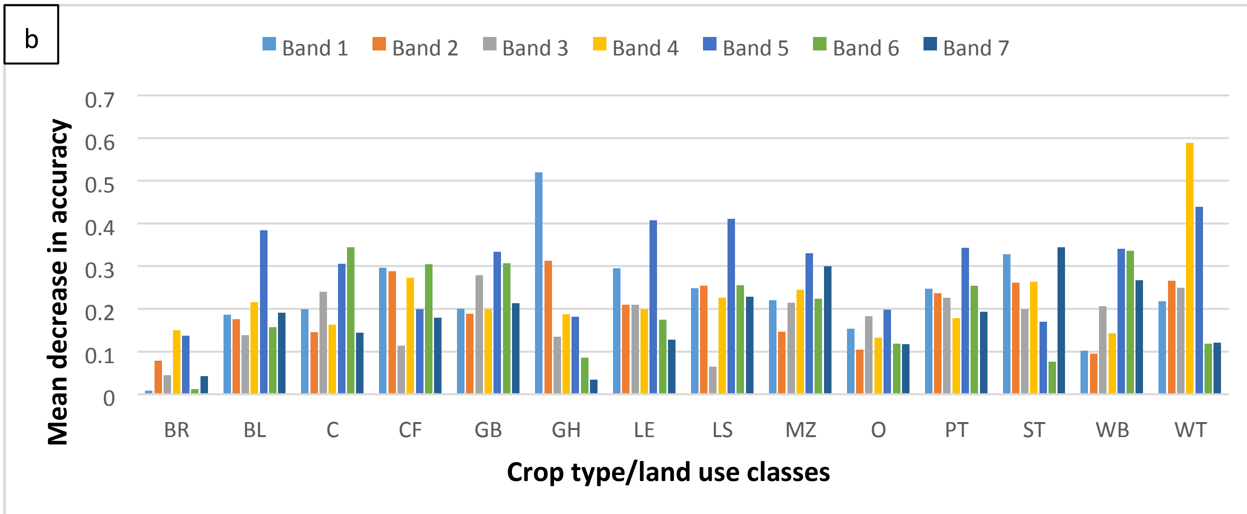
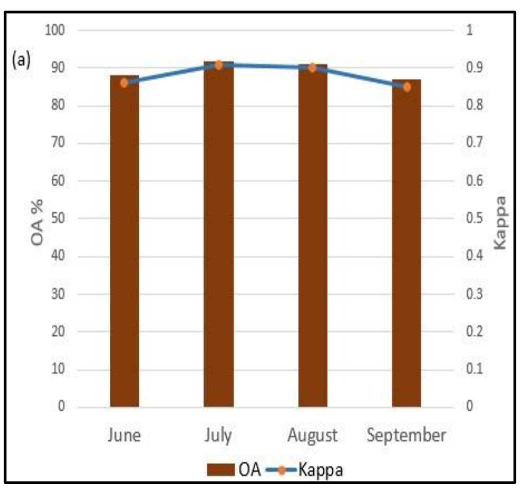


Figure 4.20: The importance of (a) Sentinel-2 and (b) Landsat-8 OLI bands in crop classification for the crop type/land use classes.

4.4 Accuracy assessment and statistical assessment

Accuracy assessment was performed for both sensors to evaluate the performance of Sentinel 2 and Landsat 8. Overall Accuracy (OA) and kappa index were calculated and comparison was done to see how the two sensors behaved during all four months using RF classification (Figure 4.21 a and b) and using SVM classification (Figure 4.22 a and b). As shown on the Figures below, July obtained the highest OA and Kappa value based on Landsat 8 data as compared to July Sentinel-2 data.



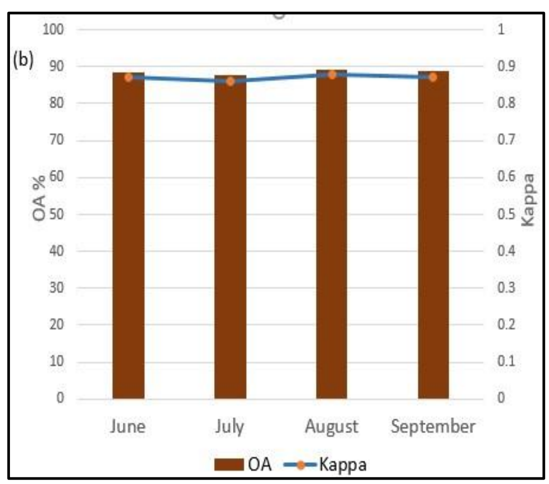


Figure 4.21: Graphs showing OA and Kappa index for RF classification using (a) Landsat-8 data (b) Sentinel-2 data.

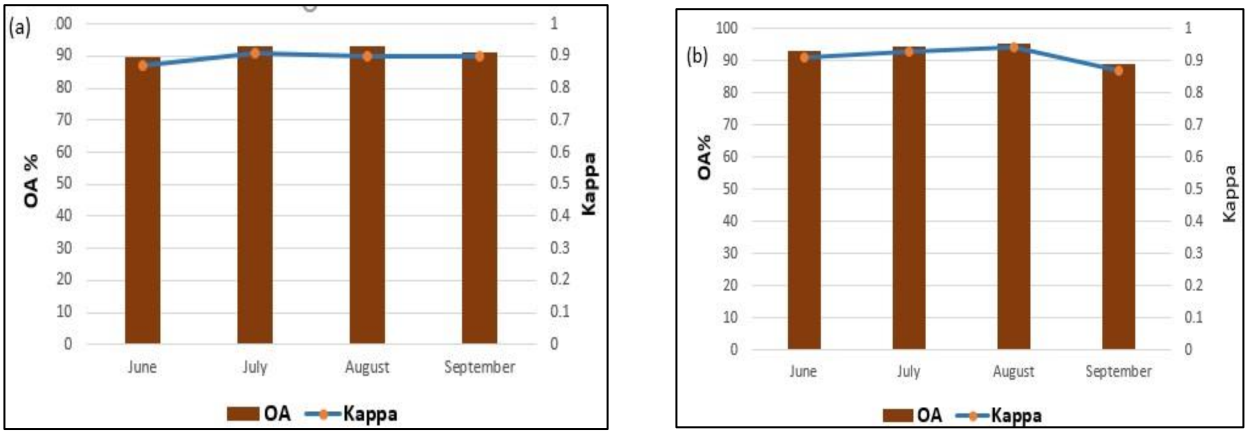


Figure 4.22: Graphs showing OA and Kappa index for SVM classification using (a) Landsat-8 data (b) Sentinel 2 data.

The efficiency and performance of both RF and SVM classifiers were evaluated using the test dataset. Table 4.4 and 4.5 present a summary of the confusion matrices obtained when the generated models were validated using the 30% test data based on L-8 OLI and S-2 MSI data. The RF classifier produced overall classification accuracy of 88%, 92%, 91%, and 87% for the months of June, July, August and September respectively and obtained kappa coefficient of 0.86, 0.91, 0.9 and 0.85, respectively. Furthermore, SVM produced OA of 90%, 93%, 93%, and 91% for June, July, August and September, while kappa coefficient is 0.87, 0.91, 0.9 and 0.9 for June, July, August and September respectively. Thus in general SVM outperformed RF in classifications based on the Landsat-8 data in all the months except for June.

Based on Sentinel-2 MSI data, the OA of RF for June, July, August and September Sentinel-2 data is 89%, 88%, 89%, and 89%, while the kappa coefficient is 0.87, 0.86, 0.88 and 0.87, respectively. Also, the OA of SVM for June, July, August and September Sentinel 2 data is 93%, 95%, 95%, and 89%, while kappa coefficient is 0.92, 0.93, 0.94 and 0.87, respectively. Overall, SVM performed better than RF in all classifications over all the four months using Sentinel-2 data.

Support vector machine algorithm performed better than RF for both sensors with the exception of June L-8 OLI and September S-2 MSI images. The rest achieved the highest overall accuracies of above 93% for both sensors with S-2 MSI for August being the highest (95%). The least (Kappa coefficient=0.85) was yielded from the September L-8 OLI imagery. The highest classification results were obtained from August S-2 data using SVM and the least classification results were obtained from the September L-8 data using RF. Therefore, RF is better when classifying Landsat8 data compared to Sentinel-2 data whilst SVM is better when classifying Sentinel-2 data compared to Landsat-8 data. The results show that Sentinel-2 performs better than Landsat-8 in classifying crop types in a heterogeneous agriculture landscape. Also, SVM is a better classifier as it efficiently classified crop types yielding the highest results as evidently shown in Table 4.4 and Table 4.5.

Generally, both RF and SVM had similar user's accuracy for orange trees (O), potatoes (P), maize (M) and bushland (BL) for both sensors (Table 4.4 and Table 4.5). The bare rock/soil showed the lowest producers' accuracy of 61.29% and user's accuracy of 61.54% for RF. 100% PA and UA were obtained from water body and settlement.

Table 4.4: Landsat-8 OLI data accuracy assessment.

	Random Forest								**Support Vector Machine**							
	June		**July**		**August**		**September**		**June**		**July**		**August**		**September**	
Class code	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**
BR	64.29	69.23	90.91	76.92	83.33	76.92	66.67	61.54	69.23	69.23	84.62	84.62	78.57	84.62	84.62	84.62
BL	84.00	87.50	88.89	100	95.83	95.83	80.77	87.50	81.48	91.67	85.19	95.83	82.14	95.83	84.62	91.67
C			85.71	100	95.83	95.83	100.00	90.48			91.67	91.67	100.00	95.83	100	100.00
CF	89.66	81.25	92.68	95	87.23	89.13	84.09	82.22	93.10	84.38	94.74	90.00	92.86	84.78	90.91	88.89
GB			84.21	94.12	100.00	91.67	93.02	97.56			94.29	97.06	97.96	100.00	97.62	100.00
GH	97.14	94.44	97.14	97.14	92.31	97.30	91.43	94.12	100.00	100.00	100.00	100.00	97.30	97.30	94.29	97.06
LE	95.65	88.00	95.65	84.62	90.00	87.10	77.42	77.42	88.00	88.00	88.89	92.31	77.78	90.32	73.68	90.32
LS	91.67	91.67	93.1	90	96.30	86.67	100.00	86.67	95.83	95.83	96.55	93.33	100.00	90.00	100	90.00
MZ					85.71	97.30	94.44	97.14					97.30	97.30	100	94.29
O	90.48	90.48	94.44	80.95	82.35	66.67	55.56	47.62	86.96	95.24	88.89	76.19	88.89	76.19	66.67	57.14
PT	75.00	75.00	63.64	70	78.57	91.67	87.50	90.32	66.67	50.00	69.23	90.00	91.67	91.67	87.1	87.10
ST	75.00	90.00	87.5	77.78	100.00	70.00	72.73	80.00	81.82	90.00	100.00	100.00	90.00	90.00	100	80.00

WB	100.00	91.67	100	95.83	100.00	100.00	96.00	100.00	100.00	91.67	100.00	95.83	100.00	100.00	100	100.00
WT	73.68	93.33	100	100	94.29	100.00	89.29	96.15	86.67	86.67	100.00	96.97	96.97	96.97	96.15	96.15
OA	**88.1579**		**91.9614 0.9116**		**91.0628**		**87.0466**		**89.91**		**93.25 0.91**		**93.00**		**91.19**	
Kappa	**0.87**				**0.9**		**0.85**		**0.87**				**0.90**		**0.85**	

Table 4.5: Sentinel-2 MSI data accuracy assessment.

	Random Forest								**Support Vector Machine**							
	June		**July**		**August**		**September**		**June**		**July**		**August**		**September**	
Class code	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**
BR	70.00	53.85	66.67	52.63	62.50	38.46 87.50 83.33	60.00	25.00	100.00	76.92	77.78	53.84615	69.23	69.23077	60.00	25.00
BL	74.19	95.83	80.00	66.67	77.78		66.67	91.67	76.67	95.83	82.14	95.83333	91.67	91.66667	66.67	91.66
C			100.00	100.00	100.00		94.74	85.71			100.00	100.00	100.00	95.83333	94.74	85.71
CF	89.29	78.13	73.58	97.50	78.43	86.96	90.00	80.00	96.30	81.25	90.24	92.50	93.62	95.65217	90.00	80.00
GB			86.84	97.06	90.20	95.83	90.24	90.24			100.00		97.96	0.00	90.24	90.24
GH	96.97	88.89	91.67	94.29	91.43	86.49	90.91	88.24	100.00	97.22	97.22	100.00	94.59	94.60	90.91	88.23
LE	86.21	100.00	95.83	88.46	96.77	96.77	70.27	83.87	92.00	9.00	92.31	92.31	96.77	96.77	70.27	83.87
LS	96.00	100.00	100.00	90.00	96.67	96.67	84.85	93.33	100.00	100.00	100.00	100.00	100.00	100.00	84.85	93.33
MZ					92.31	97.30	82.93	97.14					92.31	97.29	82.93	97.14
O	95.00	90.48	100.00	71.43	100.00	76.19	52.94	42.85	100.00	95.23	90	85.71	95.24	95.20	52.94	42.85
PT	100.00	25.00	69.23	90.00	78.95	83.33	88.00	70.96	60.00	75.00	100.00	100.00	97.22	97.20	88.00	70.96
ST	81.82	90.00	100.00	44.50	70.00	70.00	100.00	100.00	83.3333	100.00	100.00	77.00	87.50	70.00	100.00	100.00
WB	92.31	100.00	100.00	100.00	100.00	100.00	96.00	100.00	96.00	100.00	100.00	100.00	100.00	100.00	96.00	100.00
WT	92.86	86.67	91.67	100.00	97.06	100.00	92.31	92.31	93.33	93.00	97.06	100.00	100.00	96.96	92.31	92.00
OA	**88.60**		**87.70**		**89.13 0.9**		**89.00**		**92.98**		**94.53 0.93**		**95.41 0.95**		**89.00**	
Kappa	**0.87**		**0.86**				**0.82**		**0.92**						**0.82**	

4.5 Comparison between Landsat-8 OLI and Sentinel-2 MSI data in mapping crops types

The performance of Sentinel-2 MSI and Landsat-8 OLI sensors on classifying crop types and the co-existing land use was assessed. Table 4.6 shows the overall accuracy and kappa values for each sensor in all the months. Both sensors performed fairly well in all the months having OA of more than 87 %. The August 2017 Sentinel-2 MSI data produced superior classification results of 95.41% OA and kappa coefficient of 0.94 whilst the lowest results were obtained from September 2017 Landsat-8 OLI data obtaining OA of 87% and kappa coefficient of 0.85 with the exception of July and August which produced better results compared to S-2 MSI based data. Figure 4.17 shows the best classification map produced using Sentinel-2 MSI imagery whilst Figure 4.14 shows the lowest classification map produced by Landsat-8 OLI imagery.

Table 4.6: OA and *kappa coefficient* based on Landsat-8 OLI and Sentinel-2 MSI data.

		Overall Accuracy (%)		**Kappa coefficient**	
Classification		RF	SVM	RF	SVM
Landsat-8 OLI	June	88.15	89.91	0.86	0.87
	July	91.96	93.25	0.91	0.91
	August	91.06	93.0	0.9	0.9
	September	87.0	91.19	0.85	0.90
Sentinel-2 MSI	June	88.60	92.98	0.87	0.92
	July	87.70	94.53	0.86	0.93
	August	89.13	95.41	0.88	0.94
	September	89.0	89.0	0.87	0.87

CHAPTER FIVE

Discussion and Conclusion

4.1 The relevance of the approach

The use of multispectral data for crop type mapping has received considerable attention over the recent decades (Conrad *et al.*, 2010, Immitzer *et al.*, 2016). However, concerns over the accuracy levels in classifying heterogeneous landscapes coupled with high costs and availability of appropriate remote sensed data capable of discriminating the similar spectral signatures of crop types continue to be an impediment to regional agriculture mapping. Medium spatial resolution multispectral data of 30m or greater has been often used for crop mapping (Conrad *et al.*, 2010, Beeresh *et al.*, 2014, Li *et al.*, 2014), although some scholars report that 30 m resolution data is not adequate for mapping highly heterogeneous intensive agriculture landscape as it could lead to numerous mixed parcel, merging of borders of agriculture field and low precision (Conrad *et al.*, 2010, Ouzemou *et al.*, 2018). On the other hand, access and use of fine spatial resolution data of less than 10 m is a challenge due to expensive cost, unavailability, limited processing knowledge and the high dimensionality of data involved (Mutanga *et al.*, 2012). Against this background, the present study sought to utilize the opportunities created by USGS through free access of the recently launched Landsat-8 and Sentinel-2 missions to evaluate their performance in discriminating heterogeneous agriculture landscape with the aid of machine learning algorithms.

Previous studies who have used this data have augmented their experiments by using vegetation indices and phenological information from long term field survey (Palchowdhuri and King, 2018, Asgarian *et al.*, 2016) and with the use of pixel-based classification (Immitzer *et al.*, 2016, Vuolo *et al.*, 2018), and object-based image analysis (Schultz *et al.*, 2015). This study, however, sought to demonstrate that crop types can also be accurately mapped without using vegetation indices like the NDVI and phenology concept. The results of this study showed that in places where ground data is unavailable, multispectral data can be used to digitize the location of the crop types. The crop type mapping can thus be done using spectral reflectance values from the corresponding GPS locations of the digitized points. Also, a supervised model can be trained to classify the maps .

The outputs derived from the above methodology can be grouped as: thematic maps derived from
Sentinel-2 MSI and Landsat-8 imagery classification, variable importance for the 2 imagery used and accuracy assessment evaluated for both algorithms to determine the best method, bands, time and sensor to use when classifying crop type in a heterogeneous landscape.

From the classified maps, it is evident that the area has been classified into 11 classes for June, 13 classes for July and 14 classes for August and September and these classes were accurately distinguished using both algorithms. Both Landsat-8OLI and Sentinel-2 MSI obtained high classification accuracies showing the considerable capability of the moderate spatial and spectral resolutions of the data in accurately identifying and distinguishing the different crop types and LULC classes (Table 4.4 and Table 4.5). These results agree with the findings of Schultz *et al.* (2015), who managed to classify crop types using Landsat-8 data achieving an overall accuracy of 85.8%. Asgarian *et al.* (2016), also explored the use of Landsat-8 data in mapping crop type in a vastly heterogeneous and fragmented agriculture landscape in Iran. The study stresses that the heterogeneous patterns of agriculture landscapes present a challenge for accurate crop identification and mapping using the medium spatial resolution data. Nevertheless, the study concluded that Landsat-8 was capable of classifying crop types giving impressive accuracies of 91.56% for spring crops and 92.36% for summer crops.

In the present study, Sentinel-2 MSI imagery achieved high classification results compared to Landsat-8 OLI hence it is a better option to use in mapping crop types. Also, Vuolo *et al.* (2018), successfully classified nine crop types using Sentinel-2 data obtaining OA of above. The outcomes of the presents study are comparable to the findings obtained by Immitzer *et al.* (2016). They examined the use of multi-spectral Sentinel-2 imagery for classifying six crop types over the heterogeneous landscape and achieved overall accuracies of above 76%.

Regarding the use of multi-temporal data, July and August achieved the highest results in all four months. This may be attributed to the fact that this is when ground data was obtained and also that July is the peak time for development of the crops such as wheat, green beans, and orange trees . Vyas *et al.* (2005) pointed out that the use of multitemporal images resulted in higher accuracies compared to single date images. Murthy *et al.* (2003), concurs with the notion that the application of multi-date or multi-temporal data has the ability to discriminate crop types and they successfully managed to separate wheat fields in India. They realized that single date classification is constrained by the problem of spectral overlap hence they employed multi-date data and successfully managed to delineate wheat fields from another crop in India.

The study tested two familiar machine learning- RF and SVM algorithms in crop identification enabling the constraint of accurate crop type classifications. The methods used have already been proven to be efficient over heterogeneous agriculture landscapes (Ouzemou *et al.*, 2018) and has been applied to various multi-spectral data sets (Conrad *et al.*, 2010). The classification results demonstrated that both RF and SVM algorithms are valuable in mapping and understanding complex agriculture landscapes as both classifiers produced equally high overall accuracies (Table 4.6) as also noted by Nitze *et al.*, (2012). Adam *et al.*(2014) found similar results when classifying complex and heterogeneous coastal landscape. The importance of each of the bands in each sensor on the classification output was successfully determined in the study by utilizing RF's variable importance. Congalton and Green (2008), postulates that this is an important contribution for resolving classification errors which are often associated with utilizing multi-spectral imagery. The study demonstrated the value of each band in each sensor in enhancing the accuracy of crop type classification classes (Figure 4.19 and Figure 4.20). The study confirmed the high contribution of the red edge and shortwave infrared (SWIR) bands in crop type mapping as postulated by Mutanga *et al.* (2012) in vegetation mapping. This observation is consistent with results by Immitzer *et al.* (2016), who also observed the importance of these bands in Sentinel-2 data for crop mapping.

Regarding the performance of the classifier, it was observed that SVM produced slightly higher classification accuracy than RF when applied to both Sentinel-2 and Landsat-8 data. The present study found that nonlinear kernel function is efficient in SVM classification. The method was also deemed efficient for solving inseparability problems that may be found in land use classes in a heterogeneous landscape (Adam *et al.*, 2014). This result agrees with the findings of Löw *et al.* (2013), who examined the influence of feature space size on SVM and RF performance for fieldbased crop classification using multispectral RapidEye data. In that work, SVM presented a slightly superior performance with an overall accuracy of 94.6 % compared to an overall accuracy of 94.3 % for RF classification. The study emphasized that SVM is a computationally effective and accurate means for agricultural crop identification compared to RF, Also, that SVM is recommended where training data points are few.

4.2 Conclusion

There is a need to provide useable, timely and accurate data about the agriculture land environment to enhance adequate agriculture planning and land-use management. To meet this endeavor, the present study assessed the capability of the relatively new and affordable optical sensors, with advanced machine learning classifiers-RF and SVM to identify and discriminate crop types from the co-existing land use types in a highly heterogeneous agricultural environment. The outcomes of this research provide new insights into the performance of Sentinel-2 and Landsat-8 for classifying crop types for a complex agriculture landscape in an African region. Sentinel-2 data achieved superior results to Landsat-8 data classification showing that the vegetation red-edge bands are noteworthy thus agreeing with other researches that also indicated the importance of these spectral bands in vegetation classification.

With respect to evaluated classifiers, comparison of the efficiency of the RF and SVM demonstrated the similarity of RF and SVM although SVM came out as the superior classification algorithm in the present study. SVM proved to be flexible as the same parameters were used to tune the classification model in all experiments. The utilization of RF algorithm allowed for rigorous evaluation of the band importance for classification of crop types and it was easy to manipulate because the optimized parameters were minimal.

In conclusion, this research showed that the adopted approaches of using machine learning algorithms for crop type mapping were promising. The findings produced in this research are relevant not only for crop monitoring and planning for food security management but also for subsidizing actions for precision agriculture for the communal farmers in South Africa. For future research, there is a need to expand the study area so that crop type mapping can be done at a national level. Also, the confirmed methods can be employed over the whole crop growing from planting to harvest season. The alternative approaches such as diffusion of optical data and radar data as well as employing object-based classification in a heterogeneous landscape should further be explored.

4.3 Recommendations

The high classification accuracy attained in this research gives reliable information on the multispectral data that can be used for agriculture crop type mapping and the algorithms that can be used to show the bands contribution in the classification hence giving information that can be used for crop assessment. Landsat-8 and Sentinel-2 sensors offer affordable systematic coverage of the earth surface that is capable of giving regional agriculture monitoring due to its reasonable resolutions. However, the interpretation of these results can only be regarded as preliminary, since further research is needed to further expand the use of multi-spectral imagery on mapping crop types in a heterogeneous landscape. There is a need for further research to develop a technique capable of accurately analyzing and discriminating the different crop types found in small agriculture fields in the African landscape. Furthermore, the use of high spatial resolutions data results in misclassification of the crop types that have the same high spectral characteristics, especially within small fields. Therefore, further investigation is needed by employing other classification approaches like object-based classification technique. Additionally, the study recommends the use of diffusion data of multispectral data and radar data such as Sentinel-1 for crop type mapping.

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**,** 3440-3458.

Asgarian, A., Soffianian, A. & Pourmanafi, S. 2016. Crop type mapping in a highly fragmented and heterogeneous agricultural landscape: A case of central Iran using multi-temporal Landsat 8 imagery. *Computers and Electronics in Agriculture,* 127**,** 531-540.

Beeresh, H., Latha, M., Thimmaraja Yadava, G. & Dandur, N. 2014. An Approach for Identification and Classification of Crops using Multispectral Images. *International Journal of Engineering,* 3.

Blein, R., Bwalya, M., Chimatiro, S., Faivre-Dupaigre, B., Kisira, S., Leturque, H. & WamboYamdjeu, A. 2013. African agriculture, transformation, and outlook. *Johannesburg: New Partnership for African Development*.

Breiman, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

Breiman, L. 2002. Manual on setting up, using, and understanding random forests v3. 1. 2002. *URL: http://oz. berkeley. edu/users/breiman/Using_random_forests_V3,* 1.

Burges, C. J. 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery,* 2**,** 121-167.

Cai, X., Magidi, J., Nhamo, L. & van Koppen, B. 2017. *Mapping irrigated areas in the Limpopo Province, South Africa*, International Water Management Institute (IWMI).

Congalton, R. G. & Green, K. 2008. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, (Mapping Science). CRC Press: Boca Raton, FL, USA.

Conrad, C., Fritsch, S., Zeidler, J., Rücker, G. & Dech, S. 2010. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sensing,* 2**,** 10351056.

Dhumal, R. K., YogeshRajendra, K. & Mehrotra, S. 2013. Classification of Crops from remotely sensed Images: An overview. *International Journal of Engineering Research and Applications (IJERA),* 3**,** 758-761.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P. & Martimort, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment,* 120**,** 25-36.

El Hajj, M., Bégué, A., Guillaume, S. & Martiné, J.-F. 2009. Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices—The case of sugarcane harvest on Reunion Island. *Remote Sensing of Environment,* 113**,** 2052-2061.

Eriksson, P. & Reczko, B. 1995. The sedimentary and tectonic setting of the Transvaal Supergroup floor rocks to the Bushveld Complex. *Journal of African Earth Sciences,* 21**,** 487-504.

Foody, G. & Arora, M. 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing,* 18**,** 799-810.

Foody, G. M. & Mathur, A. 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and remote sensing,* 42**,** 13351343.

Forkuor, G., Dimobe, K., Serme, I. & Tondoh, J. E. 2018. Landsat-8 vs. Sentinel-2: examining the added value of Sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience & Remote Sensing,* 55**,** 331-354.

Graesser, J. & Ramankutty, N. 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sensing of Environment,* 201**,** 165-180.

Greyling, J. C. 2012. *The role of the agricultural sector in the South African economy.* Stellenbosch: Stellenbosch University.

Hornik, K., Meyer, D. & Karatzoglou, A. 2006. Support vector machines in R. *Journal of statistical software,* 15**,** 1-28.

Huang, C., Davis, L. & Townshend, J. 2002. An assessment of support vector machines for land cover classification. *International Journal of remote sensing,* 23**,** 725-749.

Immitzer, M., Vuolo, F. & Atzberger, C. 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing,* 8**,** 166.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S. & Defourny, P. 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing,* 7**,** 12356-12379.

Irons, J. R., Dwyer, J. L. & Barsi, J. A. 2012. The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment,* 122**,** 11-21.

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**,** 941-951.

Kumar, P., Prasad, R., Choudhary, A., Mishra, V. N., Gupta, D. K. & Srivastava, P. K. 2017. A statistical significance of differences in classification accuracy of crop types using different classification algorithms. *Geocarto International,* 32**,** 206-224.

Leslie, C. R., Serbina, L. O. & Miller, H. M. 2017. Landsat and agriculture—Case studies on the uses and benefits of Landsat imagery in agricultural monitoring and production. US Geological Survey.

Li, L., Friedl, M. A., Xin, Q., Gray, J., Pan, Y. & Frolking, S. 2014. Mapping crop cycles in China using MODIS-EVI time series. *Remote Sensing,* 6**,** 2473-2493.

Löw, F. 2013. Agricultural crop mapping from multi-scale remote sensing data-Concepts and applications in heterogeneous Middle Asian agricultural landscapes.

Löw, F., Michel, U., Dech, S. & Conrad, C. 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS Journal of photogrammetry and remote sensing,* 85**,** 102-119.

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.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R. & Berger, M. 2012. Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment,* 120**,** 91-101.

Maponya, P. & Mpandeli, S. 2012. Climate change and agricultural production in South Africa: Impacts and adaptation options. *Journal of Agricultural Science,* 4**,** 48.

Maponya, P. I. 2012. *Climate change and agricultural production in Limpopo Province: impacts and adaptation options.*

Mountrakis, G., Im, J. and Ogole, C. 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66**,** 247-259.

Muller-Wilm, U., Louis, J., Richter, R., Gascon, F. and Niezette, M., 2013, September. Sentinel2 level 2A prototype processor: Architecture, algorithms and first results. In *Proceedings of the ESA Living Planet Symposium, Edinburgh, UK* (pp. 9-13).

Murthy, C., Raju, P. & Badrinath, K. 2003. Classification of the wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks. *International Journal of Remote Sensing*, 24**,** 4871-4890.

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.

Nitze, I., Barrett, B. & Cawkwell, F. 2015. Temporal optimization of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation,* 34**,** 136-146.

Nitze, I., Schulthess, U. and Asche, H. 2012. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to the maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA***,** 7-9.

Odenweller, J.B., and Johnson, K.I., 1984. Crop identification using Landsat temporal-spectral profiles. *Remote Sensing of Environment*, *14*(1-3), pp.39-54.

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

Oommen, T., Misra, D., Twarakavi, N. K., Prakash, A., Sahoo, B. & Bandopadhyay, S. 2008. An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences,* 40**,** 409-424.

Ouzemou, J.-E., El Harti, A., Lhissou, R., El Moujahid, A., Bouch, N., El Ouazzani, R., Bachaoui, E. M. & El Ghmari, A. 2018. Crop type mapping from pansharpened Landsat 8 NDVI data: A case of a highly fragmented and intensive agricultural system. *Remote Sensing Applications: Society and Environment,* 11**,** 94-103.

Ozdogan, M., Yang, Y., Allez, G. & Cervantes, C. 2010. Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote sensing,* 2**,** 2274-2304.

Pahlevan, N. & Schott, J. R. 2013. Leveraging EO-1 to evaluate the capability of a new generation of Landsat sensors for coastal/inland water studies. *IEEE Journal of selected topics in applied earth observations and remote sensing,* 6**,** 360-374.

Pal, M. & Mather, P. 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing,* 26**,** 1007-1011.

Petropoulos, G. P., Kontoes, C. & Keramitsoglou, I. 2011. Burnt area delineation from a unitemporal perspective based on Landsat TM imagery classification using support vector machines. *International Journal of Applied Earth Observation and Geoinformation,* 13**,** 70-80.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J. P. 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.

Schmedtmann, J. & Campagnolo, M. L. 2015. Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control. *Remote Sensing,* 7**,** 93259346.

Schultz, B., Immitzer, M., Formaggio, A., Sanches, I., Luiz, A. & Atzberger, C. 2015. Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in southeastern Brazil. *Remote Sensing,* 7**,** 14482-14508.

Sibanda, M., Mutanga, O. & Rouget, M. 2016. Discriminating rangeland management practices using simulated hyspIRI, Landsat 8 OLI, sentinel 2 MSI, and VENµs spectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 9**,** 39573969.

Soman, K.P., Loganathan, R. and Ajay, V., 2009. *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd..

Sothe, C., Almeida, C. M. d., Liesenberg, V. & Schimalski, M. B. 2017. Evaluating sentinel-2 and Landsat-8 data to map sucessional forest stages in a subtropical forest in Southern Brazil.
Remote Sensing, 9**,** 838.

Unies, N. 1999. World Population Prospects: The 1998 Revision, vol. 1, Comprehensive Tables. *Numéro de vente: E*, 99.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. & Ceschia, E. 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199**,** 415-426.

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C. & Ng, W.-T. 2018. How much does multitemporal Sentinel-2 data improve crop type classification? *International journal of applied earth observation and geoinformation,* 72**,** 122-130.

Vyas, S., Oza, M. & Dadhwal, V. 2005. Multi-crop separability study of Rabi crops using multitemporal satellite data. *Journal of the Indian Society of Remote Sensing,* 33**,** 75.

Waals, J. H. van der (2013) *AGRICULTURAL POTENTIAL SURVEY :* PORTION 1, LOSKOP SUID 53-JS, LIMPOPO PROVINCE. Available at: [https://sahris.sahra.org.za/sites/default/files/additionaldocs/Groblersdal Draft Basic Assessment Part 2.pdf.]. Accessed on 24 February 2019.

Yang, C., Everitt, J. H. & Murden, D. 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture,* 75**,** 347-354.

Zhang, N., Wang, M. and Wang, N. (2002) 'Precision agriculture * a worldwide o v er v iew', 36, pp. 113–132.

Zhu, Z. & Woodcock, C. E. 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment,* 144**,** 152-171.

DAFF 2018. Economic Review of the South African Agriculture 2017/18. South Africa: Department of Agriculture, Forestry and Fisheries.