Zentrum für Entwicklungsforschung

Challenges and opportunities of using ecological and remote sensing variables for crop pest and disease mapping

Dissertation

Zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der

Landwirtschaftlichen Fakultät

Rheinischen Friedrichs-Wilhelm-Universität Bonn

von

Richard Kyalo

aus

Machakos, Kenya

Bonn 2019

Referent: Prof. Dr. Christian Borgemeister
Koreferrent: Prof. Dr. Klaus Greve
Tag der Promotion: 10. Juli 2019
Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

ABSTRACT

Crop pest and diseases are responsible for major economic losses in the agricultural systems in Africa resulting in food insecurity. Potential yield losses for major crops across Africa are mainly caused by pests and diseases. Total losses have been estimated at 70% with approximately 30% caused by inefficient crop protection practices. With newly emerging crop pests and disease, monitoring plant health and detecting pathogens early is essential to reduce disease spread and to facilitate effective management practices. While many pest and diseases can be acquired from another host or via the environment, the majority are transmitted by biological vectors. Thus, vector ecology can serve an indirect explanation of disease cycles, outbreaks, and prevalence. Hence, better understanding of the vector niche and the dependence of pest and disease processes on their specific spatial and ecological contexts is therefore required for better management and control.

While research in disease ecology has revealed important life history of hosts with the surrounding environment, other aspects need to be explored to better understand vector transmission and control strategies. For instance, choosing appropriate farming practices have proved to be an alternative to the use of synthetic pesticides. For instance, intercropping can serve as a buffer against the spread of plant pests and pathogens by attracting pests away from their host plant and also increasing the distance between plants of the same species, making it more exigent for the pest to target the main crop. Many studies have explored the potential applications of geospatial technology in disease ecology. However, pest and disease mapping in crops is rather crudely done thus far, using Spatial Distribution Models (SDM) on a regional scale.

Previous research has explored climatic data to model habitat suitability and the distribution of different crop pests and diseases. However, there are limitation to using climate data since it ignores the dispersal and competition from other factors which determines the distribution of vectors transmitting the disease, thus resulting in model over prediction. For instance, vegetation patterns and heterogeneity at the landscape level has been identified to play a key role in influencing the vector-host-pathogen transmission, including vector distribution, abundance and diversity at large. Such variables can be extracted from remote sensing dataset with high accuracy over a large extent. The use of remotely sensed variables in modeling crop pest and disease has proved to increase the accuracy and precision of the models by reducing over fitting as compared to when only climatic data which are interpolated over large areas thus disregarding landscape heterogeneity.When used, remotely sensed predictors may capture subtle variances in the vegetation characteristic or in the phenology linked with the niche of the vector transmitting the disease which cannot be explained by climatic variables. Subsequently, the full potential of remote sensing applications to detect changes in habitat condition of species remains uncharted. This study aims at exploring the potential behind developing a framework which integrates both ecological and remotely sensed dataset with a robust mapping/modelling approach with aim of developing an integrated pest management approach for pest and disease affecting both annual and perrennial crops and whom currently there is no cure or existing germplasm to control further spread across sub Saharan Africa.

Herausforderungen und Möglichkeiten der Verwendung von ökologischen und Fernerkundungsvariablen für die Schädlings- und Krankheitskartierung

KURZFASSUNG

Pflanzenschädlinge und Krankheiten in der Landwirtschaft sind für große wirtschaftliche Verluste in Afrika verantwortlich, die zu Ernährungsunsicherheit führen. Die Verluste werden auf 70% geschätzt, wobei etwa 30% auf ineffiziente Pflanzenschutzpraktiken zurückzuführen sind. Bei neu auftretenden Pflanzenschädlingen und Krankheiten ist die Überwachung des Pflanzenzustands und die frühzeitige Erkennung von Krankheitserregern unerlässlich, um die Ausbreitung von Krankheiten zu reduzieren und effektive Managementpraktiken zu erleichtern. Während viele Schädlinge und Krankheiten von einem anderen Wirt oder über die Umwelt erworben werden können, wird die Mehrheit durch biologische Vektoren übertragen. Daraus folgt, dass die Vektorökologie als indirekte Erklärung von Krankheitszyklen, Ausbrüchen und Prävalenz untersucht werden sollte. Um effektive Vektorkontrollmaßnahmen zu entwickeln ist ein besseres Verständnis der ökologischen Vektor-Nischen und der Abhängigkeit von Schädlings- und Krankheits-Prozessen von ihrem spezifischen räumlichen und ökologischen Kontext wichtig.

Während die Forschung in der Krankheitsökologie wichtige Lebenszyklen von Wirten mit der Umgebung schon gut aufgezeigt hat, müssen weitere Aspekte noch besser untersucht werden, um Vektorübertragungs- und Kontroll-Strategien zu entwickeln. So hat sich beispielsweise die Wahl geeigneter Anbaumethoden als Alternative zum Einsatz synthetischer Pestizide erwiesen. In einigen Fällen wurde der Zwischenfruchtanbau als ‚Puffer' gegen die Ausbreitung von Pflanzenschädlingen und Krankheitserregern vorgeschlagen. Bei diesem Anbausystem werden Schädlinge von ihrer Wirtspflanze abgezogen und auch der Abstand zwischen Pflanzen derselben Art vergrößert (was eine Übertragung erschwert).

Viele Studien haben bereits die Einsatzmöglichkeiten von Geodaten in der Krankheitsökologie untersucht. Die Kartierung von Schädlingen und Krankheiten in Nutzpflanzen ist jedoch bisher eher großskalig erfolgt, unter der Zunahme von sogenannten ‚Spatial Distribution Models (SDM)' auf regionaler Ebene. Etliche Studien haben diesbezüglich klimatische Daten verwendet, um die Eignung und Verteilung verschiedener Pflanzenschädlinge und Krankheiten zu modellieren. Es gibt jedoch Einschränkungen bei der Verwendung von Klimadaten, da dabei andere landschaftsbezogene Verbreitungs-Faktoren ignoriert werden, die die Verteilung der Vektoren und Krankheitserreger bestimmen, was zu einer Modell-Überprognose führt. Vegetationsmuster und Heterogenität auf Landschaftsebene beeinflussen maßgeblich die Diversität und Verteilung eines Vektors und spielen somit eine wichtige Rolle bei der Vektor-Wirt-Pathogen-Übertragung. Bei der Verwendung von Fernerkundungsdaten können subtile Abweichungen in der Vegetationscharakteristik oder in der Phänologie, die mit der Nische des Vektors verbunden sind, besser erfasst werden. Es besteht noch Forschungs-Bedarf hinsichtlich der Rolle von Fernerkundungsdaten bei der

Verbesserung von Artenmodellen, die zum Ziel haben den Lebensraum von Krankheitsvektoren besser zu erfassen. Ziel dieser Studie ist es, das Potenzial für die Entwicklung eines Rahmens zu untersuchen, der sowohl ökologische als auch aus der Ferne erfasste Daten mit einem robusten Mapping- / Modellierungsansatz kombiniert, um einen integrierten Ansatz zur Schädlingsbekämpfung für Schädlinge und Krankheiten zu entwickeln, der sowohl einjährige als auch mehrjährige Kulturpflanzen betrifft Keine Heilung oder vorhandenes Keimplasma zur weiteren Verbreitung in Afrika südlich der Sahara.

TABLE OF CONTENTS

1 GENERAL INTRODUCTION

1.1 Context

Agriculture is the main economic activity in sub-Saharan Africa (SSA), which supports more than 67% of the population. Development objectives for SSA, where rural population heavily depend on food production for their income, includes moving towards resource conservation and natural resource management while striving for higher agricultural production. This is aimed at increasing economic growth by 4-5 % annually to achieve food security and a modest standard of living provided for the 1.3 billion people expected in the region by 2025 (World Bank 1989). Much of the vulnerability of African agricultural systems lies in the fact that its agricultural systems remain largely rain-fed and are under-developed as the majority of African farmers are small scale farmers with limited financial resources (Benin et al. 2016). Thus, given the projected rise in population in Africa, less costly innovations are needed that mutually benefit improved farming practices that enhance crop productivity.

A wide range of naturally occurring biotic and abiotic constraints, including poor soil, water scarcity, inappropriate temperatures, crop pests, diseases and weeds are well known to reduce food crop productivity in Africa, leading to low input efficiencies, reduced crop output and ultimately food insecurity (Reynolds et al. 2015). Specifically, crop losses due to pests and disease accounts of up to more than 70 % in SSA. These values might extensively ascend under varying climatic conditions whereby new and more aggressive pests and diseases have been accounted for to influence the stability of crop yields thereby underming food security on the continent. For instance, the recent rapid spread of the voracious fall armyworm, *Spodoptera frugiperda* (J.E. Smith) (Lep.: Noctuidea), that has marched through Africa is causing a serious food crisis. However, it is not possible to make estimates on overall potential crop losses due to climate change. This is partly due to the unexpected adaptation of certain pests to a changing environment that creates uncertainties, thus increasing levels of unpredictability of spatial and temporal interaction between weather, cropping system and pest's abundance (Chakraborty and Newton 2011).

Mitigation measures demands for better understanding of pests and diseases habitats and dispersal pattern which is widely affected by a range of environmental variables at varying scales. For instance, at regional scale the suitability is majorly influenced by climate while at landscape level, climate suitability is modified by land use, land cover and topography. Suitability is further modified at local scales by vegetation and micro topography (Pearson and Dawson 2003). Morever, recent research has further linked pest and disease infestation rates to cropping patterns. A good example is the push-pull system technology which entails inter-cropping of repellent and attractant crops for insect pest control. The attractant crop draws the insect in (acts as the “pull”), whereas the repellent crop deters the insect (acts as the “push”). The push-pull system that was developed in Africa to protect maize from lepidopteran stemborers by planting a grass that is more attractive to the moths as a border and planting a repellent legume crop between the rows of maize (Khan et al. 1997).

In addition, inter-cropping has also been used as a buffer against the spread of plant pests and pathogens by increasing the distance between plants of the same species, making it more exigent for the pest to target the main crop (Nyasani et al. 2012; Songa et al. 2007). Thus, cropping systems are an important variable in control and management of pest and disease in Africa. Moreover, useful baseline information feed to monitor and understand seasonal cropping pattern changes as a function of climatic variability and climate change in Africa are needed. This information, especially if linked with yields and crop pest and disease infestation levels, may be very useful for better agricultural risk projections in future.

Besides, the land surface temperature (LST) is a well-known parameter that majorly influences pest and disease distribution because of their temperature dependence (Marinho et al. 2016; Pearsall and Myers 2001). Ground temperature is measured through point measurements that contribute to high spatial variability because the locations of ground data stations are generally very sparse in SSA. Contrary to this, the usage of remote sensing to retrieve LST can compensate spatial resolution of LST measurements derived through ground observations with less man-power and at lower cost. Several studies have shown the benefits of using remote sensing derived LST

in monitoring insect pests over the conventional use of temperature retrieved from ground weather stations which includes a higher spatial and temporal resolution and reasonable accuracies. A god example is MODIS LST which has high spatial characteristics to capture the spatial variability of temperature within a finer scale in comparison to ground weather station.

In Africa, the majority of insect pests are often monitored using direct field survey methods such as scouting and checking the plants for any damage symptoms. These techniques are essentially time-consuming, tiresome, costly and biased since only a few sites within the fields are sampled (Bock et al. 2010; Newton and Hackett 1994). Sometimes it happens that by the time ground survey is completed the damage area has already changed. Thus, complementary and synoptic pest monitoring procedures that allow the implementation of site-specific practices are required.

In this context, remote sensing is capable of offering synoptic, timely, accurate and relatively inexpensive data that can be utilized to provide explicit overviews of specific pest and disease distribution and damage or infestation level at field level. Also, remote sensing allows a wide-area insect pest and disease monitoring approach which is less spotty and spatially more effective and coherent compared to point-specific field-based survey methods (Paul et al. 2003; Abdel-Rahman et al. 2004). Further, state of the art remote sensing techniques can be utilized to derive vegetation phenology variables from remotely sensed data that offer fine scaled vegetation dynamics information for prediction of pests and diseases. This offers great opportunity of linking remotely sensed and bioclimatic variables to provide an operational system for monitoring long term changes related to the spread of pests and disease in relation to a landscape context (Makori et al. 2017).

Further, remote sensing offers great potential to carry out damage assessment promptly, with extensive accuracy and eliminating bias. If sensors are sufficiently sensitive to the spectral signature of a pest or disease, it may be possible to detect the crop damage at an early stage and monitor the same. Pests and diseases cause the damage to the crop in different ways. Insects with chewing and biting mouthparts often feed on leafy portion of plants, causing distinctive symptoms like defoliation. This

reduces the green biomass present on the plants. By contrast, insects with sucking mouthparts suck the cell sap from the leaves, causing shriveling and deformation of leaves resembling water stress damage. This characteristic damage may cause changes in reflectance and therefore be visible on satellite images. It should be possible to quantify damage with additional ground truth calibration. Once the damage is noticed it is possible to monitor the pest. In this way it would be possible to provide early warning information to areas where the pest is likely to advance. This helps to take preventive measures against the pest. Compared to ground surveys, remote sensing is cheap and availability of high-resolution satellite images in recent years has also helped to improve accuracy.

With escalating population pressure throughout the world and the need for increased agricultural production, this research sought to explore the opportunities and challanges of using ecological and remote sensing data for extracting significant constraints related to pest and disease ecology such as cropping systems which has not been explored before especially in heterogenious landscapes in Africa. In addition, the study further highlights the opportunies of improving pest and disease habitat suitability modelling with inclusion of vegetation productivity and landscape context perspective in relation to disease occurance and finally mapping the diease damage levels to ensure that timely mitigation strategies are put in place. The methodology was tested in Kenya for two diseases affecting both annual and perrenial crops for which there is presently no real resistance in the existing germplasm. These include maize lethal necrosis disease (MLN) affecting maize (annual crop) and the African Citrus greening disease (ACGD) greatly affecting citrus (perrenial crop) production in Africa and beyond. This research contributes to improved crop productivity by providing a robust modeling approach towards enhanced monitoring of ecological pattern, distribution and damage levels mapping of pests and diseases in both annual and perrenial crops.

1.2 Objectives of the study

The overall goal of this study is to explore the challenges and opportunities of using ecological and remote sensing-based variables for crop pest and disease mapping in both annual and perennial crops. The specific objectives of the study are:

1. To map maize cropping systems using RapidEye observations in highly fragmented agro-ecological landscapes in Kenya.
2. To establish an ecological niche modeling approach for predicting the spatial distribution of African citrus triozid, the main vector of ACGD in Africa.
3. To investigate the impact of cropping systems on maize lethal necrosis severity using multi-spectral high to moderate resolution satellite imagery.

1.3 Outline of the study

This thesis is divided into five chapters.

Chapter 1 provides a broad introduction, background and the rationale of the study. It also outlines the objectives and the expectations of the study.

Chapter 2 presents the results of mapping maize cropping systems using rapidEye observation in Kenya as one of the important variables in numerous crop productivity models.

Chapter 3 outlines how remotely sensed vegetation phenological metrics variables could potentially improve the habitat suitability mapping of the African citrus triozid (ACT) transmitting ACGD on citrus crops in an heterogenious landscape of Kenya. The chapter also compares the extracted habitat suitability of ACT with the surrounding landscape context based on dispersal characteristics of the pest to understand the role of land use, landcover in disease transmission and survival of the vector.

Chapter 4 explores the possibilities of mapping maize lethal necrosis disease damage levels by combining two acquisations of RapidEye imagery with resampled Landsat 8 using hierachial random forest classification algorithm.

Chapter 5 summarises and highlights the key finding of the research and links these to draw final conclusions on applicability of remote sensing variables understanding of pest and disease ecology for better and timely mitigation measures. Finally, important research gaps for future research areas are also suggested.

2 MAIZE CROPPING SYSTEMS MAPPING USING RAPIDEYE OBSERVATIONS IN AGRO-ECOLOGICAL LANDSCAPES IN KENYA

This chapter has been published as Kyalo et al. (2017) *Sensors* 17(11): E2537

2.1 Abstract

Cropping systems information on explicit scales is an important but rarely available variable in many crops modeling routines and of utmost importance for understanding pests and disease propagation mechanisms in agro-ecological landscapes. In this study, high spatial and temporal resolution RapidEye bio-temporal data were utilized within a novel 2-step hierarchical random forest (RF) classification approach to map areas of mono- and mixed maize cropping systems. A small-scale maize farming site in Machakos County, Kenya was used as a study site. Within the study site, field data was collected during the satellite acquisition period on general land cover/land use and the two cropping systems. Firstly, non-cropland areas were masked out from other land use/ land cover (LULC) using the LULC mapping result. Subsequently an optimized RF model was applied to the cropland layer to map the two cropping systems (2nd classification step). An overall accuracy of 93% was attained for the LULC classification While the class accuracies (PA: producer's accuracy, and UA: user's accuracy) for the two cropping systems were consistently above 85%. We concluded that explicit mapping of different cropping systems is feasible in complex and highly fragmented agro-ecological landscapes if high resolution and multi-temporal satellite data such as 5-meter RapidEye data is employed. Further research is needed on the feasibility of using freely available 10-20 m Sentinel-2 data for wide-area assessment of cropping systems as an important variable in numerous crop productivity models.

2.2 Introduction

Agro-ecological systems in Africa are particularly vulnerable to climate variability and climate change due to their over dependence on rainfall (Rockstrom 2003). The particular cropping system used by farmers is often a key determinant in climate-smart

agriculture concepts and crop diversification and livelihoods strategies (Mersha and Van Laerhoven 2016). While information about cropland extents or crop acreages are increasingly available and being used in food supply projections (Edgerton 2009), explicit information about the actual cropping systems is largely not utilized or available. This leads to significant uncertainties of crop production models and ultimately in food security projections for Africa (Antle et al. 2016).

The cropping system is defined as the planting sequence of crops applied to an agricultural area or field over a certain period. In agronomical terms, an agricultural field can be mono-cropped, inter-cropped, relay cropped, mixed-cropped or under crop rotation (i.e. planting different crops in sequential years) (Panigrahy et al. 2005). Mixed-cropping is a common practice on small scale farms in developing countries like Kenya. In Kenya, maize (*Zea mays L*.) is the staple food, and it is common to find it mixed with bean (*Phaseolus vulgaris L*.) (Callaghana et al. 1994). The degree of mixed cropping is often determined by the need for diversification against a backdrop of increased climate variability and the need to increase soil fertility and soil moisture regimes to sustain or increase crop productivity (Wang et al. 2014). In this study, we define mixed cropping specifically as maize grown in a spatial arrangement with other leguminous crops on the same field within the same growing season and mono-cropping as maize grown as a single crop within the same time frame and field.

High and medium spatial resolution satellite data have been widely used for agricultural land use mapping in different agro-ecological zones in Africa and beyond (Forkuor et al. 2015; Sibanda and Murwira 2011). However, many studies alluded to the challenges of accurately mapping crops and cropping systems in Africa on a landscape scale primarily due to the small scale and highly fragmented nature of cropping patterns as well as their intra- and inter-annual dynamics (Forkuor et al. 2014). The temporal and spatial highly variability of cropping systems is often a result of incoherent farmers decisions (the planting date often varies from one season or year to the next) and other localized and hard to quantify socio-economic factors (Woomer et al. 1997). Moreover, rain fed crops are largely indiscriminate from some natural vegetation communities such as grassland during the wet or growing season when both (the crops and some natural

vegetation types) have the same phenological growing cycles (Conrad et al. 2011a). Thus, landscape scale crop mapping mechanism using medium resolution data such as 30-meter Landsat, a frequently used type of dataset for crop mapping in Africa (Roy et al. 2010), has resulted in high spectral heterogeneity and poor mapping results (Husak et al. 2008). Essentially, remotely sensed data can provide spatially coherent information only on crop acreage and crop vitality on landscape scales with an advantage over traditional conventional surveying methods that are often tedious and costly (ineffective), especially if crop assessments are performed over larger areas (Conrad et al. 2011b).

Relatively newly available 5-meter RapidEye data are suitable for crop mapping in highly fragmented and dynamic landscapes because of the higher and enhanced geometrical resolution of the satellite system, particularly in small-scale farming systems in Africa where the size of field is relatively small (≤ 1.25 ha) (Cord et al. 2010; Forkuor et al. 2015; Forkuor et al. 2014a). The enhanced spectral resolution of RapidEye data in the red-edge waveband domain, for instance, allows for significantly enhanced land use classification and improved crop discrimination. This could be due to strong correlations between the vegetation spectral features at the red-edge band and chlorophyll content and also the sensitivity of the red-edge band to differences in leaf structure (Eitel et al. 2011; Schuster et al. 2012; Tigges et al. 2013). Combined with state-of-the-art and hierarchical classification approaches using robust machine learning classification algorithms and RapidEye data from different time steps, including their derived vegetation indices, explicit and permissible accurate crop type mapping results even in complex African landscapes can be generated (Forkuor et al. 2015; Forkuor et al. 2014a). Various types of non-parametric machine learning classification methods like random forest (RF) have been successfully applied to mapping crops in Africa (Conrad et al. 2011b; Forkuor et al. 2015).

Mulianga et al. (2015) characterized cropping practices (crop type and harvest mode) of sugarcane-based cropping system in Kenya using re-sampled 15-meter multi-temporal Landsat dataset and a maximum likelihood classifier. However, sugarcane is usually grown on large-scale commercial and homogeneous fields that can be easily

discriminated (Abdel-Rahman and Ahmed 2008). To the best of our knowledge, no study has yet attempted to map maize-cropping systems in heterogeneous landscapes in Africa and, moreover, no study is known that utilized RapidEye time-series data and machine learning classification approach in this regard. Accordingly, the main objective of the study was to examine the utility of random forest (RF) classifier and new-generation RapidEye imagery with enhanced waveband coverage in the red-edge spectral region for cropping systems mapping. Specifically, we aimed to develop a (semi-automated) processing scheme to find the optimal RF model parameters by analyzing the relative model contribution of the RapidEye spectral indices and individual waveband regions (bands) for crop systems mapping. Having information on the spatial distributions of maize systems would ultimately help to better understand factors that contribute to crop productivity and farm or field level yield variability (Ramert et al. 2002).

2.3 Study area

The study area is in Machakos County, about 100 km south-east of Nairobi in Kenya (Figure 2.1). The study area lies between the latitudes 1°17'53.71" S and 1°31'8.54" S and between the longitudes of 37°28'15.79" E and 37°40'33.43" E. The total study area covers about 677 km^2 with elevation ranging from 400 to 2100 m above mean sea level (MAMSL). The climate is semi-arid with a highly variable rainfall regime distributed over two rainy seasons, hence two cropping seasons namely the short rain season and the long rain season. Short rains occur from November to January, and long rains from March to June with an average rainfall ranging from 500 to 2000 mm (mean annual precipitation) and a mean annual maximum temperature of 28°C (Bryan et al. 2013).

The most widespread vegetation type in Machakos is semi-arid deciduous thicket and bush land, dominated by *Acacia* spp. (Fabaceae) and *Commiphora* spp. (Burseraceae). In drier locations below the elevation of 900 m, thorn bush grades into semi-desert vegetation. Moreover, arable land covers about 64% of the total landmass of the study area (Mwangi and Mundia 2014), with mixed cropping regularly practiced in this region (Woomer et al. 1997). The most prevalent crops in the region are maize,

bean, pigeon pea, and cowpea. Maize and bean in most cases are mixed in the long rainy season, while cowpea is mainly mixed with maize and bean in the short rainy season. Recent uncertainties in rainfall patterns have encouraged mixed cropping with majority of farmers mixing maize with bean (Macharia 2004). In addition, irrigated farming is also practiced in locations neighboring the Athi River to facilitate small-scale cultivation of vegetables, tomatoes and chili peppers.

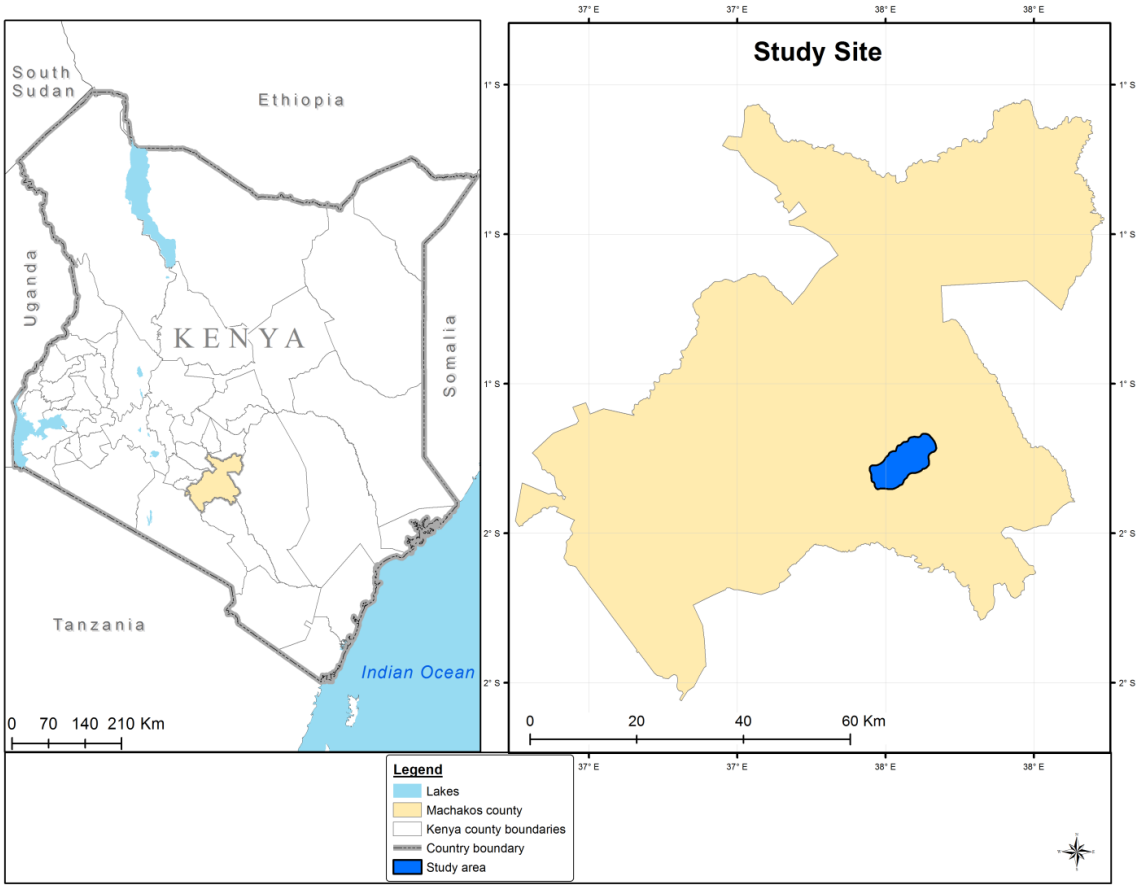

Figure 2.1 Locations of the study area in Machakos County, Kenya, with the dark blue polygon showing the study area. Grey lines illustrate sub-county boundaries.

2.3.1 Methodology

Figure 2.2 summarizes our methodological approach for mapping the two cropping systems (mono- and mixed cropping). Basically, we employed a 2-step hierarchical classification approach to map the maize cropping systems using bi-temporal RapidEye data and the RF classification algorithm. In the first step, we produced a general land use/ land cover (LULC) classification map to separate cropland from non-cropland in

order to reduce the data complexity for subsequent classification (Forkuor et al. 2014a). In the second step, we classified the extracted crop mask into two cropping systems (viz. mono- and mixed cropping).

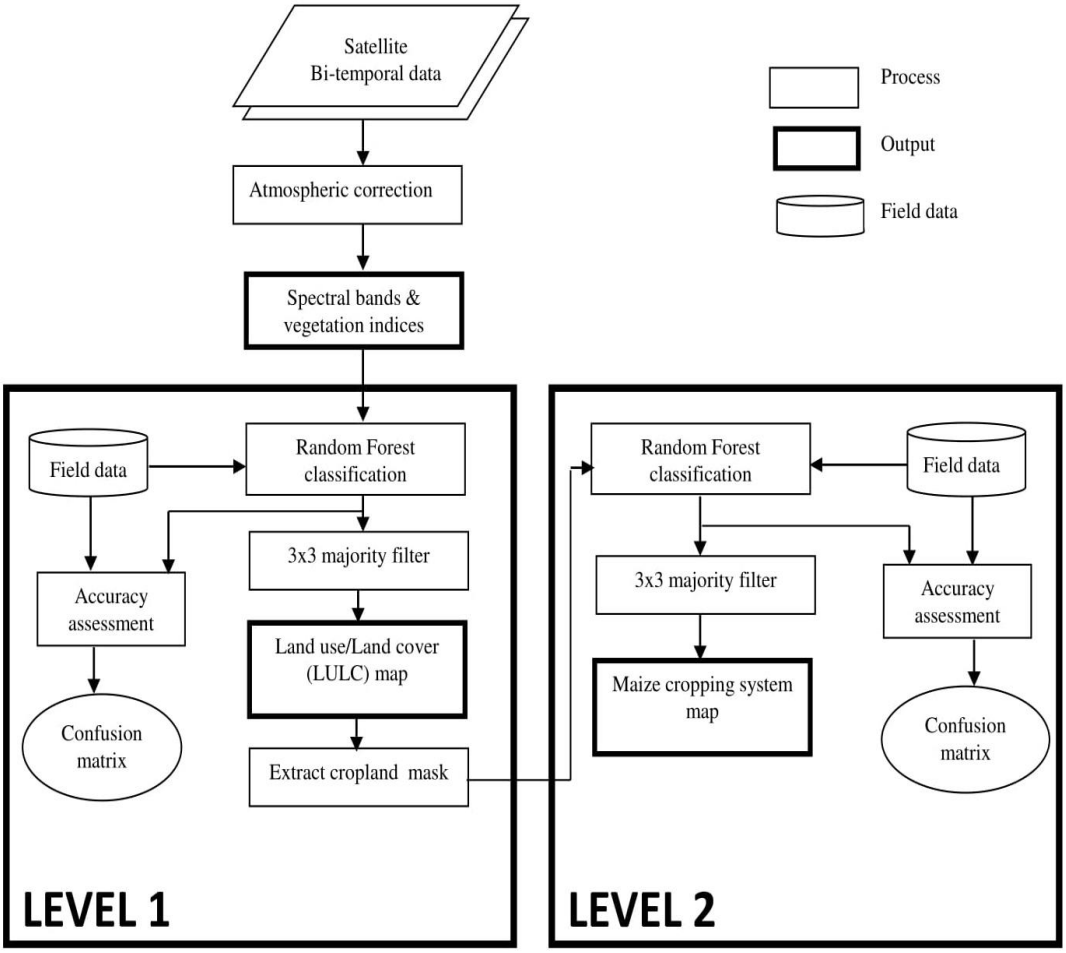

Figure 1.2 Flow chart of the hierarchical classification approach using random forest (RF) classifier and Bi-temporal RapidEye data.

2.3.2 Image acquisition and preprocessing

Two RapidEye images were acquired for the study area on the 3 January 2015 and the 27 January 2015, during the maize stem elongation (RE1) and flowering (RE2) crop phenological development stages, respectively. These two maize phenological development stages are characterized using the BBCH (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie) scale (Lancashire et al. 1991). The images were acquired at two different acquisition windows to assess the effect of crop phenology (the crop life cycle) on mapping cropping systems with an assumption that

the spectral features of the crops (mainly bean and cowpea) that are mainly planted with maize are distinguishable during the maize flowering stage.

RapidEye provides images with spatial resolution of 5 m and five spectral bands (wavelength regions) which are located at blue (440-550 nm), green (520-590 nm), red (630-685 nm), red-edge (690-730 nm) and near infrared (760-850 nm) regions of the electromagnetic spectrum. The RapidEye Ortho product (Level 3A) was utilized. To retrieve surface reflectance, atmospheric correction was applied using the atmospheric-topographic correction (ATCOR3) software to remove haze and other atmospheric interferences. ATCOR3 is an extension of model ATCOR2 that permits extended three-dimensional topographic corrections by inclusion of digital elevation model (DEM) data to remove illumination difference due to topography effects (Richter 1997). Parameters, such as satellite azimuth, illumination elevation, azimuth and incidence angle, which are used for atmospheric corrections, were obtained from respective metadata files for each image. To reduce illumination effects caused by the terrain on RapidEye imagery, topographic corrections were performed using Shuttle Radar Topographic Mission (SRTM) digital elevation (DEM) data. The 30 m SRTM DEM data, used in topographic correction processes, were re-sampled to 5 m pixel resolution using a bilinear interpolation technique. Due to different date and time acquisitions of each tile in the RapidEye mosaic, a tile specific stepwise normalization technique using multivariate alteration detection (IMAD) was used to normalize the tiles via a central normalization reference tile. The image data sets were geo-referenced to Universal Transverse Mercator (UTM, zone 36 south).

Subsequently, mosaicking was applied on all the co-registered normalized tiles and two mosaics having a size of ~ 61 by 61 km each were produced initially. To align all the corresponding pixels, the two mosaicked images were co-registered to each other using image-to-image co-registration to ascertain the alignment of corresponding pixels. Finally, regions that were covered by clouds have been masked out. We utilized 30 vegetation indices calculated from each RapidEye data set (Table 2.1) together with the respective five RapidEye bands as input into the RF classification algorithm. The inclusion of vegetation indices that are related to vegetation biochemical and biophysical traits

like chlorophyll activity and leaf area index, together with the individual spectral bands, has proved to significantly improve crop classification accuracies in heterogeneous landscapes (Forkuor et al. 2014a).

2.3.3 Field data collection

A field campaign was conducted within three days from the first image acquisition date (3 January 2015) to collect reference data on croplands which in our study area are solely mono- and mixed maize cropping systems. Further, field data were collected on non-cropland, which are composed of water bodies, artificial surfaces and natural vegetation. A stratified random sampling was followed to collect the reference data. A handheld Global Positioning System (GPS) device with an error of ±3 m was used to locate the reference control points. Once a field was identified, we delineated the field boundaries (polygon) within a minimum area of 30 by 30 m. To avoid the edge effect, we collected the polygon data five meters away from the edge of each field. Geo tagged photographs of each cropping system in the sample fields were taken from the main four cardinal directions and from the center of the fields for further inspections of the cropping systems and crop age. To mitigate the effect of soil background on the crops' spectral features, we only sampled maize fields (mono- and mixed cropping systems) that were about three weeks old at the first image acquisition date. The reference data were randomly divided into 70% training and 30% validation set. The training set was used to train the RF classifier while the validation dataset was used to evaluate the accuracy.

Table 2.1 Spectral vegetation indices used in the study. Source information is given in the last column.

Name	Index	Formula	Reference
Canopy Chlorophyll Content Index	CCCI	$((R_{NIR} - R_{red\_edge})/(R_{NIR} + R_{red\_edge}))/((R_{NIR} - R_{red})/(R_{NIR} + R_{red}))$	(El-Shikha et al. 2008)
Normalized Difference Red edge	NDRE	$(R_{NIR} - R_{red\_edge})/(R_{NIR} + R_{red\_edge})$	(Basso et al. 2011)
Transformed Soil Adjusted Vegetation Index	TSAVI	$B(R_{NIR} - B * R_{red} - A)/(R_{red} + B(R_{NIR} - A) + X(1 + B^2))$	(Strachan et al. 2002)
Soil Adjusted Vegetation Index Red Edge	SAVI-edge	$1.5 * (R_{NIR} - R_{red\_edge})/(R_{NIR} + R_{red\_edge} + 0.5)$	
Leaf Chlorophyll Index	LCI	$(R_{NIR} - R_{red\_edge})/(R_{NIR} + R_{red})$	(Pu et al. 2008)
Soil Adjusted Vegetation Index	SAVI	$1.5 * (R_{NIR} - R_{red})/(R_{NIR} + R_{red} + 0.5)$	(Roujean and Breon 1995)
Normalized Difference Vegetation Index	NDVI	$(R_{NIR} - R_{red})/(R_{NIR} + R_{red})$	(Tucker 1979)
Difference Vegetation Index	DVI	$R_{NIR} - R_{red}$	(Tucker 1979)
Rationalized Normal Difference Vegetation Red-Edge Index	RNDVI-edge	$(R_{NIR} - R_{red})/(R_{NIR} - R_{red})^{1/2}$	(Roujean and Breon 1995)
Simple Ration	SR	R_{NIR}/R_{red}	(Birth and McVey 1968)
Chlorophyll Green	Chlgreen	$(R_{NIR} - R_{green})^{-1}$	(Gitelson et al. 2006)
Chlorophyll Red-Edge	ChlRed-edge	$(R_{NIR} - R_{red\_edge})^{-1}$	(Gitelson et al. 2006)
Green Normalized Difference Vegetation	GNDVI	$(R_{NIR} - R_{green})/(R_{NIR} + R_{green})$	(Huang et al. 2007)
Simple Ratio 672/550 Datt5	SR672/550	R_{red}/R_{green}	(Datt 1998)
Simple Ratio 695/670 Carter 5	Ctr5	R_{red_edge}/R_{red}	(le Maire and Francois 2004)
Simple Ratio 710/760 Carter 4	Ctr4	R_{red_edge}/R_{NIR}	(le Maire and Francois 2004)
Wide Dynamic Range Vegetation Index	WDRVI	$(0.1R_{NIR} - R_{red\_edge})/(0.1R_{NIR} + R_{red\_edge})$	(Ahamed et al. 2011)
Enhanced Vegetation Index	EVI	$2.5 * (R_{NIR} - R_{red})/(R_{NIR} + 2.4R_{red}+1)$	(Ahamed et al. 2011)

Modified Chlorophyll Absorption Ratio Index	MCARI	$((R_{NIR} - R_{red}) - 0.2(R_{red\_edge} + R_{green}))(R_{red\_edge}/R_{red})$	(Hunt jr et al. 2011)
Rationalized Normal Difference Vegetation Index	RNDVI	$(R_{NIR} - R_{red\_edge})/(R_{NIR} - R_{red\_edge})^{1/2}$	
Disease Water Stress Index	DSWI-4	R_{green}/R_{red}	(Apan et al. 2003)
Modified Chlorophyll Absorption Ratio Index	MCARI		
Structure Intensive Pigment Index 3	SIPI3	$(R_{NIR} - R_{blue})/(R_{NIR} + R_{red})$	(Blackburn 1998)
Anthocyanin Reflectance Index	ARI-edge	$(1/R_{green}) - (1/R_{red\_edge})$	(Gitelson et al. 2001)
Disease Water Stress red-edge Index	DSWI-edge		
Structure Intensive Pigment Index 2	SIPI2	$(R_{NIR} - R_{blue})/(R_{NIR} + R_{red\_edge})$	(Blackburn 1998)
Enhanced Vegetation Index red-edge2	EVI-edge2	$2.5 * (R_{NIR} - R_{red\_edge})/(R_{NIR} + 2.4R_{red\_edge} + 1)$	
Transformed Soil Adjusted Vegetation Index Red-Edge	TSAVI-edge	$B(R_{NIR} - B * R_{red\_edge} - A)/(R_{red\_edge} + B(R_{NIR} - A) + X(1 + B^2))$	
Difference Vegetation Index red-edge	DVI-edge	$R_{NIR} - R_{red\_edge}$	
Green Leaf Index	GLI	$2(R_{green} - R_{red} - R_{blue})/2(R_{green} + R_{red} + R_{blue})$	(Hunt jr et al. 2011)

Notes: R_{blue}, R_{green}, R_{red}, R_{red_edge} and R_{NIR} are surface reflectance value at blue (band1) green (band2), red (band3), red-edge (band4) and near infrared (band5) of RapidEye. The parameters for TSAVI slope of the soil line (A) = 1.2 and intercept of the soil line (B) = 0.04 and adjustment factor (X) = 0.08

2.3.4 Variable importance measure and classification

A supervised machine learning RF classification algorithm (Breiman 2001) was used to classify the bi-temporal RapidEye image data. RF is considered a robust and efficient classification approach for crop mapping using high spatial resolution satellite data like RapidEye, especially within heterogeneous landscape (Forkuor et al. 2014). It has the potential to handle noisy and highly correlated predictor variables, which commonly occur in remotely sensed data (Curran and Hay 1986). In particular, RF is an ensemble modeling technique, developed by Liaw and Wiener (2002) to improve the classification and regression trees (CART) by combining a large set of decision trees. Each tree in the RF ensemble is built from a bootstrapped random sample containing approximately two-thirds of the training data drawn with replacement. The remaining one third of the data that is not included in the bootstrapped training sample, i.e. the out-of-bag (OOB) samples, is used to internally evaluate the classification performance. RF classifier uses two user-defined parameters (*ntree* and *mtry*). To improve the classification accuracy, the number of trees (*ntree*) grown and variables used at each tree split (*mtry*) were optimized based on the OOB error rate with a grid search and a tenfold cross validation method (Waske et al. 2009). The number of optimal trees (*ntree*) was searched between 500 to 2,500 using a 500 interval, while the optimal *mtry* was searched on the *mtry* vector of a multiplicative factor with the default *mtry* being the square root of the total number of spectral variables (indices and/or bands) (Breiman 2001). The ensemble measures the importance of each spectral variable used in the classification by utilizing the permutation of variables which calculates variable importance as the mean decrease in classification accuracy using the OOB samples.

To select the optimal combination of spectral variables that achieved significant accuracies from the important variables returned by RF classification model using the OOB error rate, we used the RF backward feature elimination method using the "varSelRF" package (Diaz-Uriarte 2017) in R statistical software (R core 2013) for the level 2 of distinguishing the two maize cropping systems. To select the most relevant spectral variables without any over-fitting, a .632+ bootstrap method with a leave-one-out cross-validation procedure and replacement from samples that are not part of the

RF classification was applied (Efron and Tibshirani 1997). The optimum numbers of spectral variables selected were employed to produce the final cropping systems map. Finally, a 3 × 3 post-classification majority filter was applied to spatially smooth the classified images' dominant classes so as to reduce salt-and-pepper effects in the classification output map.

2.3.5 Accuracy assessment

A confusion matrix was constructed to assess the accuracy of the classified maps using the overall accuracy (OA), producers' accuracy (PA) and users' accuracy (UA). The most recently proposed allocation quantity (QD) and allocation (AD) disagreements (Pontius and Millones, 2011) were also calculated from the classification confusion matrix to evaluate the reliability of the classification map and to measure the agreement between the predicted classification features and the reference field data (OOB samples). Class-wise accuracy assessment was performed for each class using F1-score (Schuster et al. 2012). This measure represents the harmonic between PA and UA for each class i as follows:

$$(F1)_i = \frac{2 \times PA_i \times UA_i}{PA_i + UA_i} \tag{2.1}$$

The advantage of using the F1 score for class accuracy evaluation is to give equal importance to both precision and recall, by combining PA and UA into a fused measure.

2.4 Results

2.4.1 Parameterization of the random forest classifiers

The RF grid search with tenfold cross validation method indicated that *ntree* value of 1,000 combined with *mtry* value of 5 was optimal for classifying general land use and land cover (LULC) classes (Figure 2.3a) to separate cropland from non-cropland. On the other hand, the *mtry* value of 15 combined with *ntree* value of 2,000 resulted in the lowest OOB error rate of 0.18% for classifying the mono- and mixed maize cropping systems as shown (Figure 2.3b).

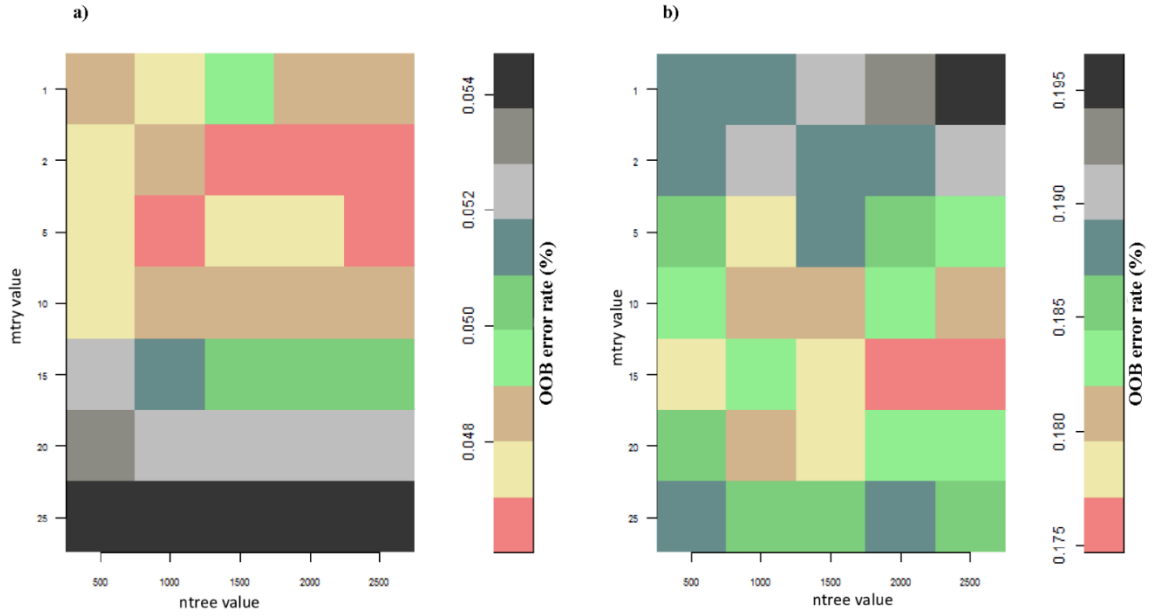

Figure 2.2 Random forest mtry and ntree optimization grid for the land use land cover (LULC) classification result (a) and for the mono- and mixed maize cropping systems mapping result (b) using the internal out-of-bag (OOB) error rate of RF resulting from a grid search with a tenfold cross validation setting.

2.4.2 Spectral variable importance for crop systems mapping

The backward variable selection method, applied on the RF variable importance ranking, resulted in selecting 15 RapidEye spectral variables (Figure 2.4) that were found to be the most relevant for mapping mono- and mixed maize cropping systems after crop masking. Using the LULC map result (Figure 2.3) Nine and six spectral variables, respectively, were selected as important variables from the two RapidEye images, captured during the stem elongation and the flowering development stages, respectively (Figure 2.4). Moreover, most of the selected variables from both observation periods were the RapidEye spectral wavebands themselves. In general, all five RapidEye bands (blue, green, red, red-edge and near infrared) were selected as useful spectral features for classifying the two maize cropping systems, while only five indices (RE1_NDVI, RE1_NDRE, RE2_DVIedge, RE1_LCI and RE1_SIPE3) were useful for separating different maize cropping systems (Figure 2.4).

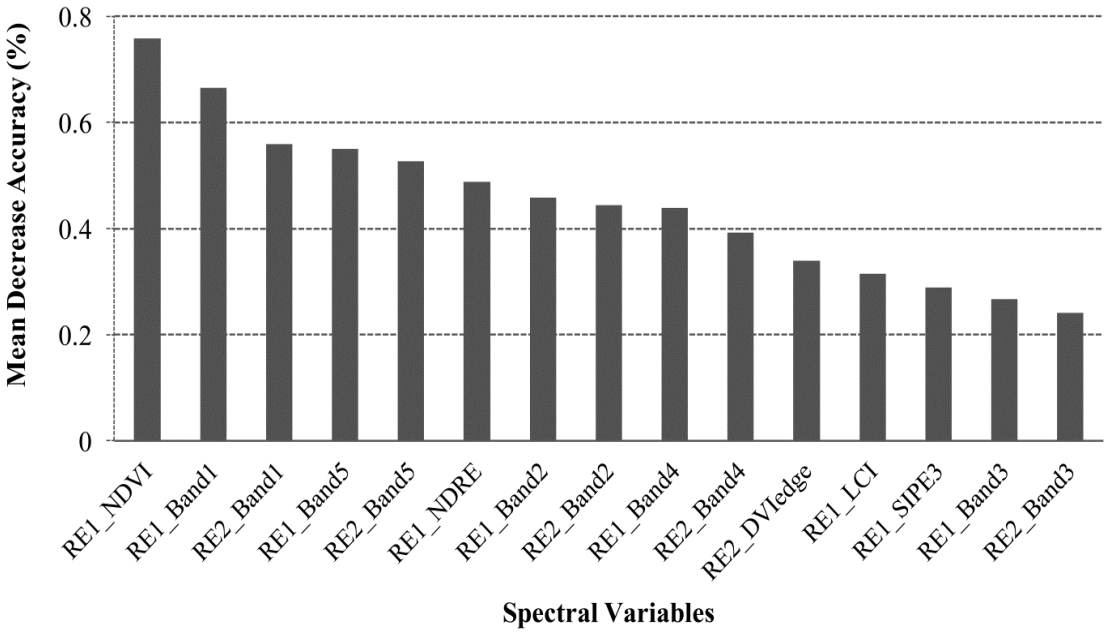

Figure 2.3 Mean Decrease in Accuracy of the 15 most important input variables that were selected using the random forest backward feature elimination function and the .632+ bootstrapping function on importance ranking.

2.4.3 Maize cropping systems mapping

Visual interpretation of the RF capability to separate the two mapped classes (mono cropping and mixed cropping) using a multidimensional class separability proximity matrix indicate that the majority of the pixels are generally well separable as shown Figure 2.5.

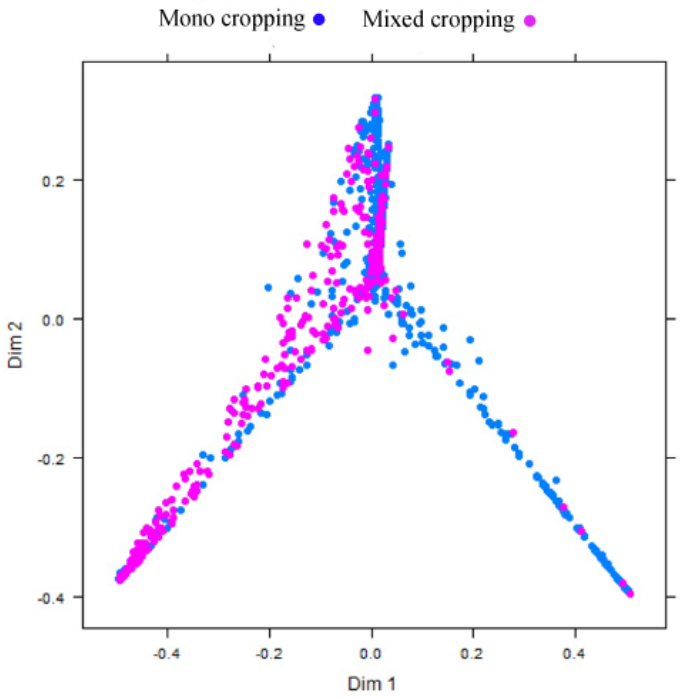

Figure 2.4 Random forest class separability proximity matrix using multidimensional scaling (MDS); Dim 1 refers to dimension 1 and Dim 2 to dimension 2.

The final thematic cropping systems map produced via the RF algorithm is shown in Figure 2.6. It shows that mixed-cropped fields are mostly present in the middle and towards the north-eastern part of the study area which is characterized by a lower altitude (around 1100 MAMSL.), while most of the mono-cropped fields are found in the south-western part of the study area at a mean altitude of 2000 MAMSL. Mono cropped fields in the higher lying areas in the range of 1400-2000 m above sea level appeared larger and less scattered (fragmented) than those on mixed cropping in the lower areas with elevation below 1400 m (Figure 2.6).

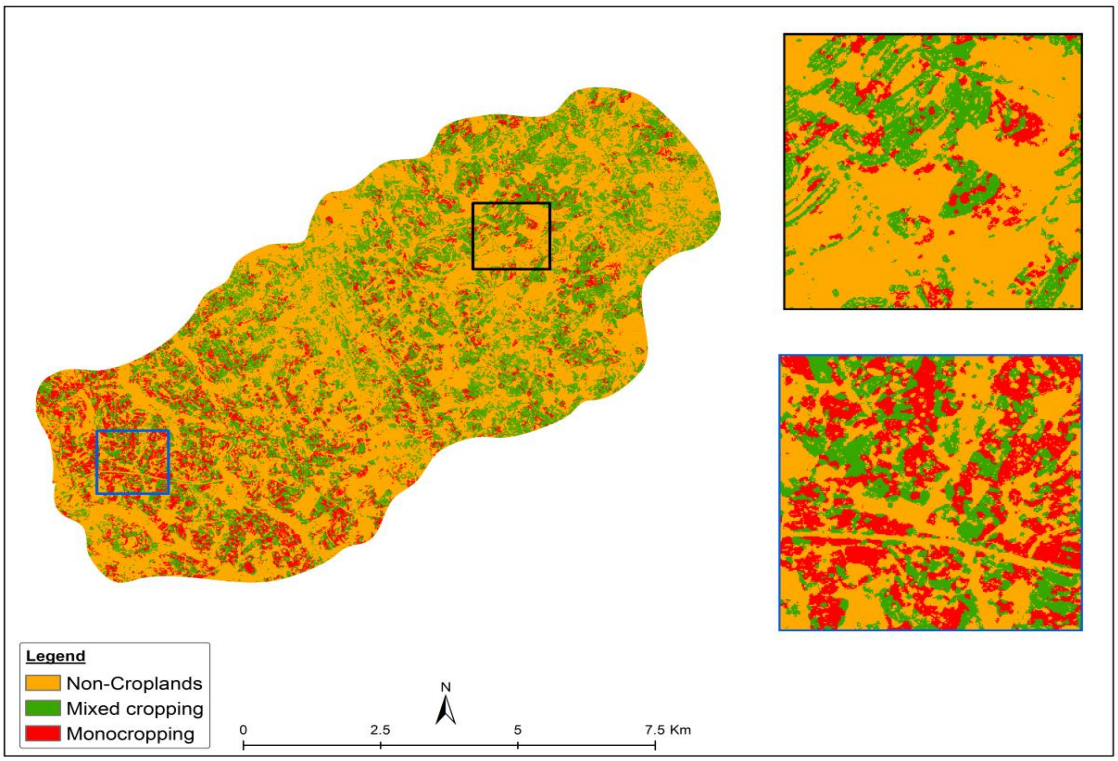

Figure. 2.5 Maize cropping systems classification map obtained using the proposed classification scheme (Figure 2.2). The two inserts, as black and blue squares, illustrate two contrasting areas in terms of cropping systems.

2.4.4 Classification accuracies

To understand the accuracies of the cropping systems map (level 2 classification) that uses the cropland mask extracted from LULC result (level 1 classification), the accuracies of LULC map are herein reported. The main results of the accuracy assessment for the hierarchical level 1 LULC classification are summarized in Table 2.2 for the three different combinations which included the RE bands, RE band with all vegetation indices and RF selected spectral variables (bands and vegetation indices). The most accurate LULC mapping result was obtained from the most important selected vegetation indices and spectral bands listed in Figure (2.4) using RF backward selection criteria with an overall accuracy of 93.2% and a kappa coefficient of 0.91. Table 2.3 presents the per-pixel evaluation confusion matrix for the LULC map. Individual accuracies (PA and UA) were consistently over 87% with F1-score averagely above 0.88 for all classes (Table 2.3). This suggests a very good concealment of croplands from other LULC classes regardless of cropland being slightly confused with the natural vegetation class.

Table 2.4 summarizes classification accuracies for three classifications calibrated for mapping the two maize cropping systems (level 2 classification). The optimized RF cropping system mapping result using only the most relevant spectral variables selected by RF gave an overall accuracy of 85.7% (kappa coefficient of 0.84) whereas the non-optimized result, using all RapidEye bands and the vegetation indices, gave a lower overall accuracy of 73.4%. In addition, individual PA and UA for the optimized RF result (Table 2.5) were consistently above 84% for both classes with a low QD score of 1% and a relatively high AD score of 13% for both cropping systems, respectively.

Table 2.2 Overall accuracies and kappa coefficient of agreement for the Land Use Land Cover classification.

Analysis	Overall accuracy (%)	Kappa coefficient
RE (bands)	87.46	0.86
RE (bands)+All RE_veg indices	86.41	0.84
RF selected spectral variables	93.20	0.91

Table 2.3. Random forest classification confusion matrix for the land use/ land cover classes (Level 1) using the 15 most important RapidEye spectral variables and 30% of the reference data.

Class	Artificial Surface	Cropland	Natural Vegetation	Water Bodies	Total	UA (%)	F1 Score
Artificial Surface	904	23	0	18	945	96.48	0.96
Cropland	11	845	89	0	945	87.84	0.89
Natural Vegetation	0	94	851	0	945	90.53	0.90
Water Bodies	22	0	0	923	945	98.09	0.98
Total	937	962	940	941	3780		
PA (%)	95.66	89.42	90.05	97.67			
OA (%)	93.20						

Table 2.4. Overall accuracies and kappa coefficient of agreement for the two maize mono- and mixed cropping systems.

Analysis	Overall accuracy (%)	Kappa coefficient
RE (bands)	80.24	0.77
RE (bands) + All RE_veg indices	73.38	0.70
RF selected spectral variables	85.71	0.84

Table 2.5 Random forest classification confusion matrix for mapping cropping systems using the most important 15 RapidEye spectral variables and 30% of the reference data.

Class	Mono Cropping	Mixed Cropping	Total	UA (%)
Mono maize cropping	486	74	560	84.97
Mixed maize cropping	86	474	560	86.50
Total	572	548	1120	
PA (%)	86.79	84.64		
OA (%)	85.71			
QD (%)	1.00			
AD (%)	13.00			

2.5 Discussion

The classification results from this study demonstrated the usefulness of the bi-temporal RapidEye imagery and RF classification tool for mapping the two major maize cropping systems in heterogeneous agro-ecological landscapes. This demonstrates the capability of high spatial resolution data with better spectral coverage such as RapidEye to distinguish different cropping systems, given both the size and shape of fields in African agro-ecological systems. The two images could have captured the spectral (phenological) profiles of the two cropping systems and resulted in a better discriminatory power between the two cropping systems (Forkuor, 2014b). In addition,

RF selected NDVI from the first RapidEye acquisition (RE1) among the most significant variable in separating the two cropping systems. This could be due to the fact that at the stem elongation phenological development stage, mono- and mixed cropping systems are easily distinguishable, while at the flowering phenological growth stage the two cropping systems seem to have similar morphological and spectral properties (Zhang et al. 2012). We observed that the most important spectral indices selected by RF from the two acquisitions are commonly related to plant biophysical properties such as Leaf Area Index (LAI) and net primary productively (Kim and and Yeom 2015). It is expected that mono- and mixed maize cropping systems can differ considerably in these traits since mono-cropped maize is known to exhibit a lower LAI especially during the flowering stage than maize mixed with legumes (among other factors, largely due to the absence of bare soil) (Sun et al., 2014). It is interesting to note that the RE2_NDVI was not amongst the selected most important variables. That could be due to fact that the second RE2 acquisition corresponded to the maize flowering stage in which the spectral contribution are not overly characterized by chlorophyll activities (NDVI) but more by spectral contributions from non-chlorophyll plant components, i.e. the cobs, wilting leaves and the maize flowers. Moreover, the RF backward variable selection process showed that the inclusion of the red-edge bands and spectral indices that use the red-edge bands (i.e., DVI-edge) were relevant for the maize cropping systems mapping result. Similarly, Schuster et al. (2012) found that the red-edge bands improved land use classification by 2.7%.

The spatial cropping system differences we observed between the low and high altitude areas (Figure 2.6) could be due to more favourable climatic conditions for crop production within mountainous regions as single (mono) crop production is more feasible in upland areas that receive higher rainfall (Seck et al. 2012). Farmers in drier areas often opt to combine maize with leguminous on the same field (mixed cropping) to improve soil fertility and soil moisture in order to attain permissible yields (Castellanos et al. 2014). The crop systems patterns also showed considerable differences in field size between the lower and higher lying areas within the study area (approximately 0.8 ha in the low altitudinal areas versus an average field size of 0.2 ha

in the higher lying areas). These field size differences could be confirmed from the field observations.

The crop system mapping results indicated comparatively high OA of 85.7% and individual class accuracies (PA, and UA) > 84% (Bassa 2012). The high accuracies could be due to, primarily, the optimal acquisition dates of the imagery, i.e., during the stem elongation and flowering crop development stages, respectively. These critical crop stages are known to produce better separation between agricultural fields and surrounding natural vegetation (Arvor et al. 2011). Specific confusions between the two cropping system classes could have numerous reasons such as heterogeneity of the landscape, variation in crop age (planting dates) and other agronomical practices (e.g., ploughing) (Forkuor et al. 2014; Ozdogan et al. 2010). Field heterogeneity is largely affected by within field spectral variations that are larger than inter-field spectral variations due to the crop morphological and physiological properties (Smith et al. 2003). This confusion is exacerbated by the amount of weed infestation per cropped area where a weed-infested mono-cropped maize field could exhibit a similar spatial arrangement and thus spectral response like a maize field inter-cropped with a legume like cowpea or bean (Romeo et al. 2012). Some sample mono-cropped fields in our study area were badly managed and infested by weeds that could have caused the spectral confusion with inter-cropped fields as previously mentioned. Another reason for heterogeneity and spectral confusion could be the fact that some farmers maintain trees within their fields and since we applied a per-pixel classification accuracy assessment, spectral confusion between trees and crops could have been exacerbated by this (Cord et al. 2010). In other words, some mono-cropped and inter-cropped maize fields could have had similar spectral features as a result of woody vegetation that can be found within maize fields. Also, in Machakos, the majority of the farmers cultivate crops around hamlets, and in many cases the cultivated fields are surrounded by pockets of natural vegetation (Vintrou et al. 2009). However, we employed an empirical classification approach that could have been more robust in terms of practical and operational cropping system mapping. Since our mapping results are produced using the most important variables (Figure 2.4). We assumed that collinearity is somewhat

accounted for (Strobil et al. 2008). Furthermore, we tested the collinearity between the indices selected for final mapping (RE1_NDVI, RE1_NDRE, RE2_DVIedge, RE1_LCI and RE1_SIPE3) and found that they were not correlated.

Maize is generally vulnerable to numerous pests and diseases (Flett et al., 2002). Choosing appropriate cropping systems can be a valuable alternative to the use of synthetic pesticides (Gianoli et al. 2006). For instance, inter-cropping has been used as a buffer against the spread of plant pests and pathogens by attracting pests away from their host plant and also increasing the distance between plants of the same species, making it more exigent for the pest to target their main crop (Seran and Brintha. 1997). A good example is inter-cropping maize with cowpea/beans has been proved to reduce the maize stem borer (Henrik and Peeter 1997) and cowpea/bean thrips (Nyasani et al. 2012). As a result, accurate baseline information on cropping systems could be used to better understand the relationship between cropping patterns with pest and disease propagation mechanisms such as the occurrence of the maize lethal necrosis (MLN) disease, reported in Kenya in 2012, which is hypothesized to be linked to the spatial distribution of the cropping systems (Mahuku et al. 2015).

Essentially, “traditional” agricultural land use mapping often renders information on the spatial distributions or acreages of certain crops without further details of the actual underlying agronomical cropping systems (Husak et al. 2008). Information on the cropping systems is vitally needed as a spatial descriptor (parameter) within commonly used crop modeling schemes such as the Decision Support System for Agro technology Transfer (DSSAT), since these crop systems are key determinants for agricultural production and food supply, given that mono-cropped systems generally may exhibit different yield cycles than mixed cropping systems. With the advent of new satellite constellations with better pixel and temporal resolution, not only crop mapping but also crop systems characterization can be performed. This will be of great use to crop scientists and decision makers. Moreover, cropping patterns can be related to climate change effects and thus to agricultural productivity and as a result the extent of food security as for instance mixed cropping systems is a key adaptation mechanism for areas experiencing considerable climate variability.

2.6 Conclusions

This study evaluated the potential of 5-m RapidEye multispectral data and the advanced RF classification technique in mapping maize cropping systems in a complex, dynamic and heterogeneous landscape. To the best of our knowledge, we produced the first cropping patterns map for a maize-based cropping system in Kenya. We conclude that RapidEye imagery acquired during stem elongation and flowering phenological development stages give satisfactory results for separating mono- and mixed maize cropping systems.

We suggest that data on cropping systems mapping, using high resolution time-series data, are useful baseline information feeds to monitor and understand seasonal cropping pattern changes as a function of climatic variability and climate change in Africa. This information, especially if linked with yields and crop pest and disease infestation levels may be very useful for better agricultural risk projections. Upholding the widely recognized role of the small-scale farming for food security in Africa, the extent, distribution and dynamics of cropping systems, as one of the important variables in crop productivity models, should further be investigated. This will result to increased farm sizes hence making it more feasible to monitor cropping systems using freely available remote sensing data sets.

However, temporal availability of high-resolution data is still restricted by the high cost of imagery from semi-commercial sensors such as RapidEye and frequent cloud cover, especially in the tropics. Upscale of the results from this study to wide-area monitoring of cropping patterns is thus still challenging. Freely available multi-temporal dataset from Landsat-8 combined with Sentinel 2a and Sentinel 2b data should be further investigated and exploited to improve cropping systems mapping in Africa and beyond. Overall, relatively accurate classification results obtained in this study provide dependable information that could be used to complement region or field-specific yield data to aid decision making in terms of improved crop productivity and food supply management.

3 IMPORTANCE OF REMOTELY SENSED VEGETATION VARIABLES FOR PREDICTING THE SPATIAL DISTRIBUTION OF AFRICAN CITRUS TRIOZID (*TRIOZA ERYTREAE*) IN KENYA

This chapter has been published as Kyalo et al. (2018) *ISPRS Int. J. Geo-Inf* 7(11): 429

3.1 Abstract:

Citrus is considered one of the most important fruit crops globally due to its contribution to food and nutritional security. However, the production of citrus has recently been in decline due to many biological, environmental, and socio-economic constraints. Amongst the biological ones, pests and diseases play a major role in threatening citrus quantity and quality. The most damaging disease in Kenya, is the African citrus greening disease (ACGD) or Huanglongbing (HLB) which is transmitted by the African citrus triozid (ACT), Trioza erytreae. HLB in Kenya is reported to have had the greatest impact on citrus production in the highlands, causing yield losses of 25% to 100%. This study aimed at predicting the occurrence of ACT using an ecological habitat suitability modeling approach. Specifically, we tested the contribution of vegetation phenological variables derived from remotely-sensed (RS) data combined with bio-climatic and topographical variables (BCL) to accurately predict the distribution of ACT in citrus-growing areas in Kenya. A MaxEnt (maximum entropy) suitability modeling approach was used on ACT presence-only data. Forty-seven (47) ACT observations were collected while 23 BCL and 12 RS covariates were used as predictor variables in the MaxEnt modeling. The BCL variables were extracted from the WorldClim data set, while the RS variables were predicted from vegetation phenological time-series data (spanning the years 2014-2016) and annually-summed land surface temperature (LST) metrics (2014-2016). We developed two MaxEnt models; one including both the BCL and the RS variables (BCL-RS) and another with only the BCL variables. Further, we tested the relationship between ACT habitat suitability and the surrounding land use/land cover (LULC) proportions using a random forest regression model. The results showed that the combined BCL-RS model predicted the distribution and habitat suitability for ACT better than the BCL-only model. The overall accuracy for the BCL-RS model result was 92% (true

skills statistic: TSS = 0.83), whereas the BCL-only model had an accuracy of 85% (TSS = 0.57). Also, the results revealed that the proportion of shrub cover surrounding citrus orchards positively influenced the suitability probability of the ACT. These results provide a resourceful tool for precise, timely, and site-specific implementation of ACGD control strategies.

3.2 Introduction

Citrus is considered one of the most important fruit crops in the world due to its contribution to food and nutritional security (Franco-Vega et al. 2016). Also, citrus is the top-ranked fruit crop with regard to its international trade value (Liu et al. 2012). The commercially important citrus species are sweet oranges (*Citrus sinensis*), lemons (*Citrus limon*), limes (*Citrus aurantifolia*), grapefruit (*Citrus paradisi*), and tangerines (*Citrus reticulata*). Globally, sweet oranges represent approximately 70% of citrus production. In 2016, the global total production of sweet oranges was about 73 million tons (FA0 2016). In Kenya, citrus is a valuable fruit crop used mainly for domestic consumption as a fresh produce with only a small quantity being processed into juice and jams (Ouma, 2008). Citrus provides some minerals and vitamins like vitamin C, carotenoids, and polyphenols that are essential for human health. In terms of the area of production, citrus (mainly oranges) ranks third (7,268 ha) after bananas (63,299 ha) and mangoes (54,332 ha) in the country (Adhikari et al. 2015).

Citrus plants can prosper in a wide range of environmental conditions from tropical to subtropical climatic conditions (Nicholas 1988). However, the best citrus production conditions are found in subtropical climate zones in elevations ranging from sea level up to 2100 m above mean sea level (m.m.a.s.l), with an optimal growth temperature ranging from 20 °C to 30 °C. In Kenya, citrus fruits' quantity and quality have been considerably declining. For instance, oranges yields at 11.73 ton ha-1 are far below (23% less) the global mean yield of 18.45 ton ha^{-1} (Asharaf et al. 2002; Waithaka 1991). Two of the major production constraints that hinder citrus production in Kenya are insect pests and diseases, among which the African citrus triozid (ACT), Trioza erytreae, plays a key role (Icipe 2015). Direct feeding by ACT results in leaf curling and,

furthermore, causes deposition of honeydews on infested plants (Khamis et al. 2017a). In Africa, ACT is known for transmission of the devastating phloem-limited bacterium *Candidatus* Liberibacter africanus (CLaf), responsible for the African citrus greening disease (ACGD) or Huanglongbing (HLB) (Zou et al. 2012). In addition to ACT, HLB is also transmitted by the Asian citrus psyllid (ACP) (*Diaphorina citri*), which is the primary vector in Asia (Boykin et al. 2012; Jagoueix et al. 1994) but was also recently discovered in eastern Africa (khamis et al., 2017). These two psyllids are distributed according to their temperature requirements, with ACT being highly temperature sensitive and thus restricted to cooler elevated areas (Aubert 1987; Catling 1973). The common symptoms of ACGD are mottling and yellowing of the leaves, reduced tree foliage which results in small and bitter-tasting fruits, and the eventual death of severely infected citrus trees (Dormann et al. 2013). In Kenya, ACGD is reported to have had the greatest impact on citrus production in the highlands, causing yield losses of 25% to 100% (Pole et al. 2010). The yield of affected trees is not only considerably reduced by continuous fruit drop, dieback, and tree stunting, but also by the poor quality of fruits that remain on the trees which are inedible.

Over the years, different approaches have been used for implementing various ACGD preventive and control measures (Alvarez et al. 2016). This includes strict regulations of nurseries through a registered disease-free certification scheme to prevent the spread of ACGD and its vectors (Grafton-Cardwell et al. 2013). Little is known on the spatial distribution of the disease vectors, yet such information could greatly assist in developing precise geo- and time-referenced vector distribution maps. Such maps can be useful in monitoring the spatial spread and suitable areas for the vectors, enabling for a more targeted implementation of interventions. Vector-transmitted disease propagation follows vector ecological principles as an indirect explanation of disease cycles, outbreaks, and prevalence (Moore et al. 2010). One of the most frequently used approaches for producing vector distribution maps is the ecological niche (EN) modeling approach (Peterson 2006). EN models statistically link spatial variabilities in a set of predictor variables to the distribution of species of interest that can be a plant disease vector like ACT (Brownstein et al. 2003; Lord 2007). The

dependence of plant disease propagation on spatio-temporal environmental niche factors of the disease vector has recently received considerable attention (Hol et al. 2013). Yet, studies focusing on the ways in which geographical environmental factors affect the habitat suitability and host-vector dynamics are still limited. In addition, there is a need for studies that employ a multi-source variable (e.g., vegetation phenology and climate) approach to predict the spatial distribution of plant disease vectors.

Models have been developed to provide information about diseases and the distribution of associated environmental variables that are used as proxies for habitat suitability. The best known EN models used in insect-based distribution modeling include Generalized Linear Models (GLM), Generalized Additive Models (GAM), Genetic Algorithm for Rule-Set Production (GARP), Boosted Regression Trees (BRT), and Maximum Entropy (MaxEnt) (Shabani et al. 2016). Studies have compared the performance of several EN modeling algorithms to predict the distribution of different species and found that MaxEnt was the best-performing model using presence-only data (Yackulic et al. 2013). In addition, MaxEnt is the most utilized EN model for estimating the distribution of plant insect pests like stink bugs (*Halyomorphahalys* spp.) (Zhu and Woodcock 2012), large pine weevil (*Hylobius abietis*), and horse-chestnut leaf miner (*Cameraria ohridella*) (Barredo et al. 2015), boreal forest insect pests (Hof and Svahlin 2016), fruit flies (Marchioro 2016), and disease vector ticks (*Ixodes ricinus*) (Alkishe et al. 2017).

For HLB, a number of studies employed mathematical, and geostatistical simulation, life table, and conceptual modeling routines (Chiyaka et al. 2012; Ramirez et al. 2016; Vilamiu et al. 2012), to study the distribution of ACP using environmental variables as predictors (i.e., temperature and rainfall) in regard to the biology of the vector (e.g., developmental stages and their populations) and host plant interactions (e.g., number of susceptible or infectious orange trees). These studies demonstrated the possibility of estimating the distribution, progression and optimal temperature ranges for ACP in countries like the USA, Mexico, Brazil, Vietnam, and Australia. In Africa, Shimwela et al. (2016) and Narouei-Khandan et al. (2016) employed two correlative MaxEnt and support vector machine modeling approaches to map the potential

distribution of ACP using global-scale environmental predictors. These two studies reported that Eastern African countries like Kenya and Tanzania would be highly suitable for the psyllid. To the best of our knowledge, no other study has employed an EN modeling approach to predict the distribution of ACGD vectors in Africa.

Yet, there is a need for an explicit ACT distribution mapping routine in countries such as Kenya, where the transmission of ACGD is mainly due to this vector. Moreover, previous studies looked at the relevance and influence of environmental variables in predicting the distribution of ACGD vectors but did not consider the expected relevance of vegetation patterns and phenology, resulting through interactions between climatic, topographic, and vegetation patterns at a landscape scale, which can considerably improve the performance of EN models like MaxEnt (Hof and Svahlin 2016). Moreover, vegetation patterns and phenology play a key role in influencing vector-host-pathogen transmission, including vector distribution, abundance, and diversity (Paull et al. 2012). These vegetation-related patterns and phenological variables can only be extracted from temporal remotely-sensed datasets. When used in EN models, the remotely-sensed vegetation pattern and phenological variables are useful additional predictors for the spatial distribution of pests and diseases since EN models rely on the correlation between a habitat's characteristics and the biophysical properties of the studied pest and disease (Zimmermann et al. 2007).

Further, much research has focused on the biology of ACT and its dispersal (Green and Catling 1971); however, little is known regarding how land use/land cover (LULC) features influence the habitat suitability of the vector and its dispersal. However, remotely-sensed datasets from different systems have been widely used for the identification and separation of citrus orchards from other LULC types for appropriate policy making and citrus production forecasting (Amoros Lopez 2011; Ozdemir 2007; Shrivastava and Gebelein 2007). More efforts concerning understanding the influence of the landscape on the survival of pests and diseases like ACT using remotely-sensed variables are crucial. For instance, the context of the landscape has been reported to affect the population of crop insects directly, or more frequently, indirectly, through its effects on the physical environment around the host plants (Plantegenest et al. 2007).

For instance, landscape heterogeneity has been reported to influence the direction and distance moved by a dispersed pest and pathogens, in addition to the infestation rate (Margosian et al. 2009). For example, Rizzo et al. (2003) reported that the proximity to the forest edge was associated with an increase in the infestation of sudden oak death disease in California. Avellino et al. (2002) tested the relationship between the landscape context and three highly differentiated focal coffee pests and pathogens. They found a positive relationship between the studied coffee pest and disease incidences and the proportion of different LULC classes at different radii around coffee sample plots. Thies et al. (2003) studied the correlation between the local proportion of destroyed oilseed canola buds and the characteristics of landscape context. They showed that an increase in the landscape complexity was associated with a decrease in damage caused to oilseed canola by *Meligethes. aeneus*. All these studies alluded to the fact that the surrounding vegetation provides a refuge for the vectors during periods of time when the conditions are unfavorable for the spreading of the disease. Despite this strong influence that the landscape properties have on the spread of pests and diseases, no research has explored the relationship between ACT habitat suitability with the surrounding landscape composition for a better understanding of the ecology and spread of ACGD.

The objectives of this study were, (i) to explore the potential and contribution of vegetation phenological variables and Land Surface Temperature (LST) derived from remotely-sensed data combined with environmental variables to predict the distribution and habitat suitability for ACT at a test site in Kenya using a MaxEnt model and, (ii) to test the effect of the surrounding landscape context on the habitat suitability of ACT. This was achieved by relating a set of bio-climatic and topographic environmental (BCL) variables and remotely-sensed (RS) variables to ACT presence-only distributions over a region-specific, i.e., representative agro-ecological gradient.

3.3 Methods

3.3.1 Study area

The study site consists of 35 administrative counties in three main agro-ecological zones in Kenya lying in low-, mid-, and high- elevation zones, see Figure 3.1. The study area covers parts of the Coastal, Eastern, Central, and Western regions of Kenya. The Central and Western regions exhibit cooler and wetter climatic conditions which are particularly favorable for citrus growing. The two regions experience a bi-modal rainfall distribution with the major crops being maize and beans, which in most cases are interspersed with mangoes and citrus trees, in addition to tea and coffee. Generally, citrus growing across the entire country is commonly practiced in small orchards and backyards, with only a few big citrus plantations in Kenya.

In the low-lying coastal region with higher humidity levels, farmers cultivate a wide range of food as well as tree crops like coconut palms, mango, citrus, and pawpaw. The major citrus-growing areas in the coastal region are Kwale Kilifi and Taita Taveta (Oosten 1989). The Eastern region is located in the hot and dry semi-arid savannah biome and has similar cropping patterns as the coastal region. It is dominated by steep slopes with elevations ranging from 500 to 1200 m above mean sea level (m. a. m. s. l.).

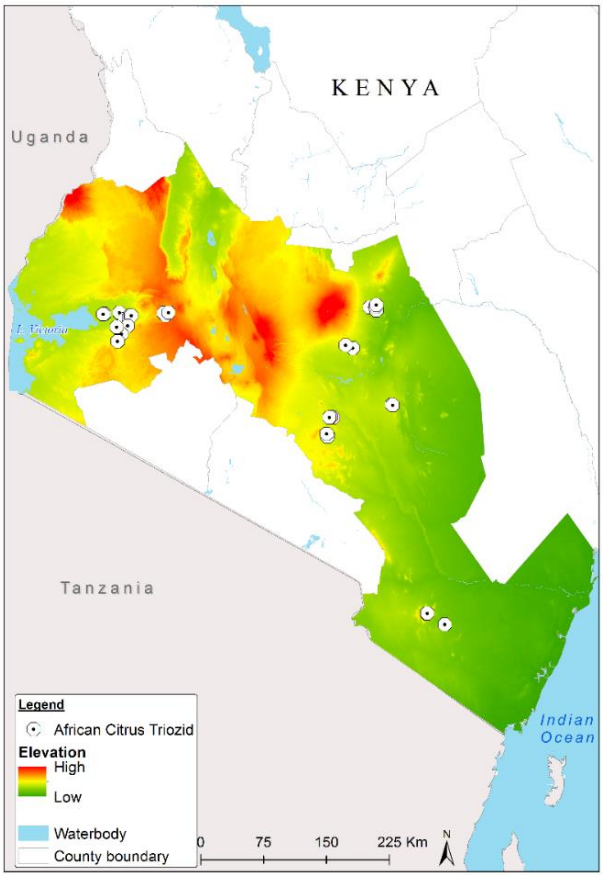

Figure 3.1 Study area (major citrus-growing regions) in Kenya where the African citrus triozid (Trioza erytreae) presence data were collected.

3.3.2 ACT occurrence data

Field surveys were conducted along a clearly defined transect within the citrus-growing regions from the lowlands to the highlands in Kenya. In general, horticultural farming in Kenya is mainly carried out by small-scale producers because of the scarcity of productive land for horticultural production. Thus, citrus is grown in a wide range of elevations ranging from the lowlands to the highlands of Kenya (George 2008). The study area was divided into three elevation zones: Low (0–500 m. a. m. s. l.), middle (501–1,000 m. a. m. s. l.), and high (>1,000 m. a. m. s. l.). Each elevation zone was regarded as a stratum; therefore, we followed a stratified random sampling protocol to collect the ACT presence data. In each stratum (i.e., elevation zone), we randomly selected citrus orchards and nurseries, including backyards of small farms with a minimum orchard-to-orchard distance of 2 km for sampling. At least 30 citrus orchards in each stratum were sampled and with the aid of a hand-held Global Positioning System (GPS) device with a

positional accuracy of ±2 m, the location of the citrus orchard, nursery, or backyard farm where ACT was present was recorded as an occurrence point. Specifically, for sample citrus orchard and backyard farms ≤ 0.5 ha, all citrus trees were inspected for ACT symptoms, while in orchards > 0.5 ha only 20 randomly selected trees were sampled in each orchard by moving across the orchard in a W-pattern (Shimwela et al. 2016). Presence-only observations (n = 47) were collected across the study area, see Figure 3.1, between January 2015 and September 2016. This number of presence-only observations is regarded as acceptable in a MaxEnt modeling routine (Amirpour et al. 2013; Peterson and Nakazawa 2008). A subset of the ACT presence-only observations was used for training the MaxEnt model (75% of the sample observations), and 25% of the sample observations were used for model evaluation (Qin et al. 2017). We also collected information on the representative sample vegetation cover and type surrounding citrus orchards and backyard farms where the ACT was present using geotagged photographs that were taken from the main four cardinal directions of the orchards for further inspection on how the landscape context could affect the presence of ACT.

3.3.3 Predictor variables

We considered 35 variables as potential predictors for estimating ACT distribution and habitat suitability. The variables were categorized into BCL and RS variables, see Table 1. For the BCL variables, we selected variables based on the ecological requirements of the vector as reported in previous studies: temperature, humidity, and elevation (Shimwela et al. 2016). Temperature and precipitation were represented by 19 “bioclimatic” variables, see Table 3.1, available from the WorldClim database (www.worldclim.org) (Hijmans et al. 2005). WorldClim projects current climatic conditions at 1-km spatial resolutions based on observations gathered from different weather stations between 1950 and 2000; the point datasets are interpolated using a thin plate smoothing spline algorithm to create a seamless raster dataset (Hijmans et al. 2005). We also used topographical variables related to the potential ACGD vectors’ habitat. This included elevation, slope, hill shade, and aspect in degrees, see Table 3.1. Hill shade was included as a proxy for relative solar radiation load that accounts for the effect of topographic shading (Pierce et al. 2005). We observed in the field that the

majority of ACT presence points were on the windward side for mountainous regions as opposed to the leeward side; hence, we included hill shade as a predictor variable in our ACT distribution model. The topographical variables were extracted from a void-filled 90 m digital elevation model (DEM) data set from the Shuttle Radar Topographical Mission (SRTM) (Jarvis et al. 2008). Using the Environment for Visualizing Images (ENVI) version 4.8 (Exelis Visual Information Solutions, Boulder, Colorado), both bio-climatic and topographical variables were resampled using a bilinear interpolation method to fit the 250m pixel size of the remotely-sensing variables (Usery et al. 2004).

RS variables on vegetation phenological metrics and vegetation productivity dynamics were derived from a Moderate Resolution Imaging Spectroradiometer (MODIS) Enhanced Vegetation Index (EVI) time-series data at a 250-m spatial resolution. MODIS products such as Normalized Difference Vegetation Index (NDVI) and EVI are the most widely used indices for monitoring of the vegetation phenological pattern (Li et al. 2010). Matsushita et al. (2007) pointed out that NDVI is easily affected by soil background and low vegetation coverage and easily saturated in high vegetation coverage. On the other hand, EVI minimizes the noise of soil background and adjusts atmospheric aerosol interference, thus improving the sensitivity of mimicking densely vegetated sites as compared with NDVI (Huete et al. 2002; Liu et al. ; Wang et al. 2006). In the present study, MODIS 16-day EVI composites for the years 2014 to 2016 from the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Center (LP DAAC—https://lpdaac.usgs.gov/) were downloaded and preprocessed using the MODISTools package in R (Tuck et al., 2014). MODISTools provides a function for mosaicking and sub-setting the downloaded data to a selected geographical extent. Then, we calculated 11 vegetation phenological metrics, see Table 3.1, using the TIMESAT software (Jönsson and Eklundh 2004). Namely, we calculated (1) start of the season (start of season) which is the time of initial vegetation green up, (2) end of the season (end of season) representing time of initial vegetation senescence, (3) the length of growing season from green up to senescence (length of season), (4) base level, which was calculated by averaging the left and right minimum values (base value) that represent the baseline of the seasonal phenology curve, (5) time for the middle of

the growing season (mid of season), (6) the highest EVI value of the season (max fitted value), (7) seasonal amplitude calculated as the difference between the peak EVI value and the average of the left and right minimum values corresponding to the amount of EVI change (amplitude), (8) the rate of vegetation green up (left derivative), (9) the rate of vegetation senescence (right derivative), (10) proxy for the relative amount of vegetation biomass without regarding the minimum EVI values (large integral), and (11) the proxy for the relative amount of vegetation biomass while regarding the minimum EVI values (small integral). All 11 vegetation phenological metrics (Jönsson and Eklundh 2004; Penatti and Isnard 2012; Wei et al. 2012) were calculated for the two growing seasons within each year. TIMESAT extracts vegetation phenological variables by fitting a local function to the time-series datasets (Cai et al. 2017). We fitted the Savitzky-Golay smoothing model function that replaces the data value by values in a window using a second-order polynomial function with optimum smoothing parameters (Chen 2004; Jönsson and Eklundh 2004). The Savitzky-Golay function reduces the effects of residual signals and smooths the time-series EVI dataset to a degree determined by the size of the smoothing window and reduces the noise caused primarily by cloud contamination and atmospheric variability (Cai et al. 2017). The start and end of season threshold parameters for the smoothing function were set at 20%, as suggested by Jonsson and Eklundh (2004), to optimize the error that could be caused by varying start and end of season dates in different locations across the study area (Makori et al. 2017). Only variables for the first season were used in this study since data from the second season were not consistent throughout all the years across the study area (Makori et al. 2017). Our study area cuts across different climatic zones in Kenya with a varying number of rainy seasons; hence, some of our sample sites commonly experience unimodal rainfall (one rainy season), while others have bi-modal rainfall (two rainy seasons) in a calendar year. This variability in the rainy seasons could have caused the variation and inconsistency in the vegetation phenological variable across the entire study area during the second rainy season. In addition to the vegetation phenological metrics, LST has proved to have a major influence on the spread and development of pests and diseases (Chabot-Couture et al. 2014). LST variables extracted from time-series MODIS data for

the years 2014 to 2016 were averaged for each year and included in the set of predictor variables. MODIS LST has a high spatial characteristic that enables the capture of the spatial variability of land surface fluxes within a finer scale as opposed to point observations taken on the ground.

Table 3.1 Predictor variables used for modeling the ecological niche for the African citrus triozid (Trioza erytreae). The variables were divided into two sets; environmental (bio-climatic and topographical) and remote-sensing variables. Bold text refers to variables which were selected through a multi-collinearity test using the Findcorrelation function in the caret package in the R software and finally used in the MaxEnt model.

Data source	Category	Variables description	Abbreviations	Units
WorldClim	Bioclimatic	Annual mean temperature	Bio 1	^{o}C
		Mean diurnal range (mean of monthly (max temp, min temp))	**Bio 2**	**^{o}C**
		Isothermality (Bio 2/Bio 7) (x100)	Bio 3	^{o}C
		Temperature seasonality (standard deviation x 100)	Bio 4	^{o}C
		Maximum temperature of warmest month	Bio 5	^{o}C
		Minimum temperature of coldest month	Bio 6	^{o}C
		Temperature annual range (Bio 5-Bio 6)	Bio 7	^{o}C
		Mean temperature of wettest quarter	Bio 8	^{o}C
		Mean temperature of driest quarter	Bio 9	^{o}C
		Mean temperature of warmest quarter	Bio 11	^{o}C
		Mean temperature of coldest quarter	Bio 11	^{o}C
		Annual precipitation	Bio 12	mm
		Precipitation of wettest month	**Bio 13**	**mm**
		Precipitation of driest month	Bio 14	mm
		Precipitation seasonality (coefficient of variation)	Bio 15	mm
		Precipitation of wettest quarter	**Bio 16**	**mm**
		Precipitation of driest quarter	Bio 17	mm
		Precipitation of warmest quarter	**Bio 18**	**mm**
		Precipitation of coldest quarter	Bio19	mm
SRTM	Topographic	**Ground height**	**Elevation**	**m**

		Sloping direction	**Aspect**	**degree**
		Steepness of the ground	Slope	degree
		Shading effect	Hill shade	n/a
MODIS EVI	Remotely sensed	Time for the start of the season	Start of season	decades
		Time for the end of season	End of season	decades
		Length of season from start to end	**Length of season**	**decades**
		Mid of the season	Mid of season	decades
		Difference between maximum and base level	Amplitude	n/a
		Average minimum EVI value	Base value	n/a
		Maximum fitted value	Max fitted value	n/a
		Rate of increase at the beginning of season	**Left derivative**	**%**
		Rate of decrease at the end of season	**Right derivative**	**%**
		Large seasonal integral	**Large integral**	**n/a**
		Small seasonal integral	**Small integral**	**n/a**
MODIS		**Land surface temperature**	**LST**	**°C**

3.3.4 Predictor variable selection

To examine the expected multi-collinearity among the predictor variables, we performed a Pearson correlation test (Figure 3.2) between all the predictor variables shown in Table 3.1. Furthermore, the 'Findcorrelation' function in the Caret package in R was used to eliminate highly correlated variables using the mean absolute error score. A correlation coefficient of $|r| > 0.7$ was set as a collinearity indicator for variables that would severely affect our model (Dormann et al. 2013). Variables that met this criterion were eliminated from the analysis and only the uncorrelated predictor variables were used in the MaxEnt model.

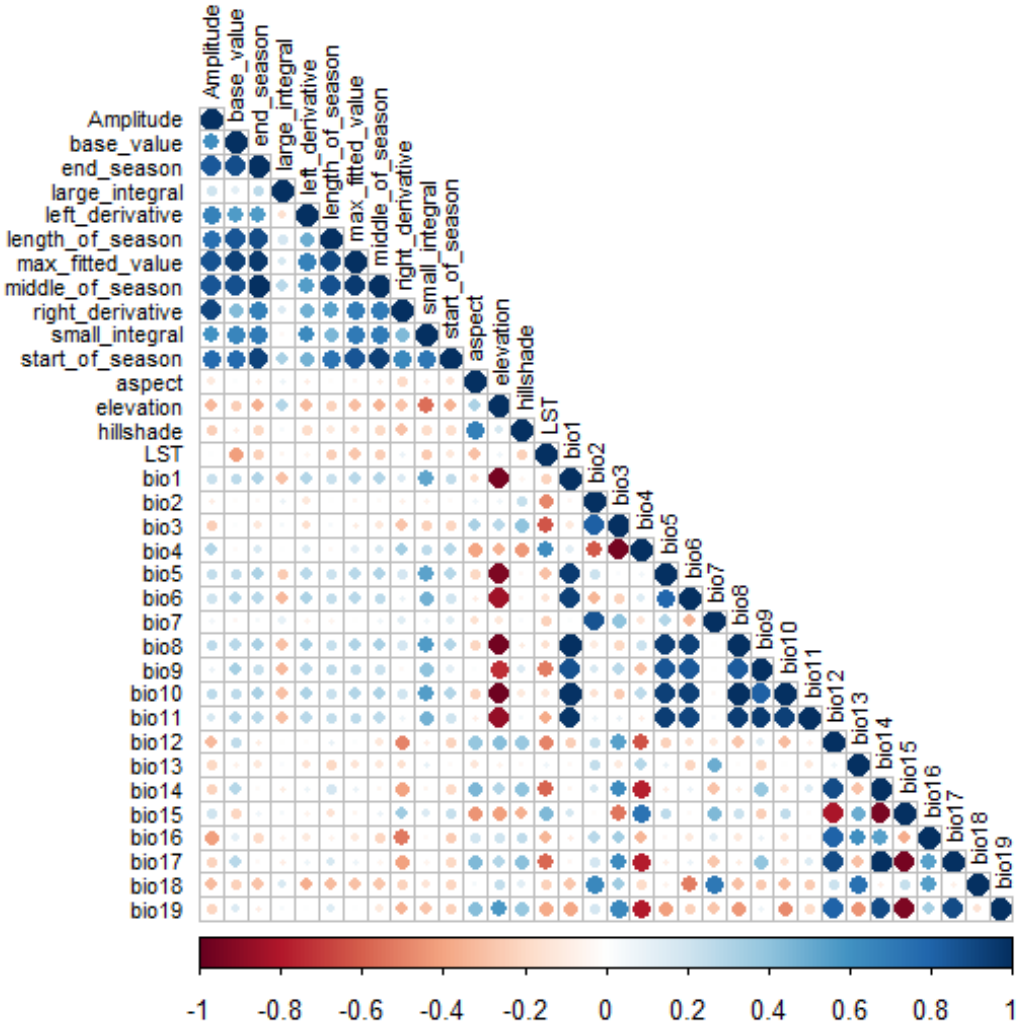

Figure 3.2. Collinearity matrix for predictor variables. Darker shades of blue and red color indicate high variable collinearity while light shades indicate low collinearity between variables.

3.3.5 EN Modeling

A MaxEnt model algorithm (Phillips et al. 2006) was used to predict the distribution and likely suitable sites for ACT. MaxEnt is a presence background machine-learning approach that estimates species' distribution that has maximum entropy subject to a set of constraints based upon a user's knowledge of the environmental conditions at known occurrence sites (Yackulic et al. 2013). Like most maximum-likelihood estimation methods, the MaxEnt algorithm adopts a uniform distribution and performs several iterations in which the weights related to the environmental variables are adjusted to

maximize the average probability of point localities. These weights are then used to compute the distribution over the entire geographical space (Buermann et al. 2008).

To minimize overfitting in the MaxEnt model, we implemented a regularization method to penalize the model in proportion to the coefficient magnitude (Royle et al. 2012). Further, we ran MaxEnt models using the default variable responses setting and a logistic output format which results in the ACT distribution suitability prediction ranging from 0 (less suitable) to 1 (highly suitable). However, a default regularization multiplier was doubled to reduce the chance of under or over prediction (Sahlean et al. 2014). In addition, we used the 10^{th} percentile training presence threshold which predicts the 10% most extreme presence observations as absent to eliminate 'outliers' from the final model (Cord et al. 2014). To study the effects of the vegetation phenology and dynamics for predicting ACT distribution, we performed two MaxEnt models, one included the environmental variables only (BCL model) and the other included both the environmental and remote-sensing variables (BCL-RS model).

3.3.6 EN Models Validation

Commonly, the accuracy of the MaxEnt distribution suitability maps is assessed using conventional accuracy measures such as the area under the curve (AUC) and chi-squared (X^2) statistics. However, these accuracy statistics are somehow biased and highly sensitive to the proportional extent of the predicted presence observations (Anderson et al. 2003), as a result of an overestimation of the pseudoabsence samples. Hence, in this study, we employed more reliable and adequate measures to evaluate the overall MaxEnt model performance. Specifically, we used true skill statistic (TSS) and Cohen's kappa coefficient (K_{hat}) to evaluate the accuracy of the ACT distribution suitability maps (Jorge 2011). As compared to TSS, kappa inherently depends on prevalence. However, an ideal measure of model performance should not be affected by prevalence but combine sensitivity and specificity (Allouche et al. 2006). Thus, TSS combines both sensitivity and specificity to account for both omission and commission errors and is not affected by prevalence and the size of the validation set and, therefore, is the best parameter to measure model performance. Both TSS and K_{hat} range from -1 to +1, where

+1 indicates perfect agreement between the observed and predicted ACT observations, whereas values <0 indicate no agreement or that most of the predicted ACT observations were produced by chance (Zhang et al. 2015). In addition, we used a Jackknife procedure to assess the relative importance of each individual predictor variable to the ACT distribution suitability model (Matyukhina et al. 2014). To test the null hypothesis that there was no statistical ($p \leq 0.05$) difference between the predictions of the BCL and the BCL-RS MaxEnt models, a two-sample t-test was performed. Herein, using 'ArcGIS create random points' tool, we generated 500 random sample points throughout the study area and compared their predictive power for each of the two models (i.e., BCL and BCL-RS).

3.3.7 Landscape Context Calculation

To describe the landscape context, we used a LULC map at a 20-m spatial resolution over the study area based on one year of Sentinel-2A observations ranging from December 2015 to December 2016 developed and validated by Climate Change Initiative (CCI) Land Cover (LC) team (ESA 2017). Since ACT is likely to spread locally up to a distance of 1500 m by natural dispersal (Van Den Berg 1988), we extracted the LULC proportion within a 1500 m radius buffer from the center for each of the 24 ACT occurrence points collected from the field which were not overlapped within each buffer, see Figure 3.3. The proportions of the four major LULC classes (tree cover, shrubs cover, grassland, and cropland) within each buffer were calculated. We hypothesize that these four major LULC classes could influence the occurrence of ACT within a landscape scale. The same buffers were also used to extract the corresponding average habitat suitability scores from the suitability map generated by the MaxEnt algorithm. Random forest (RF) regression (Breiman 2001; liaw and Wiener 2001) analysis was performed to determine the most relevant LULC classes for the ACT habitat suitability scores using the RF variable importance by-product. An RF regression model was performed using the default settings suggested by Breiman (2001), and the importance of the LULC classes was assessed using the RF mean decrease in accuracy (%) metric.

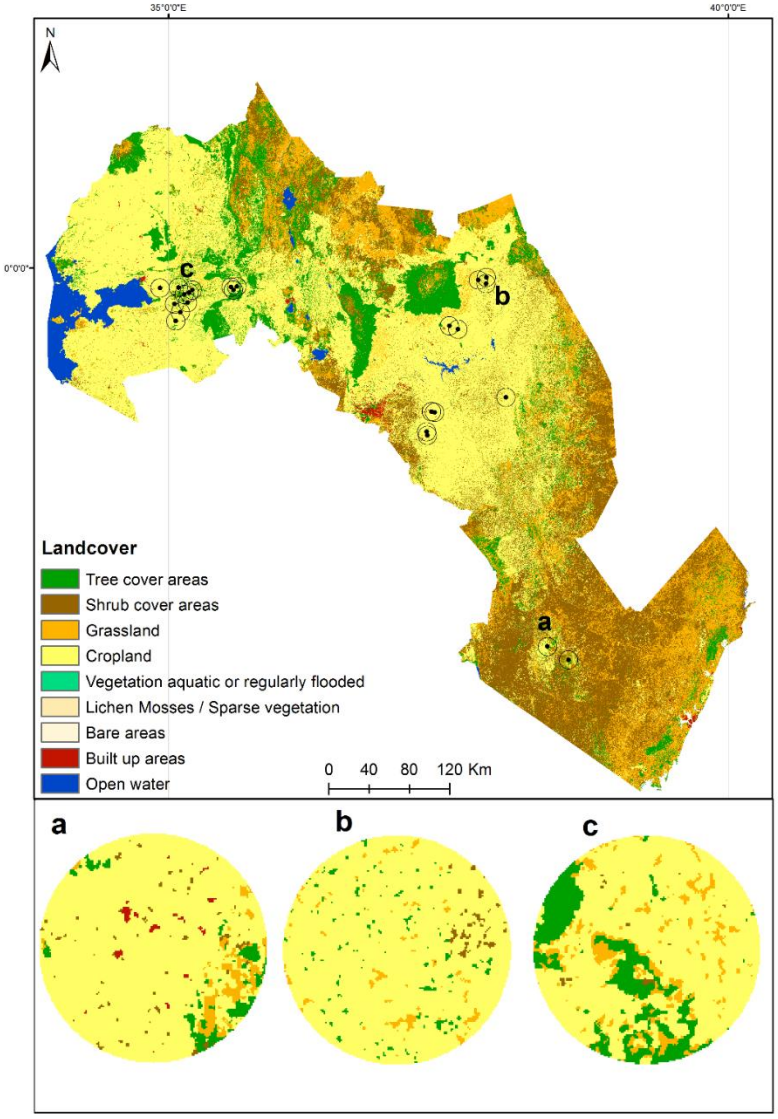

Figure 3.3 A 20-m spatial resolution land use/land cover map for the study area generated by the Climate Change Initiative (CCI) Land Cover (LC) team. Using yearly Sentinel-2 observations. (a), (b), and (c) represent zoomed buffers of 1,500 m radius each around certain representative African citrus triozid (ACT) occurrence points.

3.4 Results

3.4.1 EN Models

The Pearson correlation test for multi-collinearity resulted in selecting only six BCL and six RS uncorrelated predictor variables, respectively, see Table 3.1. The overall accuracy, TSS, and K_{hat} for both the BCL and BCL-RS MaxEnt models are shown in Table 3.2. A combined BCL-RS model gave the highest accuracy of 92% with a TSS score of 0.83

compared to the model with only environmental variables (BCL model), which had an overall accuracy of 85% and a TSS score of 0.572. TSS and K_{hat} statistics showed a prediction better than expected at random (TSS = 0.5) for both models, with the BCL-RS model performing better than the BCL model.

Table 3.2 Accuracy assessment statistics for the developed African citrus triozid (Trioza erytreae) MaxEnt models.

Model	Bio-climatic and topographical variables (**BCL, *n* = 6)**	Bio-climatic, topographical & remotely-sensed variables (**BCL-RS, *n* = 12)**
Overall accuracy	0.85	0.92
Sensitivity	0.73	0.91
Specificity	0.85	0.92
K_{hat}	0.30	0.42
TSS	0.57	0.83

3.4.2 Variable Importance

Figures 3.4 and 3.5 show the results of the jackknife test of variable importance for the BCL and BCL-RS models, respectively. Blue shades show the individual importance of each variable when used in isolation, while green shows the model performance when the variables are exempted from the model. The figures also show the variables which caused the greatest decreases in the gain when omitted, indicating that they provided a significant portion of information that was not contained in the other variables. For both models (BCL and BCL-RS), the variable with the highest gain (relevance) when used in isolation was Bio 18; therefore, Bio 18 appears to have the most useful information individually, followed by Bio 16 (for variable definitions see Table 3.1). Likewise, the variables that decreased the gain the most when they were omitted were Bio 16 and Bio

18 for the BCL model and Bio 16 and LST for the BCL-RS model. These variables appear to have the most influence on the models compared to the other variables.

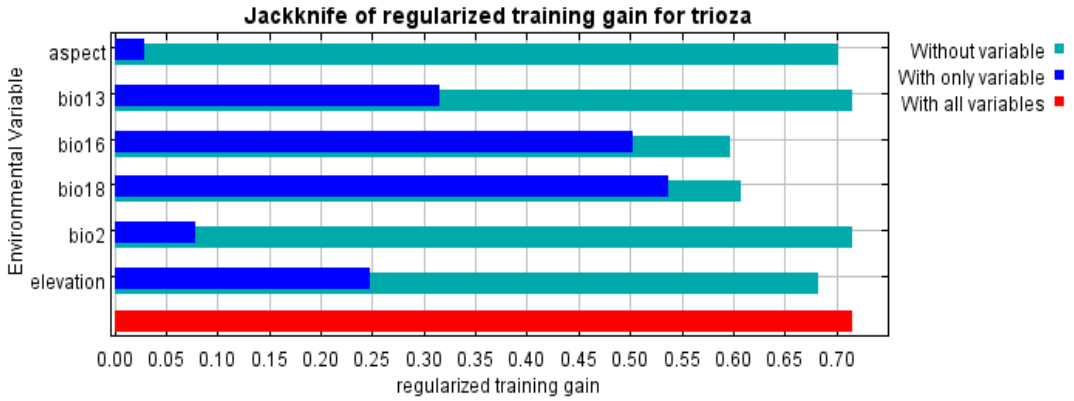

Figure 3.4 Jackknife variable importance test of regulated gains for the BCL model. The dark blue shades show the regularized training gain for the specific variable, light blue shows the relevance when the variable is omitted, while red shows the regularized training gain with all the variables combined.

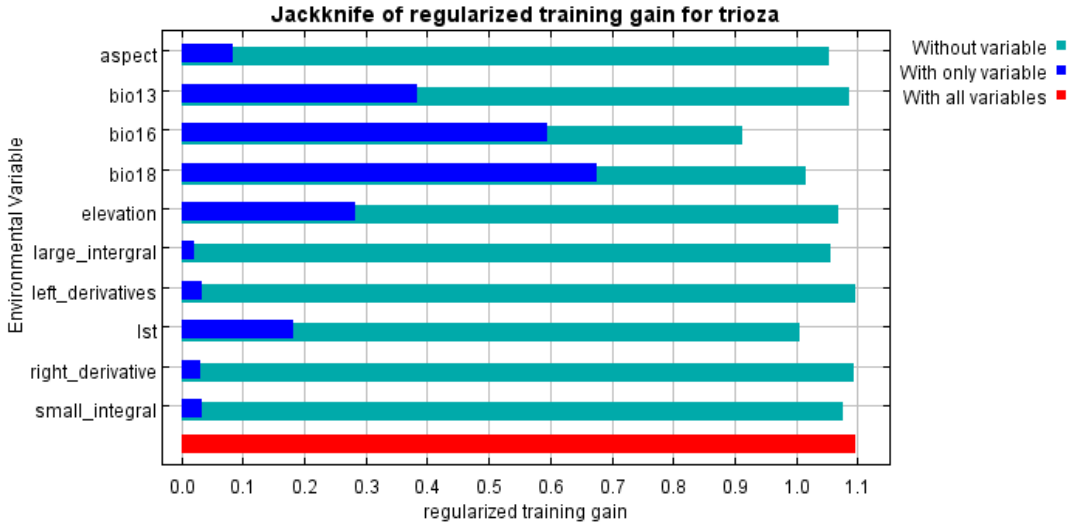

Figure 5. Jackknife variable importance test of regulated gains for the BCL-RS model. The dark blue shades show the regularized training gain for the specific variable, light blue illustrates gains without the variable, while red shows the regularized training gain with all the variables combined.

Table 3.3 presents the percentage that each variable contributed and its permutation importance in the BCL and BCL-RS models, respectively. In the BCL model, Bio 16 was the variable that contributed the most (48.3%) followed by Bio 18 (44.5%), Elevation

(4.0%), and Aspect (2.2%), respectively. Similarly, in the BCL-RS model, Bio 16 contributed the most (41%), followed by Bio 18 (36.3%), while the contributions for LST, Elevation, and Aspect ranged from 4.9% to 6.6%.

Table 3.3 Percentage contributions and permutation importance for each variable to the BCL and BCL-RS models, respectively.

Variables	Percent Contribution	Permutation Importance
BCL Model		
Bio 16	49.3	40.5
Bio 18	44.5	38.9
Elevation	04.0	18.3
Aspect	02.2	02.3
Bio 2	00.0	00.0
Bio 13	00.0	00.0
BCL-RS Model		
Bio 16	41.0	30.6
Bio 18	36.3	23.9
Land surface temperature (LST)	06.6	11.5
Elevation	05.3	07.0
Aspect	04.9	07.8
Small integral	02.8	03.9
Large integral	02.5	07.6
Bio 13	00.5	04.2
Right derivative	00.2	03.4
Left derivatives	00.0	00.1

3.4.3 Habitat Suitability Mapping

Figure 3.6 shows the predicted habitat suitability map for ACT based on the BCL, see Figure 3.6a, and the BCL-RS, see Figure 3.6b, models. The maps indicate the more suitable predicted sites with warmers colors (red) and less suitable predicted sites with cooler colors (blue). Both models show better predicted conditions in Western, Central,

and small parts of Eastern Kenya. These areas have a higher elevation above mean sea level. The least suitable sites are mostly towards the coastal region which has lower elevations.

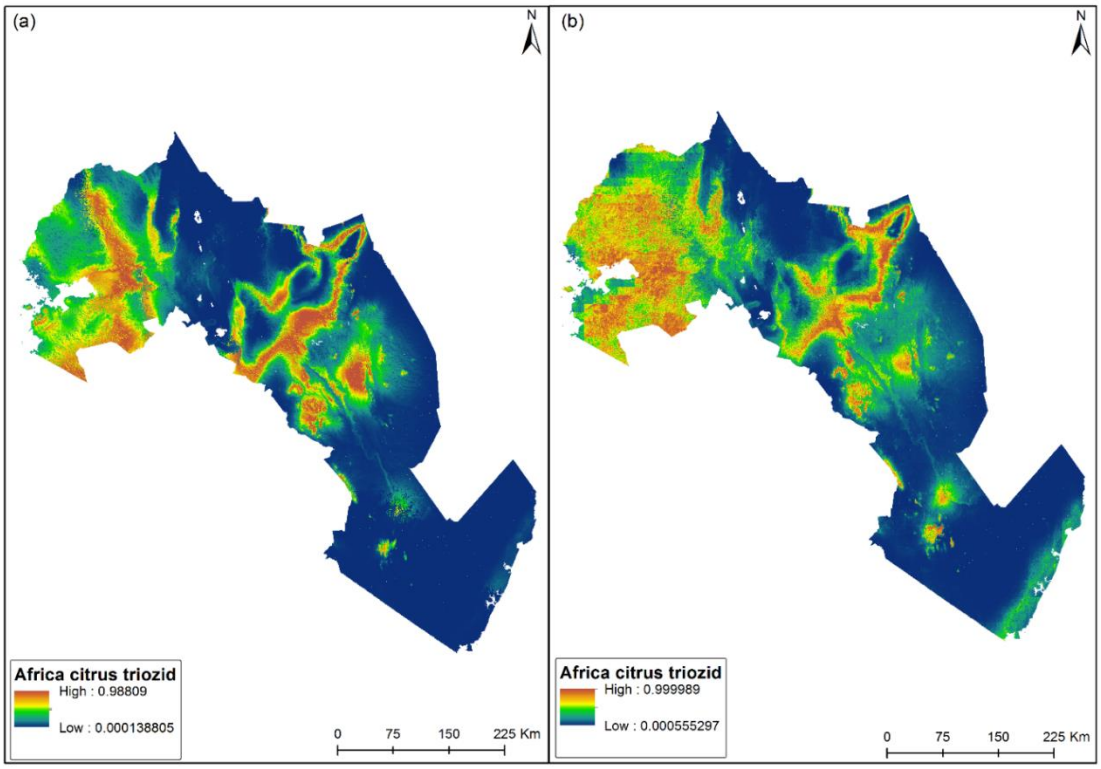

Figure 3.6 Predicted distribution suitability map for African citrus triozid (*Trioza erytreae*) using environmental (BCL model) variables (a), and environmental and remotely-sensed (BCL-RS model) variables (b). Blue indicates low distribution suitability, while red represents high distribution suitability.

The t-test result showed that the BCL-RS model produced significantly (t-statistic = 2.8279 and p = 0.005) higher AUC values compared to the BCL model. The t-test difference in the means, indicated that RS variables contributed 18% to the prediction model when combined with environmental variables.

3.4.4 Relationship between ACT Habitat Suitability and Landscape Context

We realized that there are diverse and multiscale responses of landscape context (i.e., LULC) to the habitat suitability of ACT. Using the mean decrease in accuracy (%) in the RF variable importance rank, the “shrubs” class was found to be the most relevant LULC

class to ACT habitat suitability followed by “trees”, “grassland”, and “cropland”, respectively, as shown in Figure 3.7.

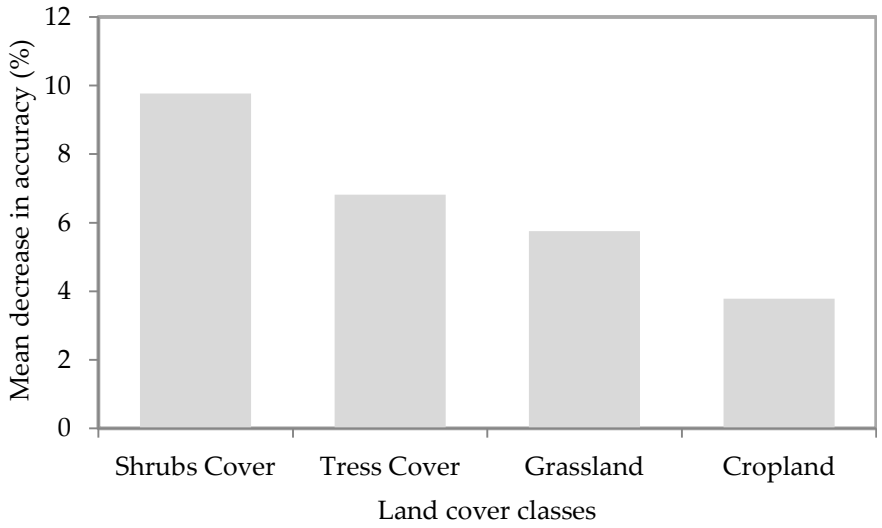

Figure 3.7 The relevance of the four major land use/land cover classes to the habitat suitability of African citrus triozid (Trioza erytreae) using a random forest variable importance rank.

3.5 Discussion

This study tests the applicability of an EN modeling approach for predicting the distribution and suitable habitat for ACT in citrus-growing regions in Kenya. This was achieved through nesting ACT habitat variables with a MaxEnt modeling framework for generating distribution information that is fundamental for prioritizing sites in which the management of ACGD is most needed or feasible (Forkuor et al. 2017). A reliable and accurate ACT distribution map is a valuable information source for monitoring vector infestation rates and disease spread. Such a spatial data set can also be used to prioritize interventions that prohibit the spread of the disease to unaffected areas (Tonnang et al. 2017). The “near-real-time” aspect of the remotely-sensing data means that the largely neglected aspect of early response can be addressed within integrated pest management (IPM) strategies (De Meyer et al. 2010).

In general, our study shows that both the uncorrelated BCL and RS variables were well-associated with the occurrence of ACT in typical Eastern African landscapes with their heterogeneous agro-ecologies. The results showed the importance of fusing

RS with BCL variables in reducing the overestimated spatial variability in the predictor variables and in enhancing the predictive power of the model (Makori et al. 2017). For the best performing models, Bio 16 (annual precipitation) had the highest contribution towards predicting the habitat suitability for ACT followed by Bio 18 (precipitation of the warmest quarter), LST, elevation, aspect, and small integral (MODIS-derived vegetation productivity). Precipitation of both the wettest and warmest quarter were important variables in defining the habitat suitability of ACT since they regulate the optimal temperature ranges within which the triozid survives (Bové 2014). In addition, precipitation and temperature regulate citrus flushing circles which are known to be highly correlated with the occurrence of ACT (Khamis et al. 2017). The significance of the precipitation related variables in describing habitat suitability for ACT was more pronounced than elevation, which has been linked with the distribution of the vector in a previous study (Shimwela et al. 2016). This could be due to the micro-climatic aspect which is not entirely dependent on elevation but also landscape heterogeneity, among other aspects. In the BCL model, Bio 16 and Bio 18 alone contributed more than 92% to the model performance, while in the BCL-RS model, the contribution from these variables was reduced to 77% indicating that inclusion of RS variables contributes immensely to the model. Since our aim was to start from known BCL variables that are commonly used to predict the spatial distribution of crop pests, then explore the contribution of RS variables to the predictive model performance, we did not create a model without bioclimatic variables. Also, we did not create any bootstrapped MaxEnt models, which could have allowed the quantification of the effect sampling variability had on our model results (ACT distribution map).

Makori et al. (2017) reported that RS information used within habitat suitability models is known to better account for explicit landscape patterns, that define habitats, thereby reducing model over-fitting and essentially increasing the accuracy and precision of habitat suitability models. In addition, our results showed that LST played a key role in defining the niche of the ACT vector. This is in agreement with previous studies which have shown LST to be a main parameter in pest modeling routines (Blum et al. 2015). The influence of RS variables in modeling the habitat suitability of ACT was

considerable. The BCL model, as shown in Figure 3.6a, had over-predicted the distribution of ACT compared to the BCL-RS model, as shown in Figure 3.6b. In our ACT prediction distribution map, areas with high occurrence probabilities are characterized by high precipitation, high elevation, lower temperature regimes, and relatively similar vegetation productivity patterns.

The ACT distribution maps using BCL-RS variables show high occurrences of ACT in specific locations of the coastal region of Kenya. This disagrees with the findings of previous studies that ACT is unlikely to be present in coastal ecosystems. This could be related to vegetation dynamics and landscape context (i.e., LULC), which are very distinct in some specific areas along the coastal region, such as Wundanyi sub-county in Taita-Taveta county where the habitat suitability was reported to be high compared to other coastal regions of Kenya using the MaxEnt model. Despite the climate conditions being very similar to other regions where the model has predicted a high suitability of ACT, vegetation patterns in regions where the habitat suitability is high are very distinct and of similar productivities since they have common climatic conditions in terms of rainfall and temperature. This is in alignment with the finding from the literature that vegetation dynamics play a key role in defining the niche of crop pests and diseases (Ratnadass et al. 2012). This result reinforced the importance of both BCL and RS variables for modeling the distribution of ACT. Further, our study is a step towards the understanding of how the spread of insect pests is enhanced by BCL (both bio-climatic and topographic) and RS (vegetation phenological variables and LST), that influence the spread and multiplication of the vector in African agro-ecosystems.

Furthermore, the results from this study revealed that landscape context should not be ignored regarding understanding the distribution and dispersal pattern of ACT. However, we did not include landscape context in our MaxEnt model since from our field observation, we realized that the majority of the citrus orchards in our study area are within a cropland class. In our case, a presence-only MaxEnt model would have extracted only ‘cropland’ features for all ACT presence points. Therefore, we opted to look at the effect of the landscape context on ACT habitat suitability based on the dispersal capability of the pest (which is 1500 m). The relationship between ACT habitat

suitability and the four major LULC classes across the major citrus-growing regions showed that there is an association between the surrounding shrub cover proportion and habitat suitability for ACT. Shrub cover near citrus orchards could provide alternative host plants for the vector during the time when citrus trees are not flushing since ACT is correlated with the flushing rhythm of the citrus host (Cook et al. 2013). In addition, from field observations, it became clear that the majority of the ACT-infected citrus trees were within shaded areas, and thus trees and shrubs surrounding the citrus orchards most likely provide more suitable temperature conditions for the survival of ACT.

To the best of our knowledge, our study is the first attempt to predict the distribution of ACT using an enhanced and optimized EN modeling algorithm with BCL and RS variables and habitat suitability relationships with the surrounding landscape classes. Previous studies have only investigated the role of various environmental variables for mapping ACP distribution, but in these studies, links between localized factors captured in more sophisticated modeling routines and better consideration of landscape patterns were not considered (Shimwela et al. 2016). Future studies should explore the relationship between vegetation phenological and other localized pest classification factors and ACT densities (i.e., number of insets per unit area) to better understand the survival and dispersal patterns of the vector as there is a need for a better and more concerted implementation of vector management practices.

3.6 Conclusions

The impact of spatially heterogeneous environmental factors on ACT population dynamics are complex to model. However, understanding the inter-relationship between vectors, hosts, and their niches environment can provide valuable information for identifying conditions suitable for pathogen introduction and transmission in citrus-growing regions. By exploring the spatial distribution of ACT, we identified a set of BCL factors that are favorable for its development, predicted its spatial occurrence, and identified potential areas that, due to their BCL conditions, would be suitable for its introduction.

The BCL-RS model showed higher accuracy metrics and was deemed appropriate for predicting the distribution and potentially suitable areas for ACT. Though less important, the influence of vegetation phenological variables and LST for determining the habitat suitability of ACT was considerable. Our results revealed that apart from the BCL variables like temperature, rainfall, and elevation, which have previously been found to define the EN of ACT, vegetation patterns and dynamics at a landscape level play a key role in influencing vector-host-pathogen transmission and distribution. The ACT distribution prediction maps are an important tool for identifying risk zones and understanding risk drivers. Also, the distribution maps can provide baseline information for the development and implementation of effective IPM strategies. Future studies should look at modeling the density of ACT on a landscape scale for the precise application of prevention and control measures.

4 ESTIMATING MAIZE LETHAL NECROSIS SEVERITY IN KENYA USING MULTI-SPECTRAL HIGH TO MODERATE RESOLUTION SATELLITE IMAGERY

4.1 Abstract

Maize Lethal Necrosis (MLN) is a serious disease in maize that significantly reduces yields up to 90% in major maize growing areas in Kenya and other countries in Africa. The disease causes chlorotic mottling of leaves and severe stunting which ultimately leads to plant death. The spread of MLN in the maize growing regions of eastern Africa has intensified since the first outbreak was reported in Kenya in September 2011. In this study, 30-meter Landsat-8 and 5-meter multi-temporal RapidEye (RE) imagery was combined with field-based assessments on MLN infection rates to map three MLN severity levels in Bomet County, Kenya. Two RE level 3A images, acquired during maize stem elongation and inflorescence stage, respectively, were combined with one cloud free Landsat-8 image (path 169, row 061) acquired during the maximum phenology stage. The Landsat was re-sampled to 5 m pixel resolution to fit the RE imagery using nearest neighbor re-sampling. Thirty spectral indices for each RE time step were computed and included in the mapping model. Machine learning using random forest classification, was used on the fused satellite data sets to create a map separating maize fields from all other land cover and land use classes. Subsequently, MLN severity levels (mild, moderate and high) were mapped using random forest. Integrating RE and Landsat-8 data improved the classification accuracy for separating maize croplands from non-cropped areas. The optimized random forest algorithm yielded an overall accuracy of 90.2%, representing high model performance, in predicting the MLN severity levels by combining mild and moderate severe classes into a single class. These results indicate the possibility of using time-series of multi-sensor satellite data (with high pixel resolutions) and machine learning to monitor the spatial distribution of disease infestation rates in fragmented agro-ecological landscapes.

4.2 Introduction

Agriculture is the primary source of livelihood for the most developing countries in Africa. In Kenya agriculture contributes up to 65% of the total labour force and contributes up to a third to the Kenyan gross domestic product (GDP) (World Bank 2008; Omiti et al. 2009). The population of Kenya is rapidly growing and expected to reach 95 million people by 2050 from a current size of 46 million. The resultant increase in food demand has profound implications on agricultural productivity and how intensive land can be used for a more sustainable way of food production (FAOSTAT, 2013).

Maize (*Zea mays* L.) is the main staple food in sub-Saharan Africa (SSA), covering over 25 million ha of small holder farmers in the region (FAOSTAT 2010). Maize is the main staple of over 85% of the population in Kenya. In 2016 Kenya produced 3.34 million tonnes down from 3.68 million tonnes in 2015, which constitutes an approximate 12% drop in production. The latter is related to a wide range of abiotic and biotic risk factors that can have far reaching consequences for the country's agricultural systems in the future (Ochieng et al. 2016). The former risk factors include uncertain changes in rainfall and temperature patterns markedly threatening food production, the latter pests and diseases that are likely to intensify under such extreme weather anticipated under climate change and in the absence of effective mitigation strategies (Myers et al. 2017).

Among the many pests and diseases affecting maize farming in Kenya and beyond, maize lethal necrosis (MLN) has emerged as a serious disease threat to maize production across SSA (Hilker et al. 2007). It was first reported in Kenya in September 2011 in the Longisa division within Bomet County (Adams et al. 2013; Wangai et al. 2018). During 2011 MLN spread to other major maize growing regions along the Rift Valley and the western part of Kenya towards Lake Victoria. In the following year the Kenyan Ministry of Agriculture reported a major drop in maize harvest caused by MLN. Since 2012 the disease has also spread rapidly into other countries in East Africa such as Tanzania, Uganda, Rwanda and Ethiopia, leading to a serious reduction in maize production across the region (Adams et al. 2014; Mahuku et al. 2015).

MLN is caused by the combined effects of a maize chlorotic mottle virus (MCMV) infection and that of any virus from the Potyviridae family, mainly the

sugarcane mosaic virus (SCMV) (Deressa 2017; Hilker et al. 2007). MCMV can be transmitted by several insect vectors among which maize thrips, leaf beetles and leaf hoppers play a major role (Jiang et al. 1990; Nault et al. 1978; Yu et al. 2014). Moreover, MCMV can also be seed-borne, leading to disease transmission, albeit at very low rates (Jensen et al. 1991). SCMV is mainly spread by several aphid species and by mechanical means (Adams et al. 2012). MLN infected maize plants show a range of symptoms, like yellowing of leaves, mottling of the leaves leading to premature plant death, failure to tassel resulting in warped maize cobs with little to no seeds (Ochieng et al. 2016), depending on the number of viruses infecting the plant and time of infection (Castillo and Hebert 1974). Early symptoms of an infection are leaves turning yellow and orange, that later change to brown before completely drying up.

Several actions have been proposed for MLN management. For instance, farmers are advised to uproot affected plants during early growing stage to ensure that the disease does not spread. Another mitigation measure is crop rotation (Uyemoto 1983) though farmers are often reluctant to change from maize to other crops because of profitability reasons and since it is the most important staple food in Kenya (Muthoni and Nyamongo 2010). With this regard, detailed information on MLN disease severity, incidence and the related effects to quality and quantity of maize production are important prerequisites for improved disease management (Mahlein 2015). For instance, visual estimates of disease symptoms in the field can determine disease severity and incidence (Bock et al. 2010; Newton and Hackett 1994), though such an approach is expensive, time consuming and often insufficiently accurate because of the human bias (Benson et al. 2015). Consequently, there is a growing demand for more precise and automated methods of plants disease monitoring to mitigate disease outbreaks by enabling timely adoption of relevant management practices (Geerts et al. 2006; Palaniswami et al. 2014).

Remote sensing techniques have demonstrated a high potential in detecting the presence and monitoring the spread of agricultural pests and diseases (Mahlein 2015) as they can induce crop physiological stress and bio-physical changes on the infested plant leaves that can alter the reflectance spectra of plants (Fang and

Ramasamy 2015). Using plant leaves and canopy spectral signature, remote sensing can complement field-based protocols in distinguishing between healthy and different levels of damage of infested plants (Albayrak 2008). Moreover, remote sensing technologies can reveal the spatial and spectral distribution of pests and diseases over large areas at relatively low cost. For instance, Song et al. (2017) evaluated sentinel-2A satellite imagery for mapping cotton root rot, demonstrating that the technique can be used for precise disease identification if the imagery is taken during the optimum root rot discrimination period. Zhang et al. (2016) used a two-date multispectral satellite imagery for accurately mapping damage caused by fall armyworm (*Spodoptera frugiperda*) in maize at a regional scale. Franke and Menz (2007) evaluated high resolution QuickBird satellite multispectral imagery for detecting powdery mildew (*Blumeria graminis*) and leaf rust (*Puccinia recondita*) in winter wheat, showing that multispectral images are generally suitable to detect infield heterogeneities in plant vigor, particularly for later stages of fungal infections, but are only moderately appropriate for distinguishing early infection levels.

This study evaluated the potential of space-borne RapidEye (RE) multi-temporal data and an advanced Random forest (RF) classification technique in mapping MLN severity levels in a complex, dynamic and heterogeneous landscape, typical for rural SSA. This information will be important not only for a better understanding of the progression of the disease over a relatively large area, but also for the formulation and implementation of site-specific strategies for effective control of MLN.

4.3 Study area

The study was conducted in Bomet and Nyamira Counties located 300 km northwest of Nairobi, Kenya. The study sites lie between 34.97°E to 35.06°E and -0.76°S to -0.83°S (Figure 4.1) with an elevation range of 1,800-3,000 meters above sea level (m.a.s.l.). Bomet falls in a semi-humid climatic zone with mean monthly temperature of 18°C and a bi-modal annual rainfall ranging between 1,100-1,500 mm (Jaetzold and Schmidt 1982). The climate is suitable for growing a wide range of crops. However, maize and tea are the most dominant crops in the region, with majority of farmers practicing a

maize based mono-cropping system, especially in the southern part of Bomet County (Abdel-Rahman et al. 2017).

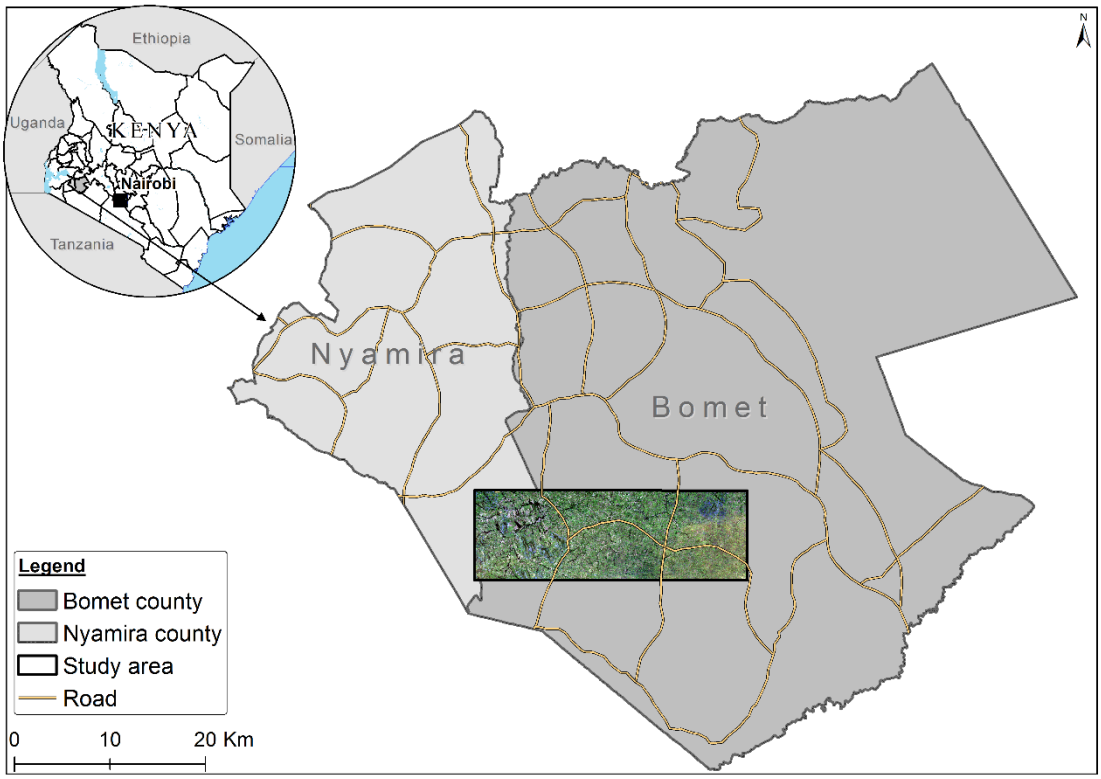

Figure 4.1. Location of the study area in the Bomet and Nyamira Counties of Kenya.

4.4 Methodology

Figure 4.2 summarizes the methodological approach for mapping severity levels of MLN. A two-step hierarchical RF classification using bi-temporal RE and Landsat-8 imagery was employed. First a land use/land cover (LULC) classification map was generated to delineate cropland from other LULC classes. We used the extracted maize crop mask from the first step to classify different MLN severity levels in maize fields (viz mild, moderately and highly infected maize plants).

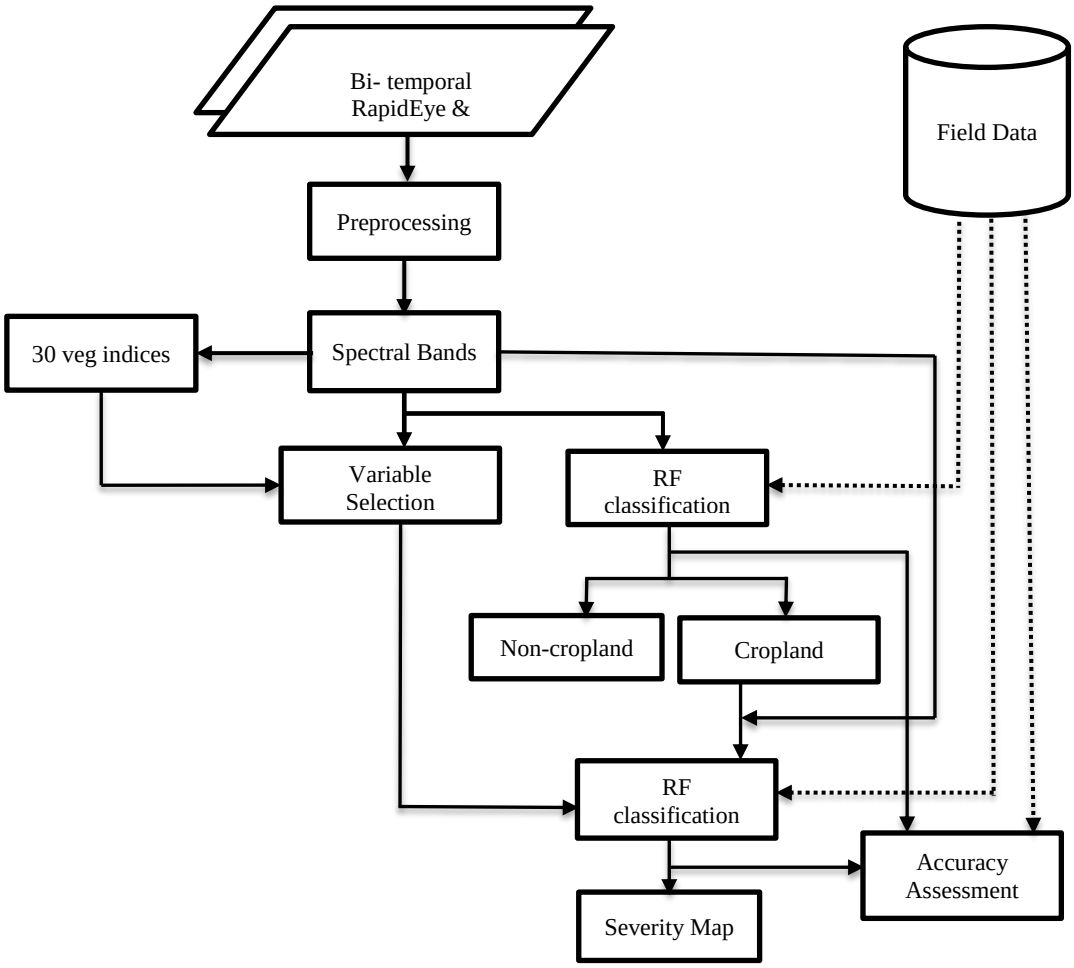

Figure 4.2 Flow chart of the two-step random forest classification for mapping maize lethal necrosis severity levels.

4.4.1 Field data collection

A field campaign was conducted to identify different LULC classes from the study area and MLN severity levels for each maize farm sampled. A stratified random sampling was followed to collect both the LULC and maize fields' severity reference data. A handheld Global Positioning System (GPS) device with an error of ±3 m was used to locate the reference control points. Once a field was identified, we delineated the field boundaries (polygon) within a minimum area of 10 ×10 m. To avoid an edge effect, we collected the polygon data five meters away from the edge of each field. To mitigate the effect of soil background on the crop spectral features, we only sampled the field crops that were about three weeks old at the first image acquisition date. The reference data for both LULC classification and MLN severity mapping were randomly divided into 70% training

and 30% validation sets. The training set was used to train the RF classifier while the validation dataset was used to evaluate the accuracy.

Disease severity scores were determined by frequent field visits of MLN affected farms by a plant pathologist to access damage levels (Campbell and Neher 1994). For each sampled farm the maize plants were grouped in specific severity levels based on damage levels and visual inspection. The severity was rated using a scale of 0-5 as described by Paul and Munkvold (2004). The six scales were 0 (no disease), 1 (10-20% leaf area affected by the disease), 2 (21-40% leaf area affected by the disease), 3 (41-60% leaf area affected by the disease), 4 (61-80% leaf area affected by the disease) and 5 (81-100% leaf area affected by the disease). For consistency MLN damage levels were classified as mild (<20%), moderate (20-80%) and high (>80%) (Nutter and Schultz 1995).

4.4.2 RapidEye data pre-processing

Two RE images were tasked for the Bomet site on the 9 December 2014 and the 27 January 2015, during the maize stem elongation and the inflorescence stages, respectively. RE is a commercial optical earth observation mission that consists of a constellation of five satellites with 5 m resolution and a swath width of 77 km with a revisit cycle of 5.5 days at nadir (RapidEye 2013). RE imagery is provided in five optical bands in the 400-850 nm range of the electromagnetic spectrum. The images used in this study were delivered as 3A orthorectified products in the form of 25 × 25 km tiles georeferenced to the Universal Transverse Mercator (UTM) projection.

Atmospheric correction was performed for each RE tile independently using the atmospheric-topographic correction (ATCOR 3) software (Guanter et al. 2009). This application provides a sensor specific atmospheric database of look-up-tables (LUT) which contain results of pre-calculated radiative transfer calculations based on the MODTRAN-5 (MODerate resolution atmospheric TRANsmission) model (Berk et al. 2008). All images were co-registered (image-to-image) to ensure the alignment of the corresponding pixels. Subsequently the RE tiles were mosaiced into a single image file for each acquisition date. For each image 30 spectral vegetation indices (SVIs) (kyalo et al. 2017) (see table 1.1) were computed and combined with the original RE bands (blue,

green, red, red edge and near infra-red) as input predictor variables for each acquisition to improve the MLN severity levels classification accuracy.

4.4.3 Landsat-8 data and pre-processing

Single date Landsat-8 Operational Land Imager (OLI) data for the study area delivered as orthorectified and level 1 terrain corrected were acquired from the United States Geological Survey (USGS) direct archive on 18 March 2015 during the maize flowering stage. The Fmask tool (Zhu and Woodcock 2012) was used for automated cloud and cloud shadow masking. All the image tiles covering the study area were converted to surface reflectance and digital numbers (DN) and converted to top-of-atmospheric (TOA) reflectance using the reflectance rescaling factors provided in the Landsat OLI file (Yackulic et al. 2013). Finally, the mosaiced image was atmospherically corrected using ENVI FLAASH (the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) which runs on a MODTRAN 4 based atmospheric correction algorithm (Cooley et al. 2002; Zhu et al. 2012). The spectral bands used in this study included blue (0.45–0.51 μm), green (0.53–0.59 μm), red (0.64–0.67 μm), near infrared (NIR; 0.85–0.88 μm), shortwave infrared 1 (SWIR 1; 1.57–1.65 μm; SWIR 2; 2.11–2.29 μm) and the panchromatic band (0.5–0.68 μm). The final atmospherically corrected tiles were mosaiced and resampled to 5 m pixel resolution using the nearest neighbor resampling technique of the Environment for Visualizing Images (ENVI) software.

4.4.4 Random forest algorithm

We employed the RF machine learning classifier (Breiman 2001; Rodriguez-Galiano et al. 2012) to predict the LULC classes, used for crop/masking mask, and also in mapping the MLN severity levels. RF was chosen as the preferred classification method since it has been proven to be robust to outliers, noise and consistently demonstrated capability to handle high dimensional datasets without suffering from overfitting (Barrett et al. 2014). RF builds an ensemble of individual decision trees from which the final prediction is based using majority voting criteria. Each decision tree is trained using a bootstrap sample consisting of two thirds of the training data drawn with replacement, and the remaining one third of the data, which is not included in the bootstrapped training

sample, is used to test the classification and estimate the out-of-bag (OOB) error (Barrett et al. 2016).

Random forest uses two user-defined parameters, the number of trees (*ntree*) and the number of variables used to split the nodes (*mtry*). The default *ntree* is 500, while the default value for *mtry* is the square root of the total number of spectral bands used in the study. To improve the classification accuracy, the two RF parameters were optimized based on the OOB error rate.

4.4.5 Variable selection and optimization

Random forest measures the importance of each predictive variable using the mean decrease in accuracy that is calculated using the OOB sample data (Georganos et al. 2018). However, the challenge was to select the fewest number of predictors that offer the best predictive power. In this regard, a backward feature elimination method (BFE) integrated with RF regression as part of the evaluation process was implemented (Guyon and Elisseeff 2003; Mutanga et al. 2012). The BFE uses the ranking to identify the sequence in which to discard the least important predictors from the input data sets. The method starts with the entire variables and then progressively eliminates the least promising variable from the list. For each iteration, the model is optimized by selecting the best *mtry* and *ntree*, using a grid search and a ten-fold cross validation method (Huang and Boutros 2016). The least promising variable is eliminated, and the root mean square error is calculated. The smallest subset of variables with lowest root mean square error (RMSE) is then selected for the final classification model.

4.4.6 Accuracy assessment

Classification accuracy of the RF classifier was assessed using an independent set of field data (30%). Overall accuracy (OA) and the F1-Score values were computed from the confusion matrices to evaluate the accuracy of generated classes. In addition, the producer's accuracy (PA) and the class specific user's accuracy (UA) was also calculated to evaluate generalization ability of the RF classifier (Congalton 2001). A confusion matrix provides information on the correct predictions by comparing the classified map with ground information collected from the field. OA refers to the ratio of correctly classified pixel to all pixels considered in the model evaluation. The F1-Score is a per

category measure which corresponds to the harmonic mean of the user's accuracies and the producer's accuracies. PA refers to the error of omission which expresses the probability of a certain class being correctly recognized, while UA is the error of commission which represents the likelihood that a sample belongs to a specific class and the classifier accurately assigns it this class. Kappa statistics were also calculated to compare the significance between different error matrices generated from the generated classification results (McHugh 2012). The Kappa coefficient measures the actual agreement between the reference data and a random classifier with a value close to one, signifying perfect agreement. To remove noise from the classification a 3×3 cell majority filter for all classified maps was applied. This approach replaces secluded cells with the classes that matches the majority of cells within a 3×3 matrix. Each filtered classified map was finally tested for accuracy.

4.5 Results

4.5.1 Random forest optimization

RF parameters (*ntree* and *mtry*) were optimized for the two-step classification for the different data sets using grid-search technique with 10-fold cross validation. The *ntree* value of 500 and *mtry* value of 3 setting yielded the least OOB error for the LULC classification. In addition, the *ntree* value of 1,000 and *mtry* of 5 yielded the best OOB error (4.8%) for the MLN severity mapping (Figure 4.3).

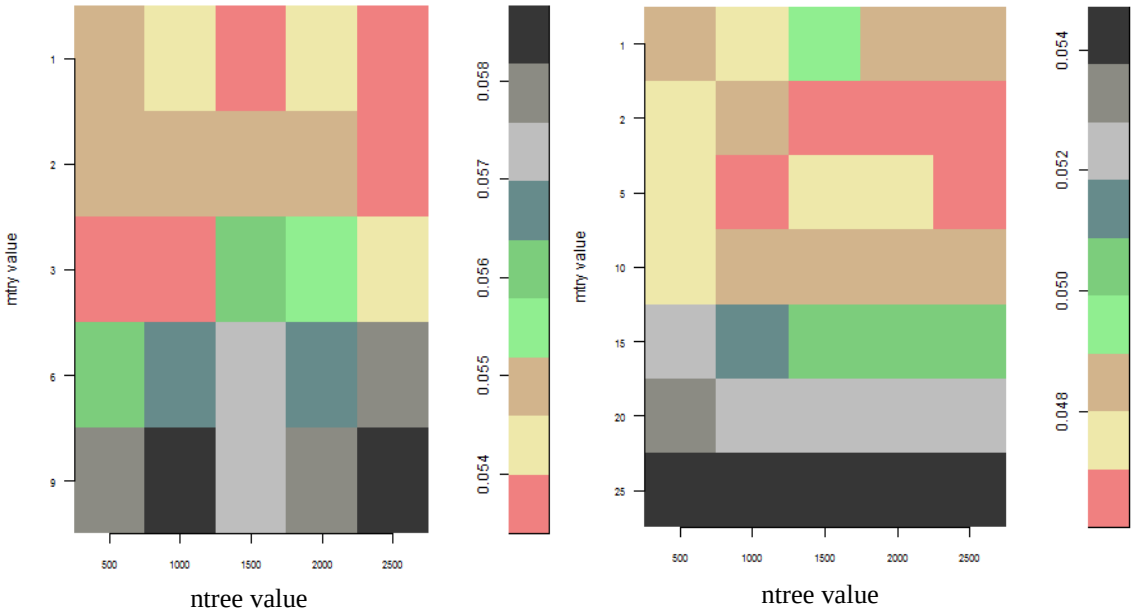

Figure 4.3 Results of Random forest optimization grid for the land use land cover classification result (a) and for the maize lethal necrosis severity mapping result in (b). The internal out-of-bag error rate was calculated using a 10-fold cross validation and the training data.

4.5.2 Crop masking

A LULC map was predicted to discriminate maize crop from other LULC classes using different combinations of the acquired RE images and the Landsat-8 image. Six major LULC classes were identified based on field observations made within the study area. Table 4.1 presents a summary of the results (overall accuracy and F1-Score) when using 30% as evaluation data. The results revealed that the use of the RE spectral bands gave an overall accuracy of 72.3% and 74.8% for single classification of RE1 and RE2, respectively. The combination of the two RE images (RE1 and RE2) spectral bands improved the overall accuracy to 80.6% while when combining the two RE images (bands) with the Landsat-8 imagery (bands) yielded a further overall accuracy improvement to 91.05%. In addition, the F1-score for each class was generally above 0.9 except for cropland and natural vegetation classes which were 0.87 and 0.85, respectively, for the optimal classification that combined RE images and Landsat-8 was selected for LULC.

Table 4.1 Overall and classwise accuracies for land use/land cover mapping using 30% of test data.

		Class wise accuracies- F1 score					
Image	**Accuracy (%)**	**Cropland**	**Forest**	**Grassland**	**Natural veg.**	**Soil**	**Tea**
RE1	72.34	0.69	0.81	0.69	0.74	0.83	0.81
RE2	74.80	0.73	0.87	0.84	0.77	0.84	0.70
RE1+RE2	80.63	0.82	0.92	0.89	0.85	0.78	0.82
RE1+RE2+ LS-8	91.05	0.87	0.93	0.95	0.83	0.96	0.93

Figure 4.4 shows the LULC map generated from the optimal combination of the two RE images with the Landsat-8 image revealing that cropland and grassland are the major classes in the study site, with few tea plantations on the northern side of the study area. However, there was slight confusion between cropland with natural vegetation and forest resulting from the presence of big trees and pockets of bushes inside the cropland as observed from the field.

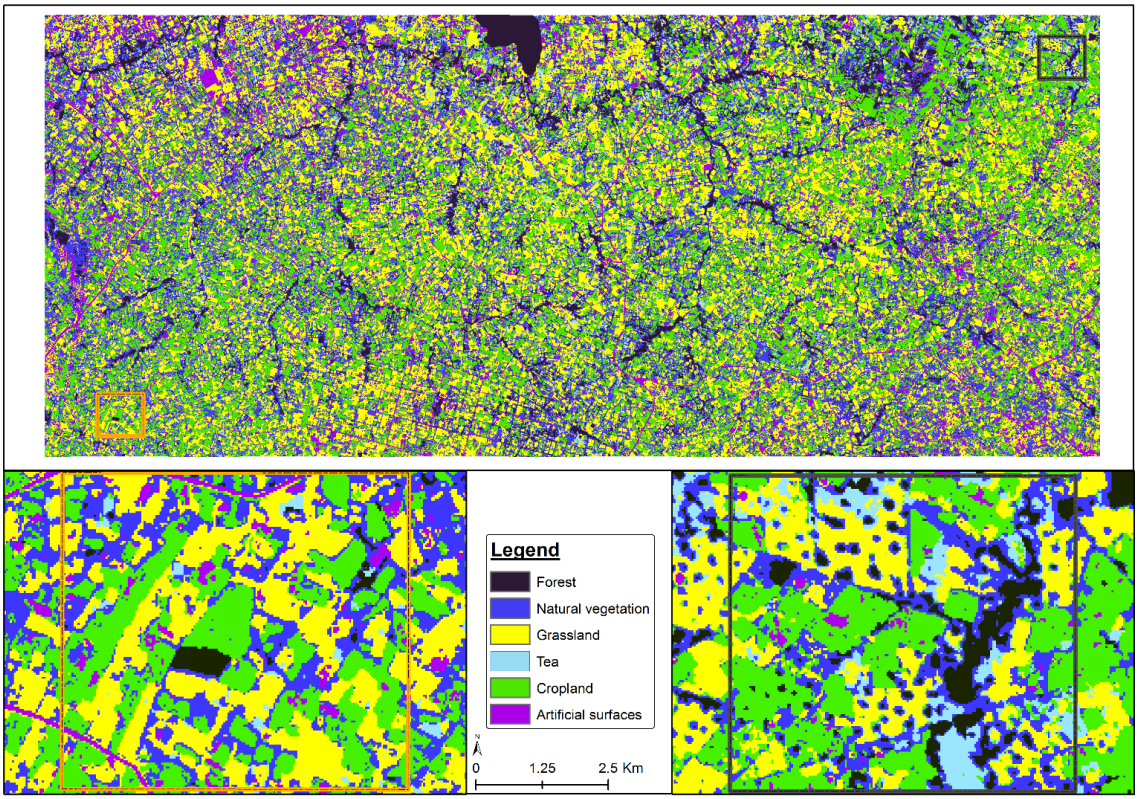

Figure 4.4 Land use land cover map obtained using Random forest classifier with two early season RapidEye images and one late season Landsat-8 image.

Table 4.2 represents the confusion matrix for the per-pixel evaluation for the LULC classification using RF. In general, all LULC classes achieved > 90% user's accuracy except for natural vegetation which had 79.89% due to spectral confusion with cropland. All LULC classes achieved > 90% producer's accuracy except for cropland and natural vegetation classes which achieved 83.23% and 86.34%, respectively. Consequently, the user's accuracy was generally > 90% for all classes except the natural vegetation class which had 79.89%. The latter can be attributed to the observed confusion between the "cropland" and "natural vegetation" classes as shown in the confusion matrix.

Table 4. 2 Confusion matrixes for land use land cover classification using Random forest classification with two early season RapidEye images (RE1 and RE2) and one late season Landsat image.

Class	Cropland	Forest	Grassland	Natural veg.	Soil	Tea	Total	PA (%)
Cropland	268	0	1	48	0	5	322	83.3
Forest	0	303	0	7	0	12	322	94.0
Grassland	2	0	304	4	12	0	322	94.1
Natural veg.	18	12	8	278	0	6	322	86.34
soil	9	0	5	3	305	0	322	94.72
Tea	0	13	0	8	0	301	322	93.48
Total	297	328	318	348	317	324	1932	
UA (%)	90.24	92.38	95.60	79.89	96.22	92.90		

Overall accuracy=91.05%. Kappa=0.89

4.5.3 Maize lethal necrosis severity variable selection

We used a RF backwards feature elimination procedure to identify the smallest set of predictor variables that resulted in the best predictive abilities of the RF model for mapping MLN severity levels. The progressive removal of the least important predictor variables resulted in the selection of seven spectral variables (indices and/or bands) which gave the least OOB error as shown in Figure 4.5. These reduced models were compared to models based on the full predictor variables dataset.

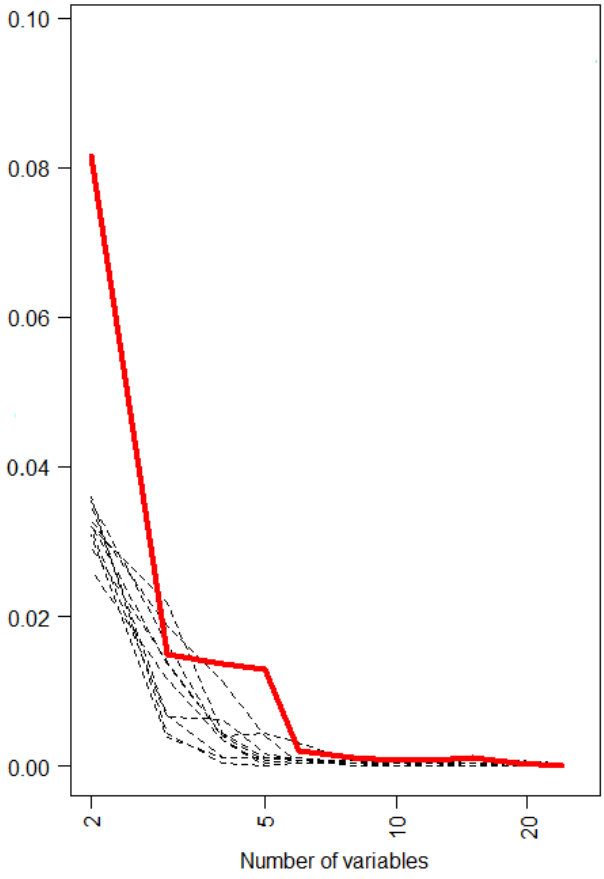

Figure 4.5 Plot for the optimal number of predictor variables selected based on the random forest backward feature elimination search function using root mean square error.

Four and three spectral variables, respectively, were selected as important variables from the two RE images, captured during the maize stem elongation and inflorescence development stages, respectively (Table 4.3). Only one spectral band (band 5) for RE2 was selected amongst the most important variables. In addition, Chlorophyll Index red edge (ChlRed-edge) vegetation indices calculated from both acquisitions were selected amongst the most significant predictor variables too. Besides, the variable importance technique in RF was used to determine the influence of each spectral variable selected on the mapping accuracy. ChlRed-edge vegetation indices from RE1 was the most important variable with a mean decrease accuracy of 0.221% followed by ChlRed-edge vegetation indices from RE2 with a mean decrease accuracy of 0.186%. Band 5 and

Transformed Soil Adjusted Vegetation Index Red Edge (TSAVI) from RE2 were the third and fourth most significant variables, respectively (Table 4.3).

Table 4.3 Spectral variables selected as the most important predictor variables for mapping maize lethal necrosis severity levels using Random forest backward feature elimination procedure.

Acquisition	Spectral variable	Abbreviation	Mean decrease accuracy (%)
RE1	Chlorophyll Index Red edge	ChlRed-edge	0.221
RE2	Chlorophyll Index Red edge	ChlRed-edge	0.186
RE2	RE Band 5 (Red edge)	Band 5	0.091
RE2	Transformed Soil Adjusted Vegetation Index Red Edge	TSAVI	0.091
RE1	Green Normalized Difference Vegetation Index	GDVI	0.085
RE1	Normalized Difference Vegetation Index	NDVI	0.084
RE1	Normalized Difference Red-Edge	NDRE	0.076

4.5.4 Maize lethal necrosis severity classification

Table 4.4 presents the accuracy assessment error matrix for the classification map generated by the RF most important spectral variables to map three MLN severity classes (i.e., mild, moderate and high). The classification achieved an overall accuracy of 73.33%. The producer's accuracy, which indicates the probability of actual areas being correctly classified, was 60.48% for the mild MLN severity, 60.95% for the moderate and 98.57% for the high severity classes. The user's accuracy attained was 66.84% for the mild, 61.24% for the moderate and 89.61% for the high severity classes.

Table 4.4 Random forest classification confusion matrix for three maize lethal necrosis severity classes (mild, moderate and high) using 7 most important spectral variables with 30% test dataset.

Class	Mild	Moderate	High	Total	Producer's accuracy
Mild	127	78	5	210	60.48
Moderate	63	128	19	210	60.95
High	0	3	207	210	98.57
Total	190	209	231	630	
User's Accuracy	66.84	61.24	89.61		

Overall accuracy=73.33%. Kappa=0.60

To improve the overall accuracy, we combined the mild and moderate severity classes which depicted high confusion as a result of similar spectral characteristics for the majority of the sampled farms. This improved the overall accuracy from 73.33 to 90.18% (Table 4.5).

Table 4.5 Error matrix generated for mapping two maize lethal necrosis severity classes (mild and high) using 7 most important spectral variables with 30% test dataset.

Class	Mild	High	Total	Producer's Accuracy
Mild	202	27	210	92.24
High	22	197	210	89.95
Total	215	205	420	
User's Accuracy	90.18	92.06		

Overall accuracy=90.18% Kappa=0.92

The final thematic MLN severity map for the two severity classes (mild and high) produced via the RF algorithm is shown in Figure 4.6. The red color represents maize farms with high severity while the blue color depicts the moderately infected fields. As shown in the zoomed portion of the map some of the maize fields harbored both mildly as well as highly MLN affected plants which agrees with the field observation (Figure 4.6).

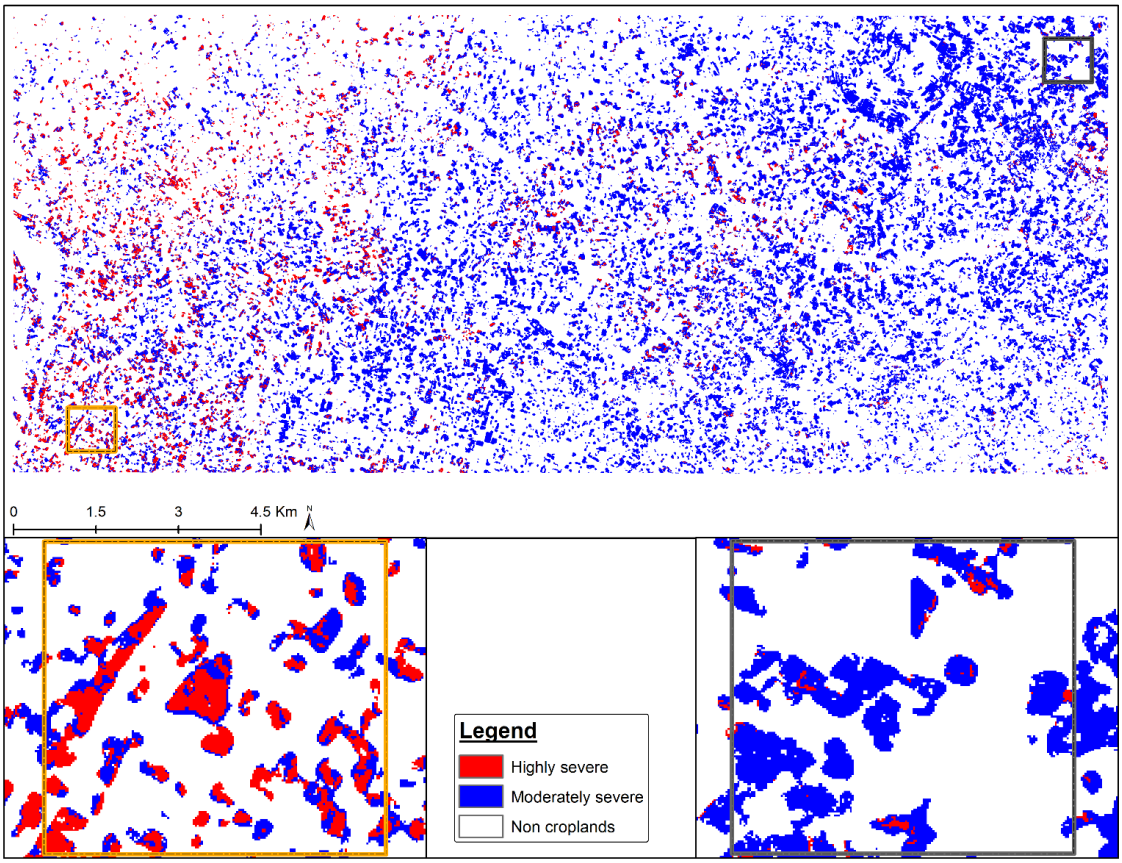

Figure 4.6 Mapping the spatial distribution of maize lethal necrosis severity levels using the seven most important spectral variables selected by random forest.

4.6 Discussion

This study explored the usefulness of bi-temporal RE imagery and a RF classification tool for mapping the MLN severity levels in heterogeneous agro-ecological landscapes in Kenya. A two-step optimized RF classification was used to extract a crop mask from LULC classification and finally generate an MLN severity map for the Bomet County and Southern part of Nyamira County, a major maize growing area in Kenya heavily affected by the disease.

Utilization of the two RE images acquired for the study area during the maize early growing stages did not yield promising results in delineating cropland from other LULC classes. This could be attributed to the late ploughing of some fields as observed during field visits. Thus use of early season RE images alone captured insignificant portions of the phenological development of the maize plants, hence the failure to produce an accurate crop mask (Forkuor 2014). Subsequently, due to the high costs of

RE imagery, the study managed two acquisitions during the maize stem elongation and inflorescence stages, respectively. Therefore, we utilized a cloud free Landsat-8 acquired during maize flowering stage resampled to RE resolution to improve LULC classification. Combining the two acquired RE images with the resampled Landsat-8 image improved the classification accuracy by provided additional information on late cultivated fields which significantly improved the accuracy for extracting crop mask from LULC classification from 80.63% to 91.05%. Similarly, Vladimir et al. (2014) used freely available Landsat-8 data with a single RE image to improve classification of small agricultural fields in northern Serbia.

The high OA of our LULC classification map supports the growing evidence that RF is a reliable classifier for heterogeneous landscapes (Nguyen et al. 2018). For instance, our results revealed a good separability for all the LULC classes apart from the slight confusion between cropland and natural vegetation classes. These results demonstrate the effectiveness of RF classifier to distinguish cropland from other LULC classes in a highly fragmented landscape. The observed overlaps between cropland and natural vegetation classes are well known and can be attributed to the spectral similarity among the vegetation and cropland caused by the presence of small pockets of shrubs within the agricultural land (Forkuor et al. 2015). In addition, the majority of farmers in our study area maintains fruit trees such as mangoes and banana trees within their fields, resulting in heterogeneity and spectral confusion between crops and other vegetation classes (Ayanu et al. 2015).

Essentially, MLN severity levels can accurately be distinguished if there are no other major stressors present that produce similar plant symptoms to those of the disease (Zhang et al. 2012). Field observations confirmed that MLN disease was the dominant stressor and that there was a minimal amount of interference from other biotic and abiotic factors in the sampled maize fields. To minimize such interference, we collected training polygons 5 m away from the farm edges to avoid edge effects and water logging which was observed to affect maize growing at the field edges in some of our sampled maize fields. Nevertheless, care was taken to ensure that infected fields

were correctly identified by visually comparing each classification map with its original NDVI and true color images in the study.

The optimal predictor variables selected using optimized RF backward feature elimination technique were four SVIs (NDVI, GDVI, ChlRed-edge and NDRE) extracted from the RE image acquired during the maize stem elongation stage combined with two SVIs (TSAVI and ChlRed-edge) and one spectral band (band 5) from the RE image acquired during maize inflorescence. These variables proved capable to discriminate two distinguishable MLN severity classes (mild and high) with highest accuracy of 90.18% and a Kappa value of 0.92.

TSAVI was selected amongst the important variables for mapping MLN severity because of its ability to minimize soil brightness influences from spectral vegetation involving red edge and NIR wavelengths (Huete 1988). Besides, TSAVI reduced soil background conditions which imposed extensive influence on partial canopy spectra and calculated SVIs. Similar results were reported by Dhau et al. (2018) who found that soil adjusted vegetation index (SAVI) was amongst the important vegetation indices for detecting and mapping of maize streak virus using RE imagery. In addition, GDVI, NDRE and NDVI were sensitive to MLN severity probably because severely infected maize plants are characterized by low chlorophyll ratio followed by ultimate variations in leaf area. Previous studies showed the importance of NDVI in monitoring of crop stress and disease detection (Eitel et al. 2011). Yet, GNDVI can better predict the leaf area index (LAI) than the conventional NDVI, while NDRE has demonstrated the ability to detect crop stress earlier than NDVI and GNDVI which are traditionally used for plant health monitoring (Wang et al. 2007). Inclusion of these vegetation indices by RF variable selection showed that changes in chlorophyll content is more sensitive to disease severity than changes in water content (Wang et al. 2015). Most notably, the presence of red edge band provided a critical and subtle measurements of vegetation properties such as chlorophyll content necessary for distinguishing between healthy and disease-affected plants (Song et al. 2017). Therefore, our study supports the conclusion that strategically positioned bands such as red edge found in new generation RE

multispectral imagery, contain more spectral information, useful for disease mapping in crop plants (Eitel et al. 2011).

Comparing the classification results generated using three MLN severity classes (mild, moderate and high) with only two severity classes by merging mild and moderate severe classes (mild and high), improved the overall accuracy by 16%. This implied that there was enormous spectral confusion between maize fields that were mildly and moderately affected by MLN. This confusion can bet attributed to the fact that disease estimation is faced with much difficulty at the onset of early symptoms due spectral similarity between slightly infected and non-infected areas (Ashourloo et al. 2014).

Based on the results from this study, a better understanding of the spatio-temporal characteristics of plant diseases is crucial to develop detection tools that are applicable for multi-temporal analyses and the temporal dimension of crop diseases. Therefore, sensor-based identification must be explored further to establish on what resolution and magnitudes disease infestation can be mapped with the specific sensors (Fang and Ramasamy 2015). Considering that the occurrence of plant diseases is dependent of explicit environmental factors and that diseases often exhibit a heterogeneous distribution, optical sensing techniques are useful in identifying primary disease foci and within field disease severity patterns (Melesse et al. 2007).

4.7 Conclusion

Monitoring of MLN severity levels is of immense practical importance, given that the disease tends to develop rapidly and that it is presently very difficult to precisely forecast its development. In this study, a method for mapping MLN severity using multi-temporal RE satellite remote sensing data and optimized machine learning algorithm was developed and tested, ensuring a systematic monitoring of MLN damage levels over a large area. Our results indicate the suitability of remote sensing data as a complementary tool for disease monitoring which could help in the development of effective disease control strategies. Although, low temporal resolution dataset with high spatial resolution is a restrictive factor for practical implementation, the launch of future

observation systems with improved repetition rates such as sentinel-2 can broaden the field of applications. Therefore, explicit geo-spatial and timely synoptic tools are needed for the monitoring of pests and diseases damage levels to facilitate better and more targeted mitigation measures in maize and other important crops. Besides, the effectiveness of remote sensing-based on the spatio-temporal dynamics of MLN must be investigated in future studies to understand linkages between maize pests and diseases hotspots with the underlying ecological factors for better and precise monitoring and management practices.

5 GENERAL CONCLUSION

Against the background of continuing population growth and limited potential to enlarge suitable cropland in sub-Saharan Africa (SSA), the rate of decrease in crop production has caused an imbalance between the food demand of the growing population and the global agricultural output. Enhancing agricultural production efficiency is among the important measures to meet the future food demand. Hence there is an increasing demand for compherensive understanding of the constraints behind low agricultural productivity in SSA. Abiotic and bitoic constraints are among the many factors affecting agricultural productiviy leading to decreased crop outputs, ultimately resulting in food insecurity. Specifically, the status of agricultural productivity in SSA is alarming, and protection against losses majorly caused by crop pests and diseases, can play a critical role in improving the situation in SSA. Yet, under utilization of smart agricultural concepts and a lack of knowledge and skills has resulted to majority of the farmers at the smallholder level being unable to diagnose crop problems sufficiently early or do not possess the technical know-how to manage them effectively. Protection of crops against plant pests and diseases has an obvious role to play in meeting the growing demand for food security. Thus, planning for agricultural adaptation and mitigation must lean on informed decision-making processes, especially on improved pre- and post-harvest pest and disease management practices.

In line with the above, this study investigated the opportunities available and challenges for using both ecological and remotely sensed based variables for better management of crop pests and diseases within the heterogenious landscapes of SSA. For instant, cropping pattern has been proved to be amongst the most significant parameter in crop productivity models by its utilization as a buffer against rapid spread of pests and diseases. Despite its importance, previously, it has not been possible to monitor cropping systems in large scale over time. Thus, results presented in Chapter 2 of this study indicates the possibility of using remote sensing imagery with permisable resolution in extracting cropping system variable in a heterogeneous landscape of SSA. However, selecting an appropriate imagery is far from being trivial. In this regard, use of

multi-temporal data sets using a robust hierarchial classification proved to extract different cropping systems with a very high classification accuracy. The same accuracy was obtained for extracting cropland from other land use landcover classes for the purpose of mapping the disease damage levels in crops (here maize). In this context, it is clear that hierarchical classification is a valuable and effective method that should be preferred in an operational cropping system mapping and disease severity monitoring context, because the findings in this thesis indicate that this concept can be more stable and consistently provides higher classification accuracies over different landscapes for both annual and perrenial crops.

It is definate from the foregoing that the many approaches to controlling crop pests and diseases causing serious losses in SSA or elsewhere in the tropics could be improved through the application of improved techniques such as ecological niche modelling (ENM) and remote sensing (RS) for large scale monitoring as depicted from the results presented in this study. ENM involves a suite of tools that link known occurrences of plant pests and disease or phenomena to geographic information system layers that summarize variations in various environmental dimensions. The results in Chapter 3 proved the importance of combining ecological variables with remotely sensed variables which were deemed appropriate for predicting the distribution and potentially suitable areas for the African citrus triozid (ACT), vector of the devastating citrus greening disease, using ENM approach as compared to the approach of using the existing ecological bioclimatic variables. Remotely-based variables further revealed that vegetation patterns and dynamics at a landscape level play a key role in influencing vector-host-pathogen transmission and distribution. Based on an ENM approach it is possible to understand the ecological niche of insect pests, and thus facilitating more precise management practices especially for scarce crops such as citrus in a highly fragmented agro-ecological landscape where disease monitoring using remote sensing techniques is not possible over a large area. Subsequently, the results in Chapter 4 depict that the biotic stress caused by the pests and disease is assumed to interfere with photosynthesis and the physical structure of the plant at tissue and canopy level, and thus affects the absorption of light energy and alters the reflectance spectrum which

can be monitored using RS imagery with permissible resolution. This implies, that research into vegetation spectral reflectance can help us gain a better understanding of the physical, physiological and chemical processes in plants due to pests and diseases attack and to detect the resulting biotic stress at regional scale using RS imagery captured during a perfect timing of different damage levels. This significanlty assists decision makers in forming strategies for timely management of pests and diseases in hotspots to prevent further spread to non-affected areas.

While the results from this study confirms that application of RS-based tools for monitoring of crop pests and diseases have a positive impact on crop productivity in SSA, there are limitations associated with the methodoly applied in this study. Amongst the main challenges is the often-limited ability to detect early incursions of crop pests and plant diseases as a result of cloud cover which prevents optical sensors from acquiring useful images during critical crop development stages, especially for annual crops. In order to overcome these challenges, satellite information with high spatial and temporal resolution are needed to monitor small farm plots. Therefore, data from Synthetic Aperture Radar (SAR) sensors that are near-independent of weather conditions must be integrated with optical images to improve the temporal coverage of available images in SSA. Another challenge related to full implementation of using ecological and remote sensed variables for enhanced crop productivity is that, in many cases, the details of ecologic parameters associated with occurrences of diseases or of species participating in disease transmission (e.g., vectors, hosts, pathogens) may be unclear because of small sample sizes, biased reporting, or simply lack of detailed geographic or ecologic analysis. This calls for a robust and effective farmer communication tools to facilitate routinely updated information on newly emerging pests and diseases to the respective authorities for mitigation measure. Concerning the operationalization of RS techniques for the monitoring of agricultural pests and diseases in a wider range of agricultural landscapes, it can be concluded from the results of this study, that despite the high performance of existing methods (e.g. ecological modeling and classifier algorithms), one important research issue remains the transferability and stability of such methods. This implies that methods developed and tested in one place

might not necessarily be applicable to another place or over extended time, respectively. Therefore, it seems worthwhile to investigate concepts to use or combine existing techniques that can adapt to specific regional or local contexts.

Even with these caveats, the results from this study have important implications. Primary among these is that the evidence presented in this study supports the belief that precise pests and diseases monitoring approaches can increase crop productivity, and, consequently, reduce food insecurity. Another significant implication is that effective policy measures to promote adoption of smart agriculture should include effective implementation of agricultural systems that are resillient to majority of stressors affecting crop yield in the heterogenious landscapes of SSA. The results from this study have proved the capability of using multi-temporal RS datasets and environmental variables to develop a robust framework for monitoring pests and diseases hotspots at regional scale characterised by the very highly heterogenious landscapes of Africa. However, it is crucial not only to adapt the concept of a better management of agriculture to existing farming practices (e.g. push-pull adoption) but also to engage the government to actively promote technology adoption and ensure that information about improved techniques for crop pests and diseases monitoring and management are effectively disseminated, and the technologies are subsequently adopted. Given the sizeable impacts on potential yield losses, economic surplus and poverty and the fact that for the majority of important pests and diseases no cure or resistant germplasms exist, the incentive for governments of promoting adoption should be considerable. Moreover, a proper monitoring mechanism should be in place to ensure that there is no further spread of pests and diseases to non-affected areas. Therefore, further research should focus on utilization of freely available recently launched RS datasets for better understanding of harmful crop pests and diseases significanlty affecting crop productivity in SSA. In addition, since the concepts and frameworks presented in this thesis are not restricted to optical data, it seems worthwhile to investigate multi-sensory RS data such as radar missions which provide spatially enhanced observations (with swaths suitable at least for local applications) with high revisit frequencies and that are not dependent on the need for cloud free

atmospherically conditions. In conlusion, this study demonstrates the potential of integrating ecological and RS datasets for better understanding of crop pests and diseases in SSA for effecive management. Thus, the results presented here could considerably contribute to the development and implementation of suitable and effective integrated pest management strategies to mitigate further spread of the aforementioned pests and diseases.

6 REFERENCES

Abdel-Rahman E.M, Landmann T, Kyalo R et al (2017) Predicting stem borer density in maize using RapidEye data and generalized linear models. Int J Appl Earth Obs Geoinf 57: 61-74.

Abdel-Rahman EM, Ahmed FB (2008) The application of remote sensing techniques to sugarcane (saccharum spp Hybrid) production: A review of the literature. Int J Remote Sens 29: 3753-3767.

Abdel-Rahman EM, Mutanga O, Adam E et al (2014) Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS J Photogramm Remote Sens 88: 48-59.

Adams IP, Miano DW, Kinyua ZM et al (2013) Use of next-generation sequencing for the identification and characterization of Maize chlorotic mottle virus and Sugarcane mosaic virus causing maize lethal necrosis in Kenya.Plant Pathology 62: 741-749.

Adams IP, Harju VA, Hodges T et al (2014) First report of Maize Lethal Necrosis Disease in Rwanda. New Dis Rep 29: 22.

Adhikari U, Nejadhashemi AP, Woznicki SA (2015) Climate change and eastern Africa: A review of impact on major crops. Food Energy Secur 4: 110–132.

Ahamed T, Tian L, Zhang Y, Ting KC (2011) A review of remote sensing methods for biomass feedstock production. Biomass Bioenergy 35: 2455-2469.

Albayrak S (2008) Use of Reflectance Measurements for the Detection of N, P, K, ADF and NDF Contents in Sainfoin Pasture. Sensors 8: 7275–7286.

Alkishe AA, Peterson AT, Samy AM (2017) Climate change influences on the potential geographic distribution of the disease vector tick Ixodes Ricinus. PLoS ONE 12: e0189092.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence kappa and the true skill statistic (TSS). J Appl Ecol 43: 1223–1232.

Alvarez S, Rohrig E, Solís D et al (2016) Citrus Greening Disease (Huanglongbing) in Florida: Economic Impact Management and the Potential for Biological Control. Agric Res 5: 109–118.

Amirpour HS, Barrios M, Farifteh et al (2013) Ecological niche modelling of bank voles in Western Europe. Int J Environ Res Public Health 10: 499–514.

Amoros Lopez J (2011) Land cover classification of VHR airborne images for citrus grove identification. ISPRS J Photogramm Remote Sens 66: 115–123.

Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: Criteria for selecting optimal models. Ecol Model 162: 211–232.

Anon (1984) Horticulture crops protection handbook Ministry of Agriculture. Nairobi, Kenya.

Antle JM, Basso B, Conant RT et al (2016) Towards a new generation of agricultural system data models and knowledge products: Design and improvement. Agric Syst 155: 255-268.

Apan A, Held A, Phinn S, Markley J (2003) Formulation and assessment of narrow-band vegetation indices from EO-1 hyperion imagery for discriminating sugarcane disease In Proceedings of Spatial Sciences Institute Conference: spatial knowledge without boundaries. Canberra, Australia.

Arvor D, Jonathan M, Meirelles MSeP et al (2011) Classification of modis evi time series for crop mapping in the state of Mato Grosso Brazil. Int J Remote Sens 32: 7847-7871.

Asharaf S, Khan AG, Ali S et al (2002) An Assessment of the Socio-Economic Factors Affecting the Adoption of Citrus Tissue Culture Technology in Kenya; Ciencia Rural: Santa Maria, Brazil.

Ashourloo D, Mobasheri MR, Huete A (2014) Developing Two Spectral Disease Indices for Detection of Wheat Leaf Rust (Pucciniatriticina). Remote Sens Ecol Conserv 6: 4723-4740.

Aubert B (1987) Trioza erytreae Del Guercio and Diaphorina citri Kuwayama (Homoptera: Psyllidae) the two vectors of citrus greening disease: Biological aspects and possible control strategies. Fruits 42: 149–162.

Avelino J, Romero-Gurdian A, Cruz-Cuellar HF (2012) Declerck FA Landscape context and scale differentially impact coffee leaf rust coffee berry borer and coffee root-knot nematodes. Ecol Appl 22: 584–96.

Ayanu Y, Conrad C, Jentsch A et al (2015) Unveiling Undercover Cropland Inside Forests Using Landscape Variables: A Supplement to Remote Sensing Image Classification. PLoS ONE 10: e0130079.

Barredo JI, Strona G, Rigo D et al (2015) Assessing the potential distribution of insect pests: Case studies on large pine weevil (Hylobius abietis L) and horse-chestnut leaf miner (Cameraria ohridella) under present and future climate conditions in European forests. EPPO Bull 45: 273–281.

Barrett B, Nitze I, Green S et al (2014) Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning Approaches. Remote Sens Environ 152: 109-124.

Barrett B, Raab C, Cawkwell F et al (2016) Upland vegetation mapping using Random Forests with optical and radar satellite data. Remote Sens Ecol Conserv 2: 212-231.

Bassa Z (2012) An assessment of land cover change patterns using remote sensing: A case study of Dube and Esikhawini KkwaZulu-Natal South Africa University of KwaZulu-Natal, Durban, 2012.

Basso B, Cammarano D, Cafiero G et al (2011) Cultivar discrimination at different site elevations with remotely sensed vegetation indices. Ital J Agron 6: 1-5.

Benin S, Wood S Nin-Pratt A (2016) Introduction in agricultural productivity in Africa: Trends, patterns, and determinants. Benin, Samuel (Ed.). IFPRI, Washington D.C.

Benson JM, Poland JA, Benson BM et al (2015) Resistance to Gray Leaf Spot of Maize: Genetic Architecture and Mechanisms Elucidated through Nested Association Mapping and Near-Isogenic Line Analysis. PLoS Genet 11(3): e1005045.

Berk A, Anderson GP, Acharya PK et al (2008) MODTRAN 5.2.0.0 User's Manual. Burlington, MA: Spectral Sciencess inc.

Birth G, McVey G (1968) Measuring the color of growing turf with a reflectance spectrophotometer. Agron J 60: 640-643.

Blackburn GA (1998) Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. J Remote Sens 19: 657-675.

Blum M, Lensky IM, Rempoulakis P et al (2015) Modeling insect population fluctuations with satellite land surface temperature. Ecol Model 311: 39–47.

Bock CH, Poole GH, Parker PE et al (2010) Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. Crit Rev Plant Sci 29(2): 59-107.

Bové JM (2014) Huanglongbing or yellow shoot a disease of Gondwanan origin: Will it destroy citrus worldwide? Phytoparasitica 42: 579–583.

Boykin LM, De Barro P, Hall DG et al (2012) Overview of worldwide diversity of Diaphorina citri Kuwayama mitochondrial cytochrome oxidase 1 haplotypes: Two Old World lineages and a New World invasion Bull. Entomol Res 102: 573–582.

Breiman L (2001) Random forests. Mach Learn 45: 5-32.

Brownstein JS, Holford TR, Fish D (2003) A climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. Environ Health Perspect 111: 1152–1157.

Bryan E, Ringler C, Okoba B et al (2013) Roncoli C Silvestri S Herrero M Adapting agriculture to climate change in Kenya: Household strategies and determinants. J Environ Manage 114: 26-35.

Buermann W, Saatchi S, Smith TB et al (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. J Biogeogr 35: 1160–1176.

Cai Z, Jönsson P, Jin H et al (2017) Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. Remote Sens 9: 1271.

Callaghana JRO, Maende C, Wyseure G (1994) Modelling the intercropping of maize and beans in Kenya. Comput Electron Agri 11: 351-365.

Campbell CL, Neher DA (1994) Estimating Disease Severity and Incidence. In: Epidemiology and Management of Root Diseases. Springer: Berlin, Heidelberg.

Castellanos A, Tittonell P, Rufino MC et al (2014) Feeding crop residue and manure management for integrated soil fertility management - a case study from Kenya. Agric Syst 134: 24-35.

Castillo J, Hebert TT (1974) New virus disease affecting maize in Peru. (Nueva enfermedad virosa afectando almaiz en el Peru.). Fitopatologia 9: 79-84.

Catling HD (1973) Notes on the biology of the South African citrus psylla Trioza erytreae (Del Guercio) (Homoptera: Psyllidae). J Entomol Soc S Afr 36: 299–306.

Chabot-Couture G, Nigmatulina K, Eckhoff P (2014) An Environmental Data Set for Vector-Borne Disease Modeling and Epidemiology. PLoS ONE 9: e94741.

Chakraborty S, Newton A C (2011), Climate change, plant diseases and food security: an overview. Plant Pathol 60: 2-14.

Chen J (2004) A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens Environ 91: 332–344.

Chiyaka C, Singer BH, Halbert SE et al (2012) Modeling huanglongbing transmission within a citrus tree. Proc Natl Acad Sci USA 109: 12213–12218.

Congalton R (2001) Accuracy assessment and validation of remotely sensed and other spatial information. Int J wildland fire 10: 321-328.

Conrad C, Colditz R, Dech S et al (2011) PLG Temporal segmentation of MODIS time series for improving crop classification in Central Asian irrigation systems. Int J Remote Sens 32: 8763–8778.

Conrad C, Machwitz M, Schorcht G et al (2011) Potentials of rapideye time series for improved classification of crop rotations in heterogeneous agricultural landscapes: Experiences from irrigation systems in central Asia In Proceedings of SPIE Remote Sensing. Prague, Czech Republic.

Cook G, Maqutu VZ, Vuuren SPV (2013) Population Dynamics and Seasonal Fluctuation in the Percentage infection of Trioza erytreae with 'Candidatus' Liberibacter Africanus the African Citrus Greening Pathogen in an Orchard Severely Infected with African Greening and Transmission by Field-Collected Trioza erytreae. Afr Entomol 22: 127–135.

Cooley T, Anderson GP, Felde GW et al (2002) FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. Proceedings of the 2002 IEEE International Geoscience and Remote Sensing Symposium, IGARSS, Toronto, 2002.

Cord A, Conrad C, Schmidt M et al (2010) Standardized FAO-LCCS land cover mapping in heterogeneous tree savannas of West Africa. J Arid Environ 74: 1083–1091.

Cord AF, Klein D, Gernandt et al (2014) Remote sensing data can improve predictions of species richness by stacked species distribution models: A case study for Mexican pines. J Biogeogr 41: 736–748.

Curran PJ, Hay AM (1996) The importance of measurement error for certain procedures in remote sensing at optical wavelengths. Photogramm Eng Remote Sens 52: 229-241.

Datt B (1998) Remote sensing of chlorophyll a chlorophyll b chlorophyll a+b and total carotenoid content in eucalyptus leaves. Remote Sens Environ 66: 111-121.

De Meyer M, Robertson MP, Mansell MW et al (2010) Ecological niche and potential geographic distribution of the invasive fruit fly Bactrocera invadens (Diptera Tephritidae). Bull Entomol Res 100: 35–48.

Deressa T (2017) Maize Lethal Necrosis Disease (MLND)-A Review. J Nat Sci Res 7: 38-43.

Dhau I, Adam E, Ayis KK et al (2018) Detection and mapping of maize streak virus using RapidEye satellite imagery. Geocarto Int 1-11.

Diaz-Uriarte (2017) R Package "varselrf". http://ligartorg/rdaz/Software/Softwarehtml. Cited 20 April 2017.

Dormann CF, Elith J, Bacher S et al (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27–46.

Edgerton MD (2009) Increasing crop productivity to meet global needs for feed food and fuel. Plant Physiology 149: 7-13.

Efron B, Tibshirani R (1997) Improvements on cross-validation: The 632+ bootstrap method. J Am Stat Assoc 92: 548-560.

Eitel JUH, Vierling LA, Litvak ME et al (2011) Broadband red-edge information from satellites improves early stress detection in a New Mexico conifer woodland Remote Sens Environ 115: 3640-3646.

El-Shikha DM, Barnes EM, Clarke TR et al (2008) Remote sensing of cotton nitrogen status using the canopy chlorophyll content index (ccci). T ASABE 51 :73-82.

ESA (2017) CCI Land Cover-S2 Prototype Land Cover Map of Africa. http://www2016africalandcover20mesrinesaint/. Cited 23 May 2018.

Fang Y, Ramasamy RP (2015) Current and Prospective Methods for Plant Disease Detection. Biosensors 5: 537-561.

FAO (2016) FAOSTAT Statistics Database; FAO: Rome, Italy.

FAOSTAT (2010) Statistical database and data-sets of the Food and Agriculture Organization of the United Nations. http://www.fao.org/docrep/015/am081m/am081m00.htm. Cited 3 September 2017.

FAOSTAT (2013) Food and Agricultural Commodities Production. Food and Agriculture Organization of the United Nations. Rome, Italy.

Flett BC, Bensch MJ, Smit E (2002) In Field guide for identification of maize pests in South Africa Potchefsroom.

Forkuor G (2014) Agricultural land use mapping in West Africa using multi-sensor satellite imagery University of Würzburg, Germany.

Forkuor G (2014) Agricultural Land Use Mapping in West Africa Using Multi-Sensor Satellite Imagery. Ph.D. Thesis, University of Würzburg, Würzburg, Germany.

Forkuor G, Conrad C, Thiel M et al (2014) Integration of optical and synthetic aperture radar imagery for improving crop mapping in Northwestern Benin West Africa. Remote Sens 6: 6472-6499.

Forkuor G, Conrad C, Thiel M et al (2015) Evaluating the sequential masking classification approach for improving crop discrimination in the Sudanian Savanna of West Africa. Comput Electron Agric: 380-389.

Forkuor G, Hounkpatin OKL, Welp G et al (2017) High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. PLoS ONE 12: e0170478.

Franco-Vega A, Reyes-Jurado F, Cardoso-Ugarte GA, et al (2016) Chapter 89—Sweet Orange (Citrus sinensis) Oils A2 In Essential Oils in Food Preservation Flavor and Safety; Preedy VR Ed Academic Press: San Diego CA USA 783–790.

Franke J, Menz G (2007) Multi-temporal wheat disease detection by multi-spectral remote sensing. Precis Agric 8(2): 161-172.

Galiano VFR, Olmo MC, Hernandez FA et al (2012) Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens Environ 121: 93-107.

Geerts S, Raes D, Garcia M et al (2006) Agro-climatic suitability mapping for crop production in the Bolivian Altiplano: A case study for quinoa. Agric For Meteo 139: 399-412.

Georganos S, Grippa T, Vanhuysse S et al (2018) Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GISc Remote Sens 55(2): 221-242.

Gianoli E, Ramos I, Alfaro-Tapia A et al (2006) Benefits of a maize-bean-weeds mixed cropping system in Urubamba Valley Peruvian Andes. Int J Pest Manage 52: 283-289.

Gitelson AA, Keydan GP, Merzlyak MN (2006) Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophys Res Lett 33(11): L1402.

Gitelson AA, Merzlyak MN, Zur Y et al (2001) Non-destructive and remote sensing techniques for estimation of vegetation status In Proceedings of Third European Conference on Precision Agriculture. Montipellier, France 301-306.

Grafton-Cardwell EE, Stelinski LL, Stansly PA (2013) Biology and Management of Asian Citrus Psyllid Vector of the Huanglongbing. Pathogens Annu Rev Entomol 58: 413–432.

Green GC, Catling HD (1971) Weather induced mortality of the citrus psylla trioza erytreae (del guercio) (homoptera: Psyllidae) a vector of greening virus in some citrus producing areas of Southern Africa. Agric Meteorol 8: 305–317.

Guanter L, Richter R, Kaufmann H (2009) On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing. Int J Remote Sens 30: 1407-1427.

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3: 1157-1182.

Henrik S and Peeter P (1997) Reduction of stemborer damage by intercropping maize with cowpea. Agric Ecosyst Environ 62: 13-19.

Hijmans RJ, Cameron SE, Parra JL et al (2005) A Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25: 1965–1978.

Hilker FM, Allen LJS, Bokil VA et al (2007) Modeling Virus Coinfection to Inform Management of Maize Lethal Necrosis in Kenya. Phytopathology 107(10): 1095-1108.

Hof AR, Svahlin A (2016) The potential effect of climate change on the geographical distribution of insect pest species in the Swedish boreal forest Scand. J Res 31: 29–39.

Hol WH, Bezemer TM, Biere A (2013) Getting the ecology into interactions between plants and the plant growth-promoting bacterium Pseudomonas fluorescens Front. Plant Sci 4: 81.

Huang BFF, Boutros PC (2016) The parameter sensitivity of random forests. BMC Bioinformatics 17(1): 331.

Huete A, Didan K, Miura T et al (2002) Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. Remote Sens Environ 83: 195–213.

Huete AR (1998) A soil-adjusted vegetation index (SAVI). Remote Sens Environ 25: 295-309.

Hunt Jr ER, Raymond E, Daughtry CST et al (2011) Long DS Remote sensing leaf chlorophyll content using a visible band index. Agron J 103: 1090-1099.

Husak GJ, Marshall MT, Michaelsen J et al (2008) Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. J Geophys Res Atmos 113: 1-8.

ICIPE (2018) SCIPM:Project by ICIPE and Partners to Improve Citrus Farming 2015.http://wwwicipeorg/news/scipm-project-icipe-and-partners-improve-citrus-farming. Cited 17 April 2018.

Iqbal MS, Jabbar B, Sharif NM et al (2017) In silico MCMV Silencing Concludes Potential Host-Derived miRNAs in Maize. Front Plant Sci 8: 372.

Jaetzold R, Schmidt H (1982) Farm management handbook of Kenya. Ministry of Agriculture. Nairobi, Kenya. 2: 397-400.

Jagoueix S, Bove MJ, Garnier M (1994) The phloem-limited bacterium of greening disease of citrus is a member of the α subdivision of the proteobacteria. Int J Syst Bacteriol 44: 379–386.

Jarvis A, Reuter HI, Nelson A et al (2008) Hole-Filled SRTM for the Globe Version 4 The CGIAR Consortium for Spatial Information (CGIAR-CSI) 2008. http://srtmcsicgiarorg/. Cited 16 May 2018.

Jensen SG, David S, Wysong EM et al (1991) Seed transmission of maize chlorotic mottle virus. Plant Dis 75(5): 497-498.

Jiang XQ, Wilkinson DR, Berry JA (1990) An outbreak of maize chlorotic mottle virus in Hawaii and possible association with thrips. Phytopathology 80: 1060.

Jönsson P, Eklundh L (2004) TIMESAT—A program for analyzing time-series of satellite sensor data. Comput Geosci 30: 833–845.

Jorge SB (2011) Mapping Species Distributions: Spatial Inference and Prediction. Q Rev Biol 86: 219–220.

Khamis FM, Rwomushana I, Ombura LO et al (2017) DNA Barcode Reference Library for the African Citrus Triozid Trioza erytreae (Hemiptera: Triozidae): Vector of African Citrus Greening. J Econ Entomol 110: 2637–2646.

Khan ZR, Midega CAO, Wanyama JM et al (2009) Integration of edible beans (Phaseolus vulgaris L.) into the push-pull technology developed for stemborer and Striga control in maize-based cropping systems. Crop Prot 28(11): 997-1006.

Kim HO, Yeom JM (2015) Sensitivity of vegetation indices to spatial degradation of RapidEye imagery for paddy rice detect: A case study of South Korea. GIsci Remote Sens 52: 1-17.

Lancashire PD, Bleiholder H, Boom TVD et al (1991) A uniform decimal code for growth stages of crops and weeds. Ann App Bio 119: 561-601.

le Maire G, Francois CD (2204) Towards universal broad leaf chlorophyll indices using prospect simulated database and hyperspectral reflectance measurements Remote Sens Environ 89: 1-28.

Li Z, Li X, Wei D et al (2010) An assessment of correlation on MODIS-NDVI and EVI with natural vegetation coverage in Northern Hebei Province China. Procedia Environ Sci 2: 964–969.

Liaw A, Weiner M (2002) Classification and regression by random Forest R News 2: 18–22.

Liu S, Liu X, Liu M et al (2017) Extraction of Rice Phenological Differences under Heavy Metal Stress Using EVI Time-Series from HJ-1A/B Data. Sensors 17: 1243.

Liu Y, Heying E, Tanumihardjo SA (2012) History Global Distribution and Nutritional Importance of Citrus Fruits. Compr Rev Food Sci Food Saf 11: 530–545.

Lord CC (2007) Modeling and biological control of mosquitoes. J Am Mosq Control Assoc 23: 252–264.

Ma H, Jing Y, Huang W et al (2018) Integrating Early Growth Information to Monitor Winter Wheat Powdery Mildew Using Multi-Temporal Landsat-8 Imagery. Sensors 18(10):3290.

Macharia P (2016) Gateway to land and water information: Kenya national report Fao (food and agriculture organization). http://wwwfaoorg/ag/agL/ swlwpnr/reports/y_sf/z_ke/kehtm. Cited 18 November 2016.

Mahlein AK (2015) Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. Plant Dis 100(2): 241-251.

Mahlein AK, Oerke E-C, Steiner U et al (2012) Recent advances in sensing plant diseases for precision crop protection. Eur J Plant Pathol 133(1):197-209.

Mahuku G, Lockhart BE, Wanjala B et al (2015) Maize lethal necrosis (MLN) an emerging threat to maize-based food security in Sub-Saharan Africa. Phytopathology 105: 956-965.

Mahuku G, Wangai A, Sadessa K et al (2015) First report of Maize chlorotic mottle virus and Maize lethal necrosis on maize in Ethiopia. Plant Dis 99(12): 1870.

Makori D, Fombong A, Abdel-Rahman et al (2017) Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models. ISPRS Int J Geo-Inf 6: 66.

Marchioro CA (2016) Global Potential Distribution of Bactrocera carambolae and the Risks for Fruit Production in Brazil. PLoS ONE 11: e0166142.

Margosian ML, Garrett KA, Hutchinson JMS et al (2009) Connectivity of the American Agricultural Landscape: Assessing the National Risk of Crop Pest and Disease Spread. BioSci 59: 141–151.

Marinho RA, Beserra EB, Bezerra-Gusmão MA et al (2016) Effects of temperature on the life cycle, expansion, and dispersion of Aedes aegypti (Diptera: Culicidae) in three cities in Paraiba, Brazil. J Vector Ecol 41: 1-10.

Matsushita B, Yang W, Chen J et al (2007) Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-Density Cypress Forest. Sensors 7: 2636–2651.

Matyukhina DS, Miquelle DG, Murzin et al (2014) Assessing the Influence of Environmental Parameters on Amur Tiger Distribution in the Russian Far East Using a MaxEnt Modeling Approach. Achiev Life Sci 8: 95–100.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochemia Medica 22: 276–282.

Melesse AM, Weng Q, Thenkabail PS et al (2007) Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. Sensors 7(12): 3209-3241.

Mersha AA, Laerhoven FV (2016) A gender approach to understanding the differentiated impact of barriers to adaptation: Responses to climate change in rural Ethiopia. Reg Environ Change 16: 1701-1713.

Michaud JP (2004) Natural mortality of Asian citrus psyllid (Homoptera: Psyllidae) in Central Florida. Biol Control 29: 260–269.

Mohajan HK (2014) Food and Nutrition Scenario of Kenya. Am J Food Nutr 2(2): 28-38.

Moore SM, Borer ET, Hosseini PR (2010) Predators indirectly control vector-borne disease: Linking predator–prey and host–pathogen models. J R Soc Interface 7: 161–176.

Mulianga B, Begue A, Clouvel P et al (2015) Mapping cropping practices of a sugarcane-based cropping system in Kenya using remote sensing Remote Sens 7: 14428-14444.

Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. Int J Appl Earth Obs Geoinf 18:399-406.

Muthoni J, Nyamongo D.O (2010) Traditional Food Crops and Their Role in Food and Nutritional Security in Kenya. J Agric Food Info 11(1): 36-50.

Mwangi EW, Mundia CN (2014) Assessing and monitoring agriculture crop production for improved food security in Machakos County. Int J Sci Res 3: 555-559.

Myers SS, Smith MR, Guth S, et al (2017) Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition. Annu Rev Public Health 38:259-277.

Narouei-Khandan HA, Halbert SE, Worner SP et al (2016) Global climate suitability of citrus huanglongbing and its vector the Asian citrus psyllid using two correlative species distribution modeling approaches with emphasis on the USA. Eur J Plant Pathol 144: 655–670.

Nault LR, Styer WE, Coffey ME et al (1978) Transmission of maize chlorotic mottle virus by chrysomelid beetles. Phytopathology 68(7): 1071-1074.

Newton AC, Hackett CA (1994) Subjective components of mildew assessment on spring barley. European Journal of Plant Pathol 100(6): 395-412.

Nguyen HTT, Doan TM, Radeloff V (2018) Applying random forest classification to map land use/land cover using landsat 8 OLI. Int. Arch. Photogramm. Remote Sens XLI11-3/W4: 363-367.

Nicholas ID (1988) 26—Plantings in Tropical and Subtropical Areas A2 In Windbreak Technology; Brandle JR Hintz DL Sturrock JW Eds Elsevier: Amsterdam, Netherlands 465–482.

Nutter Jr FW, Schultz PM (1995) Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. Can J Plant Pathol 17: 174-184.

Nyasani JO, Meyhöfer R, Subramanian S et al (2012) Effect of intercrops on thrips species composition and population abundance on French beans in Kenya. Entomolog Exp Appl 142: 236–246.

Ochieng J, Kirimi L, Mathenge M (2016) Effects of climate variability and change on agricultural production: The case of small scale farmers in Kenya. NJAS 77: 71-78.

Omiti J, Otieno DJ, Nyanamba TO et al (2009) Factors influencing the intensity of market participation by smallholder farmers: a case study of rural and peri-urban areas of Kenya. Afr J Agric Res 3: 57-82.

Oosten CV (1989) Farming Systems and Food Security in Kwale District Kenya; MOPAN Development Ed Africa Studies Centre. Leiden, Netherlands.

Ouma G (2008) Challenges and approaches to sustainable citrus production in Kenya. Afr J Plant Sci Biotechnol 2: 49–51.

Ozdemir L (2007) Separation of Citrus Plantations from forest cover using Landsat Imagery. Allg For Jagdztg 178: 208–212.

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

Palaniswami C, Viswanathan R, Bhaskaran A et al (2014) Mapping sugarcane yellow leaf disease affected area using remote sensing technique. J Sugarcane Res 4: 55-61.

Panigrahy S, Manjunath KR, Ray SS (2005) Deriving cropping system performance indices using remote sensing data and Gis. Int J Remote Sens 26: 2595-2606.

Paul J, Hatfield J, Schepers JS et al (2003) Remote Sensing for Crop Management.

Paul PA, Munkvold G A (2004) model-based approach to preplanting risk assessment for gray leaf spot of maize. Phytopathology 94: 1350-1357.

Paull SH, Song S, McClure KM et al (2012) From superspreaders to disease hotspots: Linking transmission across hosts and space. Front Ecol Environ 10: 75–82.

Pearsall IA, Myers JH (2001) Spatial and temporal pattern of dispersal of western flower thrips (Thysanoptera: Thripidae) in nectarine orchards in British Columbia. J Econ Entomol 94: 831-843.

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob Ecol Biogeogr 12: 361–71.

Penatti N, Isnard T (2012) Subdivision of pantanal quaternary wetlands: Modis NDVI timeseries in the indirect detection of sediments granulometry. Int Arch Photogramm Remote Sens Spat Inf Sci XXXIX-B8: 311–316.

Peterson AT (2006) Ecologic Niche Modeling and Spatial Patterns of Disease Transmission. Emerg Infect Dis 12: 1822–1826.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231–259.

Pierce KB, Lookingbill T, Urban D et al (2005) Simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis Landscape Ecology. Landsc Ecol 20: 137–147.

Plantegenest M, Le May C, Fabre FDR (2007) Landscape epidemiology of plant diseases. J R Soc Interface 4: 963–972.

Pole FN, Ndung'u JM, Kimani JM (2010) Citrus farming in Kwale district: A case study of Lukore location In Proceedings of the 12th KARI Biennial Conference: Transforming Agriculture for Improved Livelihoods through Agricultural Product Value Chains. Nairobi, Kenya.

Pontius RG, Millones M (2011) Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32: 4407-4429.

Pu R, Gong P, Yu Q (2008) Comparative analysis of EO-1 ALI and hyperion and Landsat ETM+ Data for mapping forest crown closure and leaf area index. Sensors 8: 3744-3766.

Qin A, Liu B, Guo Q et al (2017) Maxent modeling for predicting impacts of climate change on the potential distribution of Thuja sutchuenensis Franch an extremely endangered conifer from southwestern China. Glob Ecol Conserv 10: 139–146.

R Core Team (2013) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria.

Ramert B, Lennartsson M, Davies G (2002) The use of mixed species cropping to manage pests and diseases - theory and practice In Proceedings of the UK Organic Research Conference Organic Centre Wales Institute of Rural Studies University of Wales, Aberystwyth, 207-210.

Ramirez GRG, Medina HCPB, Trujillo TJR (2016) Agroclimatic risk of development of Diaphorina citri in the citrus region of Nuevo Leon Mexico. Afr J Agric Res 11: 3254–3260.

RapidEye (2013) RapidEye Mosaic TM product specifications. http:// blackbridge.com/rapideye/upload/RapidEye. Cited 18 May 2018

Ratnadass A, Fernandes P, Avelino J et al (2012) Plant species diversity for sustainable management of crop pests and diseases in agroecosystems: Review Agron Sustain Dev 32: 273–303.

Reynolds TW, Waddington SR, Anderson CL et al (2015) Environmental impacts and constraints associated with the production of major food crops in Sub-Saharan Africa and South Asia. Food Sec 7:795-822.

Richard K, Abdel-Rahman EM, Subramanian S et al (2017) Maize Cropping Systems Mapping Using RapidEye Observations in Agro-Ecological Landscapes in Kenya. Sensors 3(17): E2537.

Richter R (1997) Correction of atmospheric and topographic effects for high spatial resolution satellite imagery Int J Remote Sens 18: 1099-1111

Rizzo DM, Garbelotto M (2003) Sudden oak death: Endangering California and Oregon forest ecosystems. Front Ecol Environ 1: 197–204.

Robin G, Jean-Michel P, Christine TM (2010) Variable selection using random forests. Pattern Recogn Lett 31: 2225-2236

Rockstrom J (2003) Water for food and nature in drought prone tropics: Vapour shift in rainfed agriculture. Roy Soc 1997-2009.

Romeo J, Pajares G, Montalvo M et al (2012) A Crop row detection in maize fields inspired on the human visual perception Scientific World J.

Roujean J, Breon F (1995) Estimating par absorbed by vegetation from bidirectional reflectance measurements. Remote Sens Environ 51: 375-384.

Roy DP, Ju J, Mbow C et al (2010) Accessing free Landsat data via the internet: Africa's challenge. Remote Sens Lett 1: 111-117.

Royle JA, Chandler RB, Yackulic C et al (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions Methods. Ecol Evol 3: 545–554.

Sahlean TC, Gherghel I, Papeş M et al (2014) Refining Climate Change Projections for Organisms with Low Dispersal Abilities: A Case Study of the Caspian Whip Snake. PLoS ONE 9: e91994.

Schuster C, Forester M, Kleinschmit B (2012) Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. Int J Remote Sens 33: 5583-5599.

Seck PA, Diagne A, Mohanty S et al (2012) Crops that feed the world 7: Rice Food Security 4: 7-24.

Seran TH, Brintha I (2010) Review on Maize Based Intercropping Agron J 9: 135-145.

Shabani F, Kumar L, Ahmadi MA (2016) Comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. Ecol Evol 6: 5973–5986.

Shimwela MM, Narouei-Khandan HA, Halbert SE et al (2016) First occurrence of Diaphorina citri in East Africa characterization of the Ca Liberibacter species causing huanglongbing (HLB) in Tanzania and potential further spread of D citri and HLB in Africa and Europe. Eur J Plant Pathol 146: 349–368.

Shrivastava RJ, Gebelein JL Landcover classification and economic assessment of citrus groves using remote sensing. ISPRS J Photogramm Remote Sens 61: 341–353.

Sibanda M, Murwira A (2011) The use of multi-temporal MODIS images with ground data to distinguish cotton from maize and sorghum fields in smallholder agricultural landscapes of Southern Africa. Int J Remote Sens 33: 4841-4855.

Smith JH, Stehman SV, Wickham JD et al (2003) Effects of landscape characteristics on land-cover class accuracy. Remote Sens Environ 84: 342–349.

Song X, Yang C, Wu M et al (2017) Evaluation of Sentinel-2A Satellite Imagery for Mapping Cotton Root Rot. Remote Sens 9: 906.

Songa JM, Jiang N, Schulthess F et al (2007) The role of intercropping different cereal species in controlling lepidopteran stemborers on maize in Kenya. J Appl Entomol 131: 40–49.

Strachan IB, Pattey E, Boisvert JB (2002) Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. Remote Sens Environ 80: 213-224.

Strobl C, Boulesteix AL, Kneib T et al (2008) Conditional variable importance for random forests. BMC Bioinformatics 9: 307-307

Sun B, Peng Y, Yang H et al (2014) Alfalfa (medicago sativa l)/maize (zea mays l) intercropping provides a feasible way to improve yield and economic incomes in farming and pastoral areas of Northeast China. PLoS ONE 9: e110556.

Thies C, Steffan-Dewenter I, Tscharntke T (2003) Effects of landscape context on herbivory and parasitism at different spatial scales. Oikos 101: 18–25.

Tigges J, Lakes T, Hostert P (2013) Urban vegetation classification: Benefits of multi-temporal RapidEye satellite data. Remote Sens Environ 136: 66-75.

Tonnang HEZ, Hervé BDB, Biber-Freudenberger L et al (2017) Advances in crop insect modelling methods—Towards a whole system approach. Ecol Model 354: 88–103.

Tuck SL, Phillips HRP, Hintzen RE et al (2014) MODISTools—Downloading and processing MODIS remotely sensed data in R. Ecol Evol 4: 4658–4668.

Tucker C (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ 8: 127-150

Usery EL, Finn MP, Scheidt DJ et al (2004) Geospatial data resampling and resolution effects on watershed modeling: A case study using the agricultural non-point source pollution model. J Geogr Syst 6: 289–306.

Uyemoto JK (1983) Biology and control of maize chlorotic mottle virus. Plant Dis 67: 7-10.

Van Den Berg MA, Anderson SH, Deacon VE (1991) Population studies of the citrus psylla trioza erytreae: Factors influencing dispersal. Phytoparasitica 19: 283.

Vilamiu RGdA, Ternes S, Braga GA et al (2012) A model for Huanglongbing spread between citrus plants including delay times and human intervention. Proc Natl Acad Sci USA 1479: 2315–2319.

Vintrou E, Houles M, Seen DL et al (2009) A Mapping cultivated area in West Africa using modis imagery and agro-ecological stratification. In proceedings of IEEE International Geoscience and Remote Sensing Symposium. Cape Town, South Africa.

Vladimir C, Lugonja P, Brkljac BN et al (2014) Classification of small agricultural fields using combined Landsat-8 and RapidEye imagery: case study of northern Serbia. J Appl Remote Sens 9: 083512

Waithaka K (1991) Consultant's Report on Tropical Fruit Production in East and Southern Africa. FAO, Rome, Italy.

Wang F, Huang J, Tang Y et al (2007) New vegetation index and its application in estimating leaf area index of rice. Rice Sci 14: 195-203.

Wang F-m, Huang J-f, Tang, Y-l et al (2007) New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. Rice Sci 14(3): 195-203

Wang L, Zhou X, Zhu X et al (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J 4: 212–219.

Wang Z, Liu C, Chen W et al (2006) Preliminary Comparison of MODIS-NDVI and MODIS-EVI in Eastern Asia; Geomatics and Information Science of Wuhan University: Wuhan, China 31: 407–410.

Wang Z-G, Jin X, Bao X-G et al (2014) Intercropping enhances productivity and maintains the most soil fertility properties relative to sole cropping. PLoS ONE 9: e113984.

Wangai AW, Redinbaugh MG, Miano DW et al (2018) First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. Plant Dis 96(10): 1582.

Waske B, Benediktsson JA, Ãrnason K et al (2009) Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Can J Remote Sens 35: 106-116.

Wei H, Heilman P, Qi J et al (2012) Assessing phenological change in China from 1982 to 2006 using AVHRR imagery. Front Earth Sci 6: 227–236.

Wisz MS, Hijmans RJ, Li J et al (2008) Effects of sample size on the performance of species distribution model. Divers Distrib 14: 763–773

Woomer PL, Bekuna MA, Karanja NK et al (1997) Agricultural resource management by smallhold farmers in East Africa. Nature and Resources 34: 22-33.

World Bank (1989) Sub-Saharan Africa: from crisis to sustainable growth, a long-term perspective study. World Bank, Washington DC, USA.

World Bank (2008) The growth Report: Strategies for Sustained Growth and Inclusive Development. Commission on Growth and Development, Washngton D.C.

Yackulic CB, Chandler R, Zipkin EF et al (2013) Presence-only modelling using MAXENT: When can we trust the inferences? Methods Ecol Evol 4: 236–243.

Yu C, Ai-Hong Z, Ai-Jun R et al (2014) Types of Maize Virus Diseases and Progress in Virus Identification Techniques in China. J Northeast Agric 21(1): 75-83.

Zhang H, Lan Y, Suh CP et al (2012) Differentiation of cotton from other crops at different growth stages using spectral properties and discriminant analysis. T ASABE 5: 1-8.

Zhang J, Huang Y, Yang G et al (2016) Using satellite multispectral imagery for damage mapping of armyworm (Spodoptera frugiperda) in maize at a regional scale. Pest Manag Sci 72: 335-348.

Zhang J, Pu R, Huang W et al (2012) Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses. Field Crops Res 134: 165-174.

Zhang L, Liu S, Sun P et al (2015) Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. PLoS ONE 10: e0120056.

Zhu Z, Woodcock CE (2012) Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens Environ 118: 83-94.

Zhu Z, Woodcock CE (2012) Potential Geographic Distribution of Brown Marmorated Stink Bug Invasion (Halyomorpha halys). PLoS ONE 7: e31246.

Zimmermann NE, Edwards TC, Moisen GG et al (2007) Remote sensing-based predictors improve distribution models of rare early successional and broadleaf tree species in Utah. J Appl Ecol 44: 1057–1067.

Zou H, Gowda S, Zhou L et al (2012) The Destructive Citrus Pathogen 'Candidatus Liberibacter asiaticus' Encodes a Functional Flagellin Characteristic of a Pathogen-Associated Molecular Pattern. PLoS ONE 7: e46447.

7 ACKNOWLEDGEMENTS

Primarily, I would like to convey special gratitude and appreciation to International Center for Insect Physiology and Ecology (*icipe*) for scholarship awarded to conduct doctoral study and dissertation. Without the support provided through drip scholarship, the thesis would not have been successfully accomplished. I also would like to acknowledge the Center for Development Research and the University of Bonn who supported and facilitated my doctoral research.

Successful completion of the dissertation was possible with the help and assistance of several people. First and foremost, I am extremely grateful to my advisor, Prof. Dr. Christian Borgemeister, for his valuable guidance and consistent encouragement I have received during the research. I consider it as a great chance to conduct my doctoral research under his guidance and to learn from his expertise to overcome challenges despite his busy schedules. I do also hereby acknowledge my doctoral dissertation tutor Dr. Tobias Landmann, for he has given me additional inspiration to work harder and consulting. Without his valuable advices and valuable supervision, it would be impossible to finalize this doctoral dissertation. Thanks so much to Prof. Dr. Klaus Greve for his co-supervision of the doctoral research.

In addition, I would like to thank Maike Retat-Amin for their precious help and support during my stay and study in Germany. In the meantime, I would also like to acknowledge my friends and colleagues at *icipe*, especially for their unconditional support. My deepest thanks to my parents, Gideon and Esther, for their love and wholehearted support, encouragement and inspiration. My profound gratitude to my beloved wife Emily for her invaluable emotional support and high empathy during the hardest times of my study. You supported me in every possible way to see the completion of my PhD. Finally, I appreciate my son Lian Mumo, you are such an inspiration to me as a father and sincerly dedicate this PhD to you as a motivation to your endeavours. May you live to be a scientist.