The Remote Sensing of Insect Defoliation in Mopane Woodland

Samuel Adewale Adelabu

A thesis submitted to the College of Agriculture, Engineering and Science, at the University of

KwaZulu-Natal, in fulfilment of the academic requirements for the degree of

Doctor of Philosophy in Environmental Sciences

December 2013

Pietermaritzburg

South Africa

ABSTRACT

Mopane (*Colophospermum mopane*) woodlands are a source of valuable resources that contribute substantially to rural economies and nutrition across Southern Africa. However, a number of factors such as over-harvesting and climate change have brought the sustainability of the mopane woodland resources into question. Insect defoliation remains a major factor contributing to the depletion of woodland resources in rural areas resulting in low vitality and productivity of the woodland. Conventional methods (e.g. visual evaluation) have been used in monitoring insect defoliated areas in the past. These methods are costly and time-consuming, because of the need to collect data immediately before and after an extreme event. In this regard, remote sensing techniques offer a practical and economical means of quantifying woodland degradation over large areas. Remote sensing is capable of providing rapid, relatively inexpensive, and near-real-time data that could be used for monitoring insect defoliation especially in semi-arid areas where data collection may be difficult.

The present study advocates the development of techniques based on remotely sensed data to detect and map defoliation levels in Mopane woodland. The first part of the study provides an overview of remote sensing of insect defoliation, the implications for detecting and mapping defoliation levels as well as the challenges and need for further research especially within Mopane woodland.

Secondly, the study explored whether Mopane species can be discriminated from each of its co-existing species using remote sensing. This was done as a prerequisite for classifying defoliation on mopane trees. Results showed that, with limited training samples, especially in semi-arid areas, Mopane trees can be reliably discriminated from its co-existing species using machine learning algorithms and multispectral sensors with strategic bands located in sensors such as RapidEye. These positive results prompted the need to test the use of ground based hyperspectral data and machine learning algorithm in identifying key spectral bands to discriminate different levels of insect defoliation. Results showed that the random forest algorithm (RF) simplified the process and provided the best overall accuracies by identifying eight spectral wavelengths, seven of which belongs to the red-edge region of electromagnetic spectrum. Furthermore, we tested the importance of the red-edge region of a relatively cheaper RapidEye imagery in discriminating the different levels of insect defoliation. Results showed that the red-edge region played an important role in mapping defoliation levels within Mopane woodland with NDVI-RE performing better than the traditional NDVI.

Thirdly, the study tested the reliability and strength of the internal validation technique of RF in classifying different defoliation levels. It was observed that the bootstrapping internal estimate of accuracy in RF was able to provide relatively lower error rates (0.2319) for classifying a small dataset as compared to other validation techniques used in this study. Moreover, it was observed that the errors produced by the internal validation methods of RF algorithm was relatively stable based on the confidence intervals obtained compared to other validation techniques.

Finally, in order to evaluate the effects of insect defoliation on the biophysical properties of mopane canopies at different defoliation levels, the study estimated leaf area index (LAI) of different defoliation levels based on simulated data. This was done using PROSAILH radiative transfer model inverted with canopy spectral reflectance extracted from RapidEyeRapidEye imagery by means of a look-up-table (LUT). It was observed that the significant differences exist between the defoliation levels signifying reduction in the LAI as a result of the defoliation. Furthermore, results showed that the estimated LAI was in the range of those reported in literature. The NDVI-RE index was the most strongly correlated with the estimated LAI as compared to other variables (RapidEye bands and NDVI).

Overall, the study demonstrated the potential of remote sensing techniques in discriminating the state of Mopane woodland after insect defoliation. The results are important for establishing an integrated strategy for managing defoliation processes within Mopane veldt, thereby satisfying both the needs of local populations for Mopane trees and the worms.

PREFACE

The research work described in this thesis was carried out in the School of Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, from January 2011 to November 2013, under the supervision of Prof. Onisimo Mutanga (School of Environmental Sciences, University of KwaZulu-Natal; South Africa).

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

Samuel Adewale Adelabu Signed: _____ Date: _____

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

1. Prof. Onisimo Mutanga Signed: _____ Date: _____

DECLARATION 1-PLAGIARISM

I, Samuel Adewale Adelabu, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   a. Their words have been re-written, but the general information attributed to them has been referenced.
   b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged and the source being detailed in the thesis and in the References section

Signed_____

DECLARATION 2- PUBLICATION AND MANUSCRIPTS

1. S. **Adelabu**, O. Mutanga, and M. A. Cho, "A review of remote sensing of insect defoliation and its implications for the detection and mapping of *Imbrasia belina* defoliation of mopane woodland," *African J. Plant Science and Biotechnology*, vol. 6, pp. 1–13, 2012.

2. S. **Adelabu**, O. Mutanga, E. Adam and M. A. Cho, "Exploiting Machine Learning Algorithms for Tree Species Classification in a Semi-Arid woodland Using RapidEye image," *Journal of Applied Remote Sensing*, 7 (1), 073480, 2013.

3. S. **Adelabu**, O. Mutanga, E. Adam and R. Sebego, "Spectral Discrimination of Insect Defoliation levels in Mopane Woodland using Hyperspectral Data ," *IEEE Journal of Selected Topics in Earth Observation and Remote Sensing*, 7 (1), 177-186, 2014.

4. S. **Adelabu**, O. Mutanga, and E. Adam, "Evaluating the impact of red–edge band from RapidEye Image for classifying insect defoliation levels" *ISPRS Journal of Photogrammetry and Remote Sensing*. In Review

5. S. **Adelabu**, O. Mutanga, and E. Adam, "Testing the reliability and stability of the internal accuracy assessment of Random Forest when compared with independent dataset" *International Journal of Remote Sensing*. In Review

6. S. **Adelabu**, O. Mutanga, M.A. Cho, "Estimating Leaf Area Index of Insect Defoliated Canopy Levels Using Inversed PROSAIL H Simulated data". In Preparation

Signed_____

DEDICATION

This project is dedicated to *the Author and Finisher of my Faith* who in His Divine mercy

has given me the grace to witness the end of the programme. I also dedicate this project to my

lovely wife and precious daughter who have been my bedrock.

ACKNOWLEDGEMENTS

I also appreciate the effort of my heartthrob (**Mrs Oluwadolapo Adelabu**) and my precious daughter (**Moyinoluwa Adelabu**); *Dola*, you are not only my wife but a God sent helper to me and *Moyin*, it was during the course of my study that the Lord added you to us and you are so considerate even when dad is not around, ***it is my prayer that the Lord will continue to keep you.***

Finally, for everyone who I have not mentioned but have in one way or the other contributed to my success, ***I pray that the good Lord will reward you abundantly. (Amen).*** Thank you so much.

"*Let the words of my mouth and the meditation of my heart be always acceptable and pleasing unto you oh my rock*"

SAMUEL ADEWALE ADELABU

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

General Introduction

1.1. Mopane *(Colophospermum mopane)* Woodlands in Southern Africa

Mopane woodlands are a source of valuable resources that contribute substantially to rural economies and nutrition across Southern Africa (Hrabar *et al.*, 2009a). The woodland provide varied products that include construction and fence poles, wood for tools hands, carvings and utensils, firewood, rope, gum, tannin, medicines and resin, green manure, livestock browse and edible caterpillars (commonly referred to as mopane worms) (Mojeremane and Kgathi, 2005). Mopane woodland is widely distributed in Southern Africa (**Fig 1.1**). It can be found in Angola, Botswana, Malawi, Mozambique, Namibia, South Africa and Zimbabwe (Makhado *et al.*, 2012; Stack *et al.*, 2003) accounting for about 550,000km^2 in total area of southern Africa (Wessels *et al.*, 2007). Specifically, mopane woodland occurs in northern Namibia; the Caprivi strip; north-eastern Botswana; north; south and west of Zimbabwe; southern parts of Malawi; southern and Central Mozambique; southern Angola and parts of the Northern and Mpumalanga Provinces of South Africa (Mapaure, 1994).

The value of mopane woodland in Botswana alone has been estimated at £1.99 million per annum while in South Africa, the mopane veldt is worth £57 million, of which approximately 40% goes to producers who are primarily poor rural women (Stack *et al.*, 2003).

Colophospermum (*C) mopane* is the dominant tree species in mopane woodlands (Lėonard, 1949). *C. mopane* is commonly known as mopane and is considered as one of the most important tree species in the mopane woodlands. However, several other tree species grow in association with mopane trees. The resource which is extensively utilized in mopane woodland is mopane worms. Mopane worms are extensively consumed by the rural and increasing urban populations across southern African countries for their nutritional value, and also sold to generate income. Mopane woodland also provides nutritious fodder for browsers, particularly in the dry season (Timberlake, 1996). Its leaves form an important source of crude protein and are preferred by browsers during winter when the tannins have leached-out (Hrabar *et al.*, 2009a). The grasses under *C. mopane* tree are quite nutritious and are highly palatable for grazing animals.

Recent studies have however showed that the long term sustainability of the woodland and its resources are under threat (Adelabu *et al.*, 2012; Ditlhogo *et al.*, 1996). A number of studies have been carried out on different management prospects of mopane woodland in Southern Africa (Mojeremane and Kgathi, 2005; Timberlake, 1996), however, only a few have studied the impacts of mopane worms on their host (Ditlhogo *et al.*, 1996; Hrabar *et al.*, 2009a). Moreover,

2

the extent of this defoliation is still relatively unknown. In this regard, further work is needed to fully quantify the influence of the worms on its host, especially the defoliation process, given the availability of emerging specialized technology.

Figure 1. 1: Description of Study Area
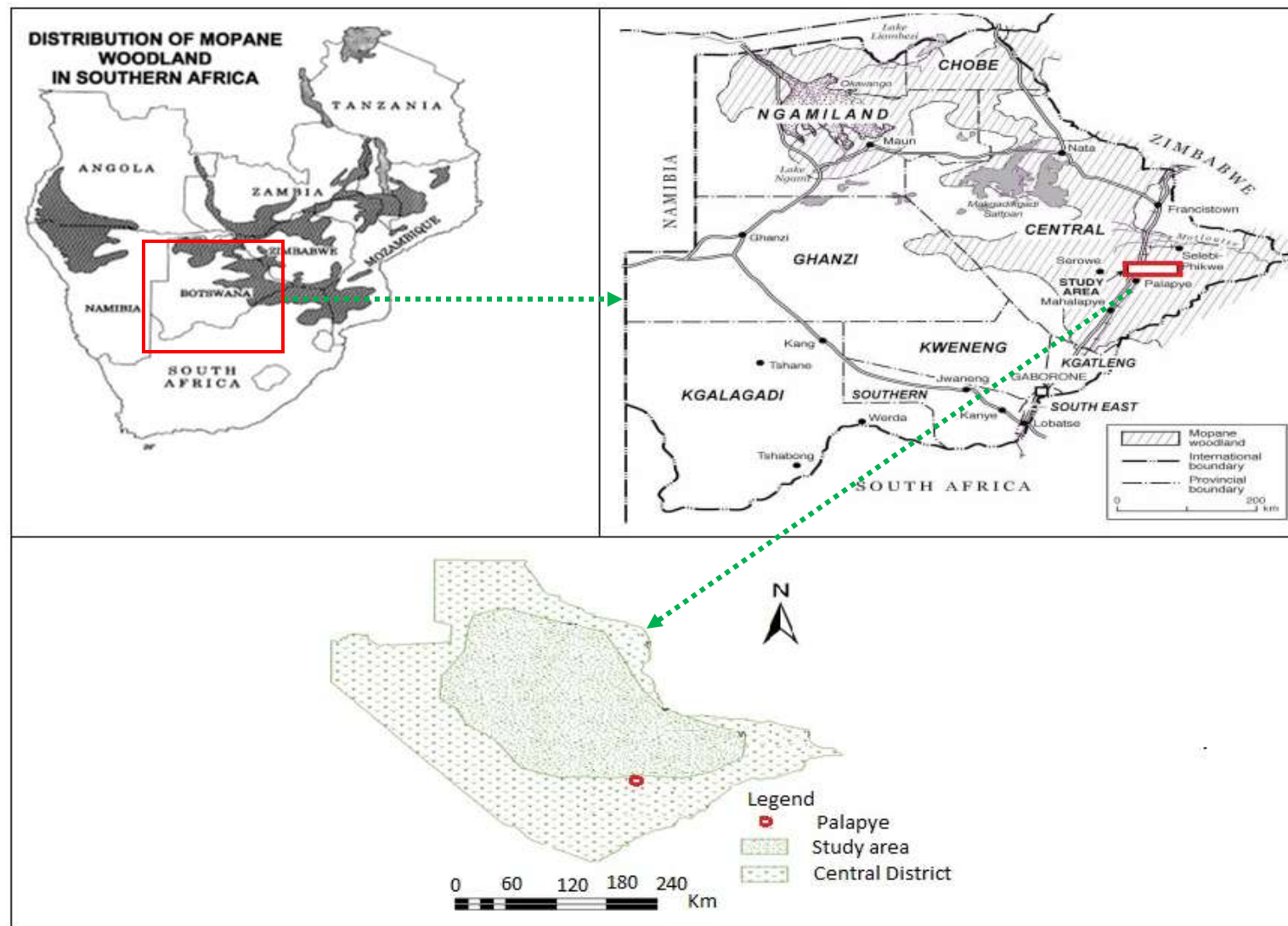
1.2.   Understanding Defoliation Process in Mopane Woodland

The most significant defoliator within the mopane woodland is the mopane worm. Details of what transpires between undefoliated mopane leaves and the worms leading to defoliation are still very sketchy. It is believed that mopane tree defoliation follows the same pattern as other insect defoliators and when the outbreak occurs, about 200 mopane worms feed on a single mopane leaf leading to 90% of mopane trees if not all, left without leaves within a mopane woodland (Ditlhogo *et al.*, 1996). Moreover, Stack *et al.* (2003) observed that mopane trees do not contain any hydrolysable tannin which is widely accepted as being the primary defence compounds against insects. This explains the close association between mopane worms and mopane trees. While mopane woodland often recover within a relatively short period after defoliation with little mortality, continuous defoliation may lead to deplorable long term effects that may be fatal.

Although, no report of eradication of mopane trees as a result of defoliation in any region of mopane veldt is known yet, studies have proved that there are long term effects. Hrabar *et al.* (2009a) noted that at present, defoliation have no effect on mopane plant size. However, the defoliation processes have a potential negative effect on stored resources which characteristically result in regrowth with smaller and or fewer leaves. Styles and Skinner (2000) further explained that heavily defoliated mopane trees tend to lose nutrient content and greatly reduced in size over years. Furthermore, the absence of mopane worms from certain regions of mopane veldt has not been satisfactorily explained. It may be due to the absence of necessary nutrients that attract the worms to the leaves of the tree. The defoliation process as a result of the worms, if not well managed can lead to the extinction of the tree and hence the worms within the region. Few of the studies on mopane woodland management have included defoliation process in their agenda (Ditlhogo *et al.*, 1996; Hrabar *et al.*, 2009a) due to complexity of the defoliation, lack of understanding, unavailability of instruments and methods in quantifying and the perceived non importance of defoliation in the management of mopane woodland. Moreover, most of these studies were carried out using field experiments which have been found to be time consuming and costly.

1.3. Remote Sensing of Mopane Woodland Defoliation

For remote sensing to be effective and accurate in detecting and mapping mopane woodland defoliation, a sound understanding of the distribution, ecology, and life cycle of both the tree and the worm is required. Knowledge of these defoliation processes facilitates the development of algorithms to detect changes in foliar characteristics using remotely sensed data. Consequently, the challenge would be to assess the defoliation process, and then associate each level of observation with different remote sensing data types in order to provide the appropriate level of detail and accuracy for detection and mapping purposes.

1.3.1. The Mopane Tree (*Colophospermum mopane)*

Mopane tree is a monotypic genus which belongs to Fabaceae (legume) family. The species was formerly placed in the genus *Copaifera* L. with the genus *Colophospermum* being created by Léonard in 1949 (Lĕonard, 1949). It can also be categorized into the casalpinioidae subfamily, which are leguminous plants mostly with leaflets like butterfly wings (Palgrave, 1983). *Colophospermum mopane* is known locally by a number of common names such as butterfly tree as it serves as host for edible mopane worm (*Imbrasia belina).*

1.3.1.1. Tree Phenology and Ecology

Colophospermum mopane is a tree or shrub with a heavy rounded but occasionally erect narrow crown (Sebego *et al.*, 2008). Most mopane trees are multi-stemmed and spread upward in a narrow V-shape (Musvoto *et al.*, 2007) and the bole can reach about 40cm but occasionally very big (Palgrave, 1983). Mopane trees differ in growth form depending on the local ecological conditions that vary from dwarf (very old and stunted tree less than 2.5m high) to cathedral mopane, which may reach heights of 18-20m. The leaves are pinnate with two large butterfly-like leaflets that may vary considerably in size on the same tree (Mojeremane and Kgathi, 2005) and within a growing season (Mapaure, 1994). *C. mopane* drops its leaves irregularly from the onset of the dry season but defoliation usually follows the trends of worm consumption and is generally leafless from August to October. However, trees may retain their leaves between successive rainy seasons, depending on the amount and distribution of rainfall (Wessels *et al.*, 2007).

1.3.2. The Worm (*Imbrasia belina)*

Imbrasia belina (Westwood), the mopane worm or Anomalous Emperor Moth, is a saturniid lepidopteran which is widely distributed throughout southern, Central and east Africa. Its

distribution in Southern Africa is in line with the mopane tree which serves as its host extending from the Northern parts of South Africa into Zimbabwe and Botswana, and west into Namibia (Mapaure, 1994). *I. belina* feeds on a number of tree species but *C. mopane* is the most suitable host in terms of developmental periods, number of emerged adults and nutritional quality (Hrabar *et al.*, 2009a).

1.3.2.1.   Life Cycle

According to Paulick (2003), *I belina* is bivoltine (produces two generation within a season) across most of its distribution with one group coming out between November and December and the other between March and April. However, univoltine season has also been observed in areas that is extremely arid (Ditlhogo *et al.*, 1996). A solitary 50 to 200 clusters of eggs are usually laid by adult moths over a period of approximately two month (**Fig 1.2**). Pupation only takes place after 10days when the larva has already emerged and it has passed through five more stages. Paulick (2003) observed that the caterpillars usually feed together in phases 1 to 3. Phase 1 involve the worms feeding on the mopane leaves in order to grow bigger and store energy. At this stage, it is believed that the worms body mass will increase by an approximately 4000 fold. After this phase, the worm moves down the tree into the soil making hole in the ground. At the pupa phase, the worm spends approximately six months in the ground. The final phase occurs at the beginning of the summer when its main function is to mate and die within two – three days (Hrabar *et al.*, 2009b).

Phase 1: EGG
An egg is laid and then 10 days later hatches

Phase 2: Worm
The worm feeds on the mopane leaves of the mopane tree to grow bigger and store energy. They grow 9-10cm

When the mopane worm is fully grown it moves down the tree and make hole in the ground

Phase 3: PUPA
This is the resting stage. The worm will spend 6 month in the ground protected by cocoon

Phase 4: Moth
The moth appears at the beginning of summer and dies after 2-3 days. Its job is to mate. the female lay between 50-200 eggs

The moth is a type of emperor moth called Gonimbrasia belina

Figure 1. 2: Life cycle of the mopane worm modified from Paulick (2003)

1.3.2.2. Population Ecology

Currently, there is no sufficient evidence to establish the population dynamics of *I. belina* (Greyling *et al.*, 2001) but researchers have considerably linked it to the outbreak dynamics similar to those known species of woodland lepidopteron in the Northern Hemisphere. Greyling *et al.* (2001) found that *I. belina* conforms to the picture of a 'typical' outbreak lepidopteron closely that is, it lays eggs in masses, and have large, brightly coloured larvae which start feeding early in the season and which are reasonably polyphagous.

1.3.3. Implication of Remote Sensing for Detecting Defoliation Levels in Mopane Woodland

Having understood the ecology of mopane tree and the worms that feed on it, the major focus will then be on how applicable remote sensing is for defoliation monitoring in mopane woodland. Currently, mopane woodland defoliation is mapped and monitored based on field

observations (Hrabar *et al.*, 2009a). The results from this is however debatable as field observations are known to be dependent on the proficiency of the observer and hence subjective (Coops *et al.*, 2004). However, monitoring of forest health using remote sensing has been given great attention in recent years with diverse range of imageries and modeling techniques (Coops *et al.*, 2004; Pontius *et al.*, 2005; Radeloff *et al.*, 1999). The advantages of applying remote sensing for monitoring insect defoliation includes the ability to acquire relatively cheap and rapid methods of acquiring up to date information over a large geographical area (de Beurs and Townsend, 2008). Also, remote sensing has an edge over other methods because it offer an effective way to acquire data from remote areas and also provides spectral information in the form of individual bands, combination of bands and relevant vegetation indices that could be used for monitoring forest health (Hall *et al.*, 2007; Rullan-Silva *et al.*, 2013; Townsend *et al.*, 2004). To date, no studies have used remote sensing for the detection, mapping and monitoring of insect defoliation in mopane woodland. Nonetheless, there are factors needed to be considered before remote sensing can be used for monitoring insect defoliation in mopane woodland.

Firstly, three stages of defoliation could be observed in mopane woodland. These are the undefoliated, partly defoliated and refoliating stages of defoliation. The time series of defoliation in mopane woodland is represented by its canopy impacts. While the non-impacted undefoliated canopies (UD) represents time before defoliation, the partly defoliated (PD) and refoliating (R) canopies represent during and after defoliation respectively. During the early stages of defoliation in mopane woodland (when the worms are feeding on the tree), the canopy of the tree appears green and cannot be visually differentiated from the healthy trees (Ciesla, 2003). Leaves become pale green gradually leading to total defoliation. At this stage, discriminating the defoliated part of the tree using remote sensing is dependent on detecting the little changes that might have occurred in the spectral reflectance of the tree. However, this challenge in defoliation process of mopane woodland could potentially be reduced through the use of hyperspectral data because of the presence of continuous narrow bands that offer in-depth information which cannot be visually identified and will have been masked by broadband scanners. (Kumar *et al.*, 2001; Mutanga and Skidmore, 2004a; Mutanga *et al.*, 2009). In addition, the time period of defoliation in mopane woodland is relatively short (2-3weeks), hence the period of detection by remote sensing will be restricted. Thus analyzing images of outbreaks may be challenging as changes in canopy characteristics may not be easily detected. To this end, remote sensing sensors that

incorporate the period of defoliation (i.e. short temporal resolution) and also spectrally capable of quick detection of subtle changes in canopy reflectance are needed.

Nevertheless, the focus of the present thesis was limited to hyperspectral data captured in the field due to cost and availability of airborne hyperspectral imagery. Furthermore, the study investigates the potential of down-scaling the hyperspectral data to readily available multispectral image in classifying the defoliation levels. Although hyperspectral data could potentially be used for discriminating the defoliation levels, it has been observed to be difficult in processing as compared to multispectral data due to geometrical and statistical properties associated with high dimensional data (Ismail and Mutanga, 2011). The challenge would therefore be to develop and test robust methods and techniques such as random forest and support vector machine for the effective processing and classification of hyperspectral data for defoliation monitoring. The use of random forest and support vector for classifying insect defoliation is expected to limit the errors that is usually obtained from traditional classifiers such as maximum likelihood. Additionally, multispectral imagery with strategic band could assist where hyperspectral data are not available. Likewise, the study test the capability of leaf area index derived from radiative transfer model in distinguishing the different defoliation levels in the woodland. The knowledge of biophysical variables such as LAI may help to understand the influence of the insects on the canopy at each of the defoliation levels. Radiative transfer model approach for estimating LAI will further help in generating LAI data as data collection becomes more and more difficult to collect especially in developing world (Weiss *et al.*, 2000).

1.4. Aim

From the observations above, the main aim of the study was to classify and monitor the defoliation levels of Mopane trees (*Colophospermum mopane*) by Mopane worms (*Imbrasia belina*) in Eastern Botswana using remote sensing data.

1.5. Research Objectives

The main objectives of the study was to

1. Review the different approaches including the remote sensing platforms and techniques that have been used for assessing insect defoliation and their implications for detecting

and monitoring mopane worm defoliation of mopane woodland highlighting their strengths and weakness.

2. Discriminate *Colophospermum mopane* and its co-existing species in a semi-arid forest using spectral information provided by new multispectral sensors together with advanced classification algorithms.

3. Discriminate the levels of change in forest canopy cover detectable after insect defoliation by ground based hyperspectral measurements in mopane woodland.

4. Explore and evaluate the benefit of the RapidEye red edge channel for discriminating different levels of insect defoliation in mopane woodland.

5. Test the reliability and robustness of the internal accuracy estimate in random forest (RF) ensemble classifier in discriminating different levels of insect defoliation in mopane woodland.

6. Quantify the impact of insect defoliation on the leaf area index of mopane canopies by estimating leaf area index at different defoliation levels using radiative transfer model.

1.6. Description of Study Area

The study was based in the eastern part of Central District of Botswana. The approximate location of the study area is between 27°E and 27°33`E and 22°23`S and 22°52`S. The study area is located at about 230km north of Gaborone (the capital city) and approximately 70km north of the tropics of Capricorn. The study area include a number of small villages and a major centre, Palapye (**Fig 1.1**) that was of importance for this study for logistical purposes as well as being a trading place for commercial, agricultural and veld products such as mopane worms. The area immediately west of Palapye forms a transition zone into the Kalahari, and it is locally known as the Sandveld (Sebego *et al.*, 2008). Some few kilometres south of Palapye lies the most southerly area where the mopane species can be found in its natural environment. The selection of this area was based on two major reasons: firstly, it is an area with typical pockets of mopane woodland where the species is found in all its various growth forms (tree, tall shrubs and short shrubs). Secondly, 2km north of Radisele lies the southern limit of mopane in Botswana (the other part being in the northern province of South Africa). This area is anticipated as appropriate for investigations on the environmental factors limiting the distribution of mopane. According to Botswana's department of meteorological services, summer months are from mid-September to mid-April. Summers are very hot in Palapye, with daytime temperatures reaching up to 40°C.

Thunderstorms are plentiful in the summer. Rainfall levels are usually a minimum of 300mm annually. Winters in Palapye are mild and dry.

1.7.   Thesis Outline

In order to achieve the objectives of this study, the thesis is organized as compilation of 6 research articles that have been submitted to peer reviewed international Journals. Of these, three have already been published; two are still in review while one is in preparation. Each chapter has been written as a stand-alone article that can be read independently from the rest of the thesis but that draws conclusion which is linked to the overall research objective. Consequently, a number of replications and overlaps crop up in the "Introduction" and "Methods" section of different chapters. The thesis consists of 8 Chapters:

Chapter 1 serves as an introduction to the study.

Chapter 2 contains a detailed literature review of the current state of insect defoliation mapping and monitoring using remote sensing. It also highlights the relevant application of remote sensing in mapping insect defoliation in mopane woodlands. The research gaps and challenges in the application of remote sensing in insect defoliation mapping are also introduced. In order to be able to classify insect defoliation levels in mopane woodland, there is need to be able to discriminate mopane trees from other tree species present in the woodland. Chapter 3 therefore contains an investigation into the ability of new generation RapidEye image to discriminate between mopane trees and its co-existing species. Furthermore, the capabilities of relatively new machine leaning classifiers (RF and SVM) were compared for tree species classification in a semi-arid environment with limited training samples.

In chapter 4, we tested the ability of ground based hyperspectral data in discriminating different insect defoliation levels. The chapter determines if there is any significant difference in the mean reflectance between the different defoliation levels at each measured wavelength from 325nm to 1075nm. For the wavelengths that are significantly different ($p < 0.001$), it was tested whether some wavelengths have more discriminating power than others and which wavelength combinations can yield the lowest misclassification rate.

Chapter 5 investigates the potential of downscaling the sensors used in Chapter 4 to cheaper and readily available new generation sensors such as RapidEye in classifying the insect defoliation

levels in mopane woodland. Specifically, the study examined the importance of the new red-edge band present in RapidEye image in classifying insect defoliation levels using two different machine learning algorithms. Chapter 6 is based on testing the robustness of internal accuracy estimate of one of the classifier used for classification in Chapter 4 and Chapter 5 for insect defoliation mapping.

In Chapter 7, LAI of insect defoliated canopy were estimated using radiative transfer models to understand the impacts of defoliation on LAI.

Finally, a synthesis of the study is provided in Chapter 8. The findings are summarized and conclusions are derived from the preceding chapters. Some relevant recommendations for future research on the applications of remote sensing for insect defoliation mapping are highlighted. A single reference list is provided at the end of the thesis.

CHAPTER TWO

Literature Review

This chapter is based on:

ABSTRACT

Forest health, especially insect defoliation monitoring using direct sampling and visual estimation has been only moderately successful due to its cost, time required for sampling, and most importantly the need to collect data immediately before and after an extreme event. However, remote sensing techniques offer timely, up-to-date, and relatively accurate information for sustainable and effective management of forest health. In this paper, we discuss the different approaches including the remote sensing platforms and techniques that have been used for assessing insect defoliation and its implications for detecting and monitoring mopane worm defoliation of mopane woodland, highlighting their strengths and weakness. Research gaps in the detection of insect defoliation with remote sensing are highlighted and future directions of research are also proposed.

Keywords: forest health, mopane worm, multispectral, hyperspectral, chlorophyll content

2.1 Introduction

An important component of forest ecosystem is its health status and the impact it has on sustainable growth. Recent evidence suggests that new damaging agents are appearing at an increasing rate which could affect the future sustainability of forest industries (Wulder and Franklin, 2003). While many of the past impacts of damaging agents such as insects on forest woodland have been disastrous, mopane worm, an important defoliator of mopane woodland in Southern Africa exhibit a different scenario. As a result, *Imbrasia belina* (mopane worm) is widely distributed and consumed in Southern Africa because of its nutritional values and sold to generate income (Timberlake, 1996).

However, while the depletion of worms derived from mopane woodland have been reported in different areas, none of these depletions have been attributed to the impacts of the worms on the vitality and productivity of their host. Mopane tree defoliation is one of the serious impacts of the worm on its host. Furthermore, the absence of mopane worms from certain regions of mopane veldt has not been satisfactorily explained. It may be due to the absence of necessary nutrients that attract the worms to the leaves of the tree. Defoliation process as a result of the worms if not well managed can in the long run lead to the extinction of the tree and hence the worms within the region. In an effort to minimize the potential loss of *I. belina* (hereafter referred to as mopane worm) in mopane woodland of Southern Africa, an integrated management strategy is needed combining detection, mapping and monitoring methods. Moreover, resource managers need to know the impacts, vulnerability and suggest possible management practices that will enable efficient and sustainable use of the resources emanating from mopane woodland.

Information on the extent and severity of mopane defoliation is required for a wide variety of forest planning, management, and modeling activities. Mapping mopane defoliation will also aid sketch mapping surveys and also help in reporting and assessing the impacts of the defoliation on the health and productivity of the woodland. Currently, there are no specific methods of mapping mopane defoliation.

The objective of this paper therefore is to discuss the different approaches including the remote sensing platforms and techniques that have been used for assessing insect defoliation and its implications for detecting and monitoring mopane worm defoliation of mopane woodland

highlighting their strengths and weakness. Firstly, we review the effects of insect defoliation on trees and the conventional ways in which they are assessed. Thereafter we consider different remote sensing platforms that have been used in detecting insect defoliation, highlighting the strengths and weaknesses in detecting, mapping and monitoring mopane woodland defoliation. Thirdly, the remote sensing techniques that can be used for accurate monitoring are presented. Finally, we discuss various challenges that might occur while using remote sensing to detect, map and monitor defoliation in mopane woodland by mopane worms, suggesting possible solutions to them.

2.2 Effects of Insect-Caused Tree Defoliation on Vegetation Productivity

The primary function of leaves in plants is to manufacture sugars and carbohydrates (Morgan *et al.*, 2010). Sugars and carbohydrates are the basic food or energy that plants use for all metabolic activities such as growth, root development, flower and seed production, disease resistance. Leaves also provide many indirect benefits such as emitting oxygen, screening out particulates and other air pollutants, intercepting precipitation to minimize erosion and shading the ground to modify surface temperatures (Morgan *et al.*, 2010). When insect defoliation occurs in a particular tree, the effects range from a slight reduction in vigor to total death (Hall *et al.*, 2003a). Insect defoliation harms plants by eliminating or limiting their food production capability (Hall *et al.*, 2003a). The refoliation process, which frequently occurs immediately after defoliation, also requires energy for budbreak and leaf expansion, which causes further depletion of stored food reserves (Hall *et al.*, 2003a).

The inability of the tree to manufacture food (energy) together with the depletion of stored food weakens the tree and results in reduced growth, stunted, pale-green new leaves and possibly twigs and branch dieback (Kantola *et al.*, 2010). Insect defoliation also affects the morphological and physiological characteristics of trees, and it is these characteristics that govern how trees absorb and reflect light (Hall *et al.*, 2003a). The production of protective substances that aid in disease resistance may be inhibited (Hall *et al.*, 2003a). It is predicted that the frequency and severity of insect defoliation outbreaks could increase in response to climatic warming, further magnifying their effects (Fraser and Latifovic, 2005).

Mopane tree defoliation follows the same pattern as other insect defoliators and when the outbreak occurs, about 200 mopane worms feed on a single tree leading to 90% of mopane trees

if not all left without leaves within a mopane woodland (Ditlhogo *et al.*, 1996). Moreover,Stack *et al.* (2003) observed that mopane trees do not contain any hydrolysable tannin which is widely regarded as the primary defence compounds against insects. This explains the close association between mopane worms and mopane trees. While mopane woodland often recover within a relatively short period after defoliation with little mortality, continuous defoliation may lead to deplorable long term effect that may be fatal.

Although, no report of eradication of mopane trees as a result of defoliation in any region of mopane veldt is known yet, studies have proved that there are long term effects. Hrabar *et al.* (2009b) noted that at present, defoliation has no effect on mopane plant size, however, it has potential negative effects on stored resources which characteristically result in regrowth with smaller and or fewer leaves. Styles and Skinner (2000) further explained that heavily defoliated mopane trees tend to lose nutrient and greatly reduce in age over years. Having discussed the (possible) effects of mopane worm defoliation on mopane woodland, it is important to highlight the linkage between (biophysical and biochemical) indicators of mopane woodland productivity and remote sensing. Knowledge of this will help accurate detection of defoliation level within the woodland.

2.3 Conventional Methods of Assessing and Monitoring of Insect-Caused Tree Defoliation

Defoliation is a general stress response, and it is closely linked to biophysical and biochemical indicators hence they are used as conventional methods of detecting insect defoliation (Lee *et al.*, 2010). Measuring forest biophysical characteristics aims at documenting forest integrity in many aspects, such as structural, functional and species diversity (Kumar *et al.*, 2001). However, these measurements often depend on extensive and expensive fieldwork, encompassing a restricted study area. Remote sensing enables monitoring studies in a wide area at constant time periods (Seidl *et al.*, 2011).

The alliance between remote sensing techniques and biophysical indicators could be valuable to studies on detecting, mapping and monitoring defoliation process in forests. To fully understand how remote sensing can be used for detecting, mapping and monitoring insect defoliation especially the mopane woodland defoliation, we need to discuss the biophysical and biochemical variables that affect mopane trees focusing on the two main ones: Leaf Area Index

(LAI) and chlorophyll content. Understanding these biophysical indicators could help in providing more information on the techniques and platforms of remote sensing that can be applied.

2.3.1 Leaf Area Index

The use of biophysical change metrics, such as LAI change, has been proved to provide a more flexible and general defoliation mapping method (Hall *et al.*, 2003a). Moreover researchers have stated that insect defoliation thresholds that are based on LAI rather than percent defoliation are more meaningful (Malone, 2001). LAI is an important variable explaining canopy primary production and can be used to infer processes such as photosynthesis, transpiration, evapotranspiration and estimate net primary production (NPP) of terrestrial ecosystems (Yao *et al.*, 2008). LAI is defined as one-half the total surface area of leaves per unit ground area. The estimation of LAI from remote sensing measurements has received much attention. For example, a simplified semi-empirical reflectance model for estimating LAI of a green canopy was introduced by (Clevers, 1997). The widely used crown condition variables are closely related to LAI, but this has received little attention. As such, LAI is increasingly desired as a spatial data layer (i.e., map), to be used as input for modeling biogeochemical (Thenkabail *et al.*, 2000).

The LAI measurements are relevant for comparing the condition of differently damaged stands and can therefore be used in forest monitoring practice (Thenkabail *et al.*, 2000). Measuring LAI on the ground is difficult and requires a great amount of labor and cost (Kumar *et al.*, 2001). To produce a LAI map of a large area, a model relating field data with remote sensing data is typically developed, the model is inverted, and the remote sensing data are then used to extrapolate that relationship to the landscape (Hall *et al.*, 2003a). Many studies have sought to establish relationships between LAI and remote sensing data (Kumar *et al.*, 2001; Thenkabail *et al.*, 2000; Yao *et al.*, 2008). Most of these studies have relied on empirical relationships between the ground-measured LAI and observed spectral responses, although several have used canopy reflectance models (Kumar *et al.*, 2001; Thenkabail *et al.*, 2000).

Although, LAI has not been used in detecting and mapping the forest health levels of mopane woodland, especially during mopane defoliation, it is hypothesized that data from LAI could evaluate the vegetation levels before, during and after defoliation. When forest health deteriorates and the deterioration is affecting canopy volume it would be detected as LAI change.

Therefore, the healthier the vegetation, the higher the LAI, since LAI increases with healthy status of plants (Sanz-Cortiella *et al.*, 2011). Hence, a forest that is highly defoliated is expected to have low LAI. LAI during the healthy state of mopane woodland (without defoliation) is expected to be high since its canopies at this stage are still very green and have not been attacked by the worms. However it may be difficult to differentiate the early defoliation stage of mopane woodland from the healthy state using LAI since the canopy at this stage is visually indistinguishable from healthy trees of the green stage (Ismail *et al.*, 2008). Combination of LAI as well as the variation in biochemical concentration in leaves could help in dissociating this level. On the basis of the above discussion, measurement of LAI can help develop a background to which remote sensing techniques could be applied for detecting, mapping and monitoring defoliation process in mopane woodland.

2.3.2 Chlorophyll Content

Chlorophyll (Chl) content is another biophysical variable for detecting insect defoliation on forest (Thomas *et al.*, 2008). Chl content is a good indicator of vegetation status and gross primary productivity because of its direct role in photosynthesis (Gitelson *et al.*, 2006). Results in the past have showed that Chl content was much lower in woodlands that have insect defoliation when compared with healthy woodland (Gitelson *et al.*, 2002). When forests are subjected to insect defoliation, many physiological changes occur, including: reductions in photosynthetic activity (Zarco-Tejada *et al.*, 2000), inhibition of Chl formation (Sims and Gamon, 2002), and an increasing breakdown of the chlorophyll molecule (Gitelson *et al.*, 2006). Efficient field measurements of these Chl related changes have been approximated using measures of Chl fluorescence (a measure of photosynthetic activity (Zarco-Tejada *et al.*, 2000).

However, this has been costly and time consuming. Recently, a relatively cheaper and less time consuming approach of detecting defoliation using Chl content over large areas involves remote sensing technology (Thomas *et al.*, 2008). Narrow wavebands near 700 nm where changes in Chl absorption are easily detectable have been recommended for early detection of forest damage (Pontius *et al.*, 2005).

The Chl in green leaves absorbs light for photosynthesis at wavelengths from 650–660 nm (Thomas *et al.*, 2008). For this reason, the red region of the spectrum is most useful for detecting the absorption of visible light by the Chl pigments. The healthiest vegetation will perform

photosynthesis efficiently, which requires an abundance of Chl pigments. The healthiest vegetation, hence, will absorb the greatest amount of red light. Most of the infrared light incident on a green leaf is reflected at wavelengths from 0.7–1.2µm due to leaf internal scattering. The near-infrared region of the spectrum is most useful for detecting the reflection of infrared light by the leaves (Pontius *et al.*, 2005). The healthiest vegetation will have many leaves and will, therefore, reflect the greatest amount of near-infrared light. Hence, healthy vegetation is highly reflective in the near infrared region and highly absorbent in the red region.

Also narrow-band hyperspectral instruments have the capability to identify early signs of defoliation in some cases even when symptoms are not visible to the human eye (Mohammed *et al.*, 1995; Pontius *et al.*, 2005; Zarco-Tejada *et al.*, 2000). Physiologically, this can be explained by the tendency of defoliated forest to reduce photosynthetic activity and hence Chl content. Even subtle changes in Chl content can alter reflectance patterns in the visible and near-infrared (NIR) portions of the spectrum (Pontius *et al.*, 2005).

While Chl content can be measured directly using Chl meter such as Minolta SPAD-502 (Konica Minolta, Osaka, Japan), most studies using Chl content in monitoring defoliation make use of models that are derived from empirical relationships between the ground-measured Chl content and observed remote sensing variables. Moreover, with high forest canopy cover such as mopane woodland, relationships between the reflected electromagnetic radiation and leaf chemistry tend to break down (Pontius *et al.*, 2005). However, Chl content derived from hyperspectral remote sensors may be a good indicator of defoliation at the green stage of defoliated mopane woodland since the photosynthesis activities at this stage is relatively higher than when they are totally defoliated. It must however be noted that although Chl content is dependent on other factors such as drought, pest infestations and diseases, mopane woodland insect defoliation will be different in that defoliation follows good-healthy growth patterns that would otherwise not occur if there was a severe drought pest or diseases infestation. Moreover, insect defoliation occurs in patches (within the healthy forests) but drought or pest or diseases infestation would impact on all vegetation.

Changes in Chl content and LAI have been related to variation in photosynthetic activities of deciduous trees (Koike, 1987). Although they (LAI and Chl content) are not direct measurements of vegetation productivity and physiological activities, they represent important determinants of

productivity and physiological capacity of plants (Sims and Gamon, 2002). Infact, relationship between the two can actually provide information on the health status of a particular tree (Kodani *et al.*, 2002). It is expected that the knowledge about the dynamics will establish the impact of defoliation on the tree.

2.4 Relevance of Remote Sensing in Assessing and Monitoring Insect-Induced Tree Defoliation

The most reliable method of measuring defoliation is by direct sampling (ground based measurement), which is obviously unreasonable because of its cost, time required for sampling, and most importantly the need to collect data immediately before and after an extreme event (de Beurs and Townsend, 2008). For large areas, aerial survey is more efficient than ground based measurement. However, ground based estimates provide better tree specific information (Ciesla and Acciavatti, 1982). Information on defoliation prior to 1947 was limited to records from ground observations, memoranda and letters (Dolph, 1980). Since 1947, when an aerial survey program was initiated, detailed information and forest pests especially in North America and Europe have been collected annually. The remote sensing approach in assessing and monitoring insect defoliation has been to relate differences in spectral response to chlorosis (yellowing), foliage reddening, or foliage reduction over time, assuming that these differences can be interpreted, classified, or correlated to damage caused by insect activity (Franklin, 2001). Remote sensing has been used to generate more spatially precise and detailed defoliation maps from which its impact on the forest resource could be determined.

The range of remote sensing applications has included detecting and mapping defoliation, characterizing patterns of disturbance, modeling and predicting outbreak patterns, and providing data to pest management decision support systems (Lee *et al.*, 2010). The possibility of forecasting the susceptibility and vulnerability of forested areas to insect defoliation has also been reported as a tool to provide mitigation options to forest managers (Luther *et al.*, 1997). These applications were intended to produce information products that support pest management planning. The advantages of applying remote sensing for monitoring insect defoliation includes the ability to acquire relatively cheap and rapid method of acquiring up to date information over a large geographical area (de Beurs and Townsend, 2008). Also remote sensing has an edge over other methods because it is the only practical way to obtain data from inaccessible regions; it has

ability to map both small and large scales for easy identification and its ability to image defoliation in different spectral forms.

One of the earliest research of using remote sensing to monitor defoliation was conducted in north Central Washington and Central Idaho in USA using aerial photographs (Heller *et al.*, 1981). Ciesla and Acciavatti (1982) determined that high altitude panoramic color infrared photography acquired during the time of peak defoliation could consistently differentiate between heavy defoliation, moderate defoliation, and no defoliation. Ever since, the use of remote sensing technology to detect, map and manage forest defoliation over large region has been a subject of intense interest (de Beurs and Townsend, 2008).

In mapping areas covered with mopane woodland, Sebego and Arnberg (2002) used coloured infrared photographs and they discovered that though mopane woodland extent and distribution can accurately be mapped using colour infrared photographs, it may not be able to discriminate defoliated from undefoliated mopane woodland. Moreover, using aerial photographs alone was also time consuming and expensive since photographs need to be taken for every events of the defoliation process. It must however be noted that aerial photographs can form a bases on which other forms of remote sensing platform can be used in monitoring mopane defoliation.

2.5 Developments in the Remote Sensing of Insect-Induced Tree Defoliation

The increasing availability of remote sensing and geographic data has not only helped the detection, mapping, monitoring and management of the health of forest ecosystems especially those affected by insect defoliation, but also proved to be important for the protection of natural resources and the economy worldwide (Kantola *et al.*, 2010). Different platforms of remote sensing have been used in the past for forest defoliation monitoring with varying success. A review of the remote sensing methods and platforms that have been used for insect defoliation illustrates the degrees at which they have been successful in obtaining information of operational relevance (i.e., used by those in forest management) (**Table 2. 1**). We discuss the various platforms and their implications in mapping mopane woodland.

Table 2. 1: Sample of Multispectral Remote Sensing Studies Applied in Defoliation

Sensor	Study Area	Image Data Date	Defoliators	Comment	Reference
Landsat-1	Pennsylvania, USA	1975	Gypsy Moth	Classification results were subjectively analyzed and found to be representative of actual ground cover. However, errors of commission in which agricultural cover types were classified as heavy defoliation decreased classification performance.	Williams (1975)
Landsat TM	Wisconsin, USA	1990-1995	Jack Pine Budworm	Classification was successful with single-date imagery but was not tested with other methods such as change detection.	Radeloff *et al.* (1999)
MODIS EVI	Siberia	2002	Silk Moth	Very effective in mapping large-scale conifer mortality and also for near real time monitoring but does not provide links with finer resolution validation data	Kovacs *et al.* (2005)
SPOT	Virginia	1989	Gypsy Moth	Questions regarding the completeness of this classification, citing the unknown effects of terrain and forest type were raised.	Clerke and Dull (1990)
Landsat TM	Canada	July (1999&2001)	Aspen	Good for identification but could not differentiate defoliation where vegetation is dense.	Hall *et al.* (2003b)
Landsat TM & SPOT	Michigan, USA	June 1988	Gypsy Moth	Both can classify defoliation only on a large scale	Joria and Ahearn (1991)
SPOT VGT	Canada	1998-2000	Hemlock Looper	Good for near real-time and identifying occurrence of defoliation but less reliable for classifying intensity of defoliation	Fraser and Latifovic (2005)
Landsat TM	Australia	March 2008	Beetles and Sawfly	Improved accuracy when advance analytical techniques was applied but still coarse for small scale monitoring	Somers *et al.* (2010)
MERIS	Canada	2003-2005	Aspen	Better than Landsat, SPOT, MODIS but unable to evaluate small scale details	Van der Sanden *et al.* (2006)
MODIS	USA	2000-2001	Gypsy Moth	Demonstrated significant relationships between defoliation and vegetation indices estimated at the plot scale.	de Beurs and Townsend (2008)

2.5.1 Broadband Sensors

Various images from remote sensing broadband sensors have been found to effectively monitor insect defoliation in woodland (**Table 2.1**). The resulting data usually classify defoliation in terms of light, moderate and heavy defoliation. It has been demonstrated that data from Landsat and other synoptic scale sensors have an appropriate spatial resolution for monitoring many types of insect defoliation. The advantages and pitfalls of Landsat data were recognized early. Williams (1975) expressed concerns about Landsat-1's ability to effectively monitor insect defoliation with only eighteen-day temporal coverage and the greater than 50% chance of cloud cover during an acquisition over Pennsylvania. Williams and Stauffer (1978) used Landsat imagery acquired before and during gypsy moth defoliation. The investigators recognized that agricultural features could be mistaken for insect defoliation. Moreover, Williams *et al.* (1979) evaluated different types of vegetation indices on Landsat imagery acquired before and during peak defoliation to differentiate between defoliation and healthy forest, however, they could not be distinguished from healthy forest. Radeloff *et al.* (1999) used Landsat Thematic Mapper (TM) TM data to identify the forest attributes that affect jack pine budworm population levels and separate the spectral signatures of these attributes from those of actual jack pine budworm defoliation in Wisconsin.

Hall *et al.* (2003a) also used Landsat multi-temporal change detection approach to map defoliation in insect defoliated forest of Canada with results showing consistency with other studies earlier carried out using the same platform. The various studies have shown the utility of Landsat multi-temporal imageries to identify those characteristics of a forest that make it susceptible to insect defoliation, as well as to identify the actual insect defoliation but could not clearly differentiate defoliation where vegetation is highly saturated. Therefore, it may be difficult for Landsat images to be used for detecting defoliation in heavily populated mopane woodland due to the short window for monitoring and the coarse temporal resolution of Landsat relative to cloud cover.

As an alternative, other remote sensing platforms have demonstrated since the early-1990s, to be effective for insect defoliation detection and mapping. Two of those are the Systeme Probatoire d'Observation de la Terre (SPOT) and National Oceanic and Atmospheric Administration Advance Very High Resolution Radiometer (NOAA AVHRR) imageries (Fraser and Latifovic, 2005; Kovacs *et al.*, 2005). Clerke and Dull (1990) determined the extent and

severity of gypsy moth defoliation in Virginia using imagery acquired by SPOT. SPOT data acquired before and during defoliation was used to classify insect defoliation. Based on ground truth data and aerial photography, the range of ratio values corresponding to heavy, moderate, and light defoliation were defined. Clerke and Dull (1990), however, raised questions regarding the completeness of this classification, citing the unknown effects of terrain and forest type on the extent and severity of gypsy moth defoliation.

Dull *et al.* (1990) used SPOT imagery, high altitude panoramic color infrared photography, and traditional aerial sketch-mapping results to determine the extent of gypsy moth defoliation in northern Virginia. This study illustrated the importance of maintaining a GIS database to track defoliation extents, spray block extents, pheromone trap data, and egg mass survey results. This database could be used to efficiently determine the defoliated area of each county, the defoliated area of each property owner, and the defoliated area of each spray block. This information could make the evaluation of treatment success, as well as any treatment decisions very simple. Joria and Ahearn (1991) used a digitized USGS map to determine the locations of forested areas in Michigan and concentrated the study on only those forested areas using both Landsat TM and SPOT imageries. Landsat TM was found to be better than SPOT for differentiating the defoliation classes.

Data from SPOT Vegetation (VGT) at 1 km resolution was used for mapping defoliation and mortality of coniferous forests due to the eastern hemlock looper, with commission errors of 60% and omission errors of 33% respectively, and with reduced errors when aggregating the data into larger mapping units (Fraser and Latifovic, 2005). The authors also indicated the potential for near real-time monitoring, however with potentially greater errors. Fraser and Latifovic (2005) suggested the combination of SPOT VGT and NOAA AVHRR data for establishing a general system for large-scale (5–10 km^2) forest change detection. While the general occurrence of defoliated areas can be identified, the classification of the intensity of defoliation has been less reliable using SPOT and NOAA AVHRR (de Beurs and Townsend, 2008).

Insect defoliation outbreaks have also been investigated using Moderate Resolution Imaging Spectro-radiometer (MODIS) data (Kharuk *et al.*, 2007). de Beurs and Townsend (2008) conducted a thorough analysis of MODIS daily, 8-day and 16-day composite data for detecting gypsy moth defoliation in oak forests. Their study demonstrated significant relationships between defoliation and vegetation indices estimated at the plot scale. They concluded that

MODIS data represent an important tool for insect damage detection at the regional scale. Furthermore, Cook *et al.* (2008) studied the effect of insect defoliation on forest production efficiency and net carbon exchange using models driven with MODIS data. In contrast to other multi spectral remote sensing platforms, MODIS data have a lower spatial resolution, and are therefore more appropriate for regional-scale analyses. In addition, MODIS data are available at a significantly higher temporal resolution (daily) while preserving the spectral bands that are available in the Landsat data.

However, a common problem in using MODIS data is that evaluation of coarse-resolution damage maps is difficult due to the general lack of spatially explicit reference data. Many of the cited studies have evaluated their classifications against sketch maps from aerial surveys or Landsat change maps. These are themselves estimates that may be limited in temporal and attribute accuracy.

Relatively new Medium Resolution Imaging Spectrometer (MERIS) data according to Van der Sanden *et al.* (2006) proved to be better in detecting and monitoring insect defoliation than Landsat, MODIS and SPOT because of its ability to image large areas at medium spatial resolution. MERIS data were concluded to generally depict areas of tent caterpillar defoliation in Canadian aspen forests (Van der Sanden *et al.*, 2006), however, no formal evaluation was made due to lack of accurate ground data. Since early 2009, the German RapidEye satellite represents the first high-resolution multispectral satellite operationally providing the red edge spectrum, followed by satellites such as WorldView-2 in 2010 and Sentinel-2 in 2014.

Currently, RapidEye and WorldView-2 are the only commercial multispectral satellite to provide global, high-resolution access to the Red-Edge spectral band as part of its 8-band multispectral capabilities (Cheng and Chaapel, 2008). This Red Edge band has been used to track stress-induced changes in plants, hence it is very important band to consider when detecting and monitoring the health of forest. Until now, the only satellite imagery available that contains Red-Edge data is MERIS with medium spatial resolution (300 m) (Van der Sanden *et al.*, 2006). MERIS can provide some insights into the conditions of an entire field, but is unable to provide the segmentation necessary to evaluate small scale details, like the health of individual trees in an orchard, hence the advantage of RapidEye and World View-2 with higher spatial resolution.

Although images from SPOT, MODIS and MERIS proved a better alternative to Landsat in terms of the spatial, temporal and spectral resolution, they are also limited in that they do not contain specific windows such as red–edge (except for MERIS which still has low spatial resolution) which is a very important band in studying defoliation (Pu *et al.*, 2003). Moreover, most of the researches on insect defoliation using multispectral images have been carried out in coniferous forest. Therefore, there is need to test the effectiveness of these multispectral sensors for monitoring insect defoliation in broadleaved forests.

2.5.2 Hyperspectral Remote Sensing

Further advances in satellite remote sensing and imaging spectrometry have given rise to hyperspectral imagery, which has been demonstrated to be a reliable and relatively accessible technology to study forests damaged by insects (Coops *et al.*, 2004; Mutanga *et al.*, 2009; Stone and Coops, 2004). Hyperspectral sensors, also known as imaging spectrometers are instruments specifically made to acquire data at high spectral and moderate spatial resolution, thereby allowing reflectance, radiance and emittance spectra to be constructed in such a way that it permits physical measurements of the Earth's surface. Unlike multispectral imagery, a hyperspectral image provides hundreds of contiguous bands across the visible (VIS), near-infrared (NIR) and shortwave infrared (SWIR) regions of the electromagnetic spectrum, offering unprecedented detailed spectral reflectance data from land surface features. Since major leaf components (e.g. pigments, water, carbon, nitrogen) produce distinctive reflectance signals at specific wavelengths of the aforementioned regions, hyperspectral imagery allows for the measurement of biochemical and biophysical attributes of the plant associated with its structure, physiology and phenology, and therefore with its health status (Asner, 1998; Cho and Skidmore, 2006; Lucas *et al.*, 2004; Mutanga and Skidmore, 2004a; Mutanga *et al.*, 2004; Treitz and Howarth, 1999).

There is mounting evidence that hyperspectral instruments have the capability, not only to assess defoliation, but also to identify the early signs of defoliation; in some cases before visual symptoms are apparent (Ismail *et al.*, 2007; Ismail *et al.*, 2008; Mohammed *et al.*, 1995). This can be explained by the tendency of defoliated leaves to undergo reduction in photosynthetic activity and to lose chlorophyll. These changes alter reflectance at chlorophyll-sensitive wavelengths (Vogelmann *et al.*, 1993). Researches on defoliation conducted using hyperspectral sensors are not limited in the literature (**Table 2.2**). Although broad-band sensors detect defoliated and non-defoliated plants, hyperspectral imagers, have the spectral detail to potentially

distinguish between soil, dead, and senescent trees. Hyperspectral data are demonstrated to discriminate plant physiological condition (Pontius *et al.*, 2005), even at early phases of senescence (Campbell *et al.*, 2004).

Some previous studies have used hyperspectral remote sensing to detect plant defoliation due to water deficit (Stimson *et al.*, 2005), insect damage (Radeloff *et al.*, 1999), pest outbreaks (Wolter *et al.*, 2008), and pollution (Campbell *et al.*, 2004). For instance, Pontius *et al.* (2005) used AISA Eagle sensor to map hemlock decline in USA. They found that unlike multispectral sensors, hyperspectral sensors were able to classify defoliation in a given forest to 11-class rating system with 88 percent accuracy making it possible for land managers to assess and monitor detailed changes in forest health. Others, such as Somers *et al.* (2010) used Hyperion sensor to monitor forage defoliation in southern Australia and observed a good relationship with ground measurement. Additionally, the high spatial resolution data from hyperspectral sensors gives the ability to detect tree-level (rather than stand level) characteristics, which reduce confusion caused by mixed pixels (e.g. crown shading, soil, non-tree vegetation, (Greenberg *et al.*, 2006). While some of these studies reveal substantial improvement over multispectral sensors, others believe that for accurate detection and monitoring, there will be need for the combination of both sensors.

However, just like multispectral sensors, only a few studies have used hyperspectral sensors to monitor defoliation in broadleaved forests (Coops *et al.*, 2003; Santos *et al.*, 2010). To the best of our knowledge, no studies have been carried out to map and monitor mopane woodland defoliation using hyperspectral scanners. Hyperspectral sensors could actually provide more information of insect defoliation in broadleaved forest because of the presence of high spatial and different bands to which defoliation can be monitored. Therefore, further research need to be conducted on the use of Hyperspectral remote sensors for effective management of insect defoliation in broadleaved forests such as mopane woodland.

Table 2. 2: Sample of Hyperspectral Remote Sensing Studies Applied in Defoliation

Sensor	Study Area	Defoliators	Comment	Reference
AISA EAGLE	USA	Hemlock Looper	Good for identifying defoliation at tree level using biophysical indicators such as chlorophyll content	Pontius *et al.* (2005)
HYPERION	Australia	Beetles and Sawfly	Good results with ground measurement	Somers *et al.* (2010)
HYPERION	Chile	Aphid	Able to detect defoliation also at tree level	Peña and Altmann (2009)

2.6 Trade-Offs Between Sensor Resolutions for Monitoring Insect Defoliation

We have shown that remote sensing is effective for mapping insect defoliation. However, three issues appear fundamental to the successful use of remote sensing to assess and monitor insect defoliation: the spectral, spatial characterization of defoliation and the timing of image acquisition.

First, a remote sensing spectral basis for damage class limits (e.g., light, moderate, and severe) is required to achieve consistent detection and mapping of defoliation severity. Field and aerial surveys tend to rate areas defoliated into categories that remote sensing studies have attempted to emulate. Broad damage class limits are not conducive for consistent defoliation mapping because they may not correspond to differences in spectral response values that are spectrally or statistically separable on the image. The two factors that drive the spectral response of a sensor include its radiometric resolution and the range of sensitivity to the electromagnetic spectrum. Defoliation tends to result in either physical loss of leaf area or leaf color change, which results in physical differences in spectral response when compared to pre-defoliation images. Several consecutive years of defoliation, however, tend to result in physiological weakening, top kill, and mortality for some defoliators. Understanding the role these factors may play in the resulting spectral responses recorded in the image is important to successful use of remote sensing for mapping defoliation. Thus, remote sensing observations from airborne or satellite sensors that can be used for monitoring defoliation must be over a more continuous scale of spectral responses that can potentially capture a finer scale of defoliation levels rather than the broad classes that are typically used (Franklin, 2001; Hall *et al.*, 2003a) hence the recent use of hyperspectral sensors.

In addition to the spectral observations of defoliation, the size of the outbreak area must also be large enough to be detectable with the airborne or satellite sensor employed. The spatial

resolution of the sensor and the areal coverage of an image are also important considerations in the selection of the appropriate sensor. As a result, with both sensor spectral and spatial resolution considerations, the remote sensing of a defoliation problem is more complex than a simple change in foliage condition.

Thirdly, the timing of image data acquisition should coincide with the period when spectral changes resulting from defoliation are most observable; for timing of data acquisition is notably one of the most difficult to achieve with satellite remote sensing because of the need for cloud-free conditions during the suitable range of dates for image acquisition. Most remote sensing studies tend to rely on pre- and post-outbreak images to detect spectral response differences resulting from insect defoliation. The opportunities to acquire imagery ranging from high (e.g., submeter pixel size) to low spatial resolution (e.g., 1-km pixel size) are obviously increasing at an unprecedented rate that should help ensure that future image data will be available during the narrow time periods necessary to capture damage from insect defoliation.

This section has outlined the remote sensing data used in monitoring defoliation from inception highlighting the strengths and weaknesses and its implications in mapping and detecting defoliation within mopane woodland. Logical questions that follow include: What has been the primary methods used in defoliation surveys and which remote sensing methods have been employed in mapping defoliation and to what extent have they been successful?

2.7 Developments in Remote Sensing Techniques for Assessing Insect Defoliation

A number of studies have demonstrated the potential of measuring defoliation from remotely sensed observations using different techniques (Coops *et al.*, 2004; Stone and Coops, 2004). The studies use map sketching and linear regression modeling between field assessments of vegetation characteristics related to biophysical variable indicators of defoliation such as LAI and Chl content. They also relate the field measurements with vegetation indices calculated from the images to detect and monitor defoliation from a range of damaging agents including fungal infections and insect predation. Therefore, we discuss the different analytical approaches that have been used for detecting, mapping and monitoring insect defoliation and their implications in mopane defoliations monitoring.

2.7.1 Vegetation Indices

Most of the vegetation indices developed to detect defoliation in woodlands are based on Chl and water content. Vegetation indices can be used to measure changes in leaf area resulting from

defoliation (Nelson, 1983). Previous studies have used vegetation indices or other measures to examine canopy defoliation by a variety of insects. Nelson (1983) calculated the difference between vegetation indices on two Landsat dates using simple ratio indices such as red-green ratio index, and then empirically determined a threshold to separate defoliated from non-defoliated pixels. This technique was found to be superior to competing techniques for the most accurate assessment of defoliated areas. It has been reported that plants under defoliation display a decrease in canopy reflectance in the lower portion of the near infrared, a reduced absorption in the Chl active region, and subsequently a shift in the red edge (Carter and Knapp, 2001).

One simple vegetation index that has also been used in the past is the Normalized Difference Vegetation Index (NDVI) (Rouse *et al.*, 1973). As defoliation occurs and leaf area decreases, the NDVI value will also decrease. It has recently been shown that the Wide Dynamic Range Vegetation Index (WDRVI) performs better than the NDVI in estimating defoliation in high-density vegetation (Gitelson *et al.*, 2006; Gitelson *et al.*, 2002). While the NDVI becomes saturated with high densities of photosynthetic green biomass and the relationship between NDVI and LAI is non-linear (Mutanga and Skidmore, 2004a), the WDRVI increases the sensitivity of the NDVI, and hence makes the WDRVI - LAI relationship linear. More complex vegetation indices correct for variations in soil background and for atmospheric scattering. The Enhanced Vegetation Index (EVI) (Huete *et al.*, 2002) is the standard vegetation index for MODIS. EVI will decrease in response to defoliation (Huete *et al.*, 2002).

However, due to errors (saturation in areas with significant forest cover) encountered when using these indices, scientist have developed better indices which includes the short wave infrared band (SWIR). Water strongly absorbs radiation in SWIR portion of electromagnetic spectrum making SWIR reflectance to be very sensitive to the amount of water in vegetation. SWIR reflectance is generally low for high leaf water content and increases with decrease in water content. The sensitivity of the SWIR has led to the development of a number of vegetation indices that are responsive to vegetation defoliation based on SWIR and NIR reflectance.

Normalized Differential Water Index (NDWI) and Normalized Differential Infrared Index (NDII) (Gao 1996) were developed from hyperspectral data as the difference between NIR reflectance and a SWIR band and were found to be good detectors of defoliation. In most cases however, most of these indices have been modified to meet the need of researches into insect defoliation in modern times. All of these indices have been used in mapping and detecting

32

defoliation of different insect defoliator with varying outcomes (Coops *et al.*, 2004; Pu *et al.*, 2003). While some have reported success, some of the indices have failed to efficiently discriminate defoliated areas from other spectrally active end members within particular woodlands. A common problem in dense vegetation stands is the high degree of light absorption making vegetation indices insensitive to biomass changes. Knowing fully the limitation of vegetation indices, scientists have developed and improved techniques that can accurately estimate biomass in more densely vegetated areas using hyperspectral derived vegetation indices rather than focusing on the red and NIR bands alone (Mutanga and Skidmore, 2004a). For detecting and mapping defoliation in mopane, we suggest the use of three categories of vegetation indices which are described below.

2.7.1.1 Multiple Ratio Indices

Multiple ratio indices such as Normalized Differential Vegetation Index (NDVI), modified normalized difference vegetation index (mNDVI) have been found to be good detectors of defoliation (Sims and Gamon, 2002). While, NDVI is a vegetation index derived from the ratio of red and NIR bands and has been found to be highly correlated with biophysical indicators that depicts defoliation (LAI and Chl content) (Dye and Tucker, 2003; Zhou *et al.*, 2003), mNDVI modifies NDVI by including the reflectance at high 445nm (at which Chl absorption produce minimal reflectance) (Sims and Gamon, 2002). The mNDVI compensates for high leaf surface scattering that NDVI does not account for (Peña and Altmann, 2009). It is our opinion that mopane worm defoliation of mopane woodland is expected to yield a decrease in NDVI, mNDVI during the defoliated stage when the leaves are almost absent. The limitation of these indices in the context of detecting and mapping defoliation of mopane woodland will be their sensitivity to optical properties of reflecting soil background since for a given amount of vegetation, soil substrates results in higher vegetation index values which may not necessarily mean lack of defoliation (Sims and Gamon, 2002). To minimize the effect of soil background, other vegetation indices have been proposed.

2.7.1.2 Simple Ratio Indices

Simple ratio indices were developed to reduce or eliminate soil influence on solar reflectance values when monitoring forest health (Huete *et al.*, 2002; Pen˜uelas *et al.*, 1995). The simple ratio indices measured with sufficient precision is quite sensitive to vegetation changes during the time of peak growth. However, an inherent drawback of these indices is the loss of

uniqueness in information due to the fact that different leaves can have different spectral responses, but have band ratio values that are similar. The introduction of multiple ratios such as NDVI has covered the limitations of simple ratio although they both saturate when LAI is very high. For monitoring mopane defoliation, two simple ratios have been suggested, the Red Green Ratio Index (RGRI) and Red NIR Index (RNI). The two indices have been found to accurately determine the defoliation level of forest to a certain level since they include the combination of bands where healthy and unhealthy vegetation can be easily differentiated (Pen˜uelas *et al.*, 1995). They also eliminate topographic (irradiance) and atmospheric effects (Pen˜uelas *et al.*, 1995). For mopane defoliation, the ratio is expected to be high when the woodland is photosynthetically active i.e., the healthy stage and *vice versa*.

2.7.1.3 Red Edge Indices

Recently, new vegetation indices based on red-edge region have been used to track insect defoliation (Van der Sanden *et al.*, 2006). Red-edge is defined as the rise of reflectance at the boundary between the chlorophyll absorption feature in VIS red wavelengths and leaf internal structure scattering in NIR wavelengths. The position of the red edge is consistent among different species and generally ranges from 680 to 780 nm (Cho and Skidmore, 2006). Red edge indices are constructed with bands sensitive to the Chl content and internal structure of the leaf, and therefore have proven to be closely related to foliage biomass quantity, growth and developmental stage and health status of the plant (Gitelson and Merzlyak, 1994; Sims and Gamon, 2002). Zarco-Tejada *et al.* (2000) describe Chl content as a potential indicator of defoliation process because of its direct role in the photosynthetic processes of light harvesting and initiation of electron transport and its responsiveness to a range of changes in plant health status at any particular time.

Red edge indices such as Curvature index, Vogelman index, NDVI-RE and the red edge inflection points have been used to relate biophysical indicators that are used to measure the health status of woodland (Zarco-Tejada *et al.*, 2000). Vogelman red edge index was discovered to be associated with leaf area and chlorophyll content while curvature index was used to track changes in Chl content and it was found to be strongly correlated with Chl index (Zarco-Tejada *et al.*, 2000). NDVI-RE have recently been successfully tested in relations to changes canopy cover (Souza *et al.*, 2010). The red edge inflection point (REIP) has also been found to correlate significantly with LAI and hence could be used for monitoring the health status of woodland.

In monitoring mopane worm defoliation of mopane woodland therefore, it is expected that the changes in the values of the red edge indices mentioned above will indicate the changes in the health status of the canopy at the different stages of defoliation. For instance, decrease in the values of the red edge indices is expected during the refoliating stage when the leaves are coming up after insect defoliation.

2.7.1.4 Pigment Content Indices

When the rate of photosynthesis decreases due to plant stress, the foliage exhibits higher concentrations of carotenoid relative to chlorophyll pigments, while higher foliar investments of xanthophyll cycle pigments result as a response to low light use efficiency. Vegetation indices based on bands sensitive to these leaf pigments have also been demonstrated to be closely correlated with vegetative growth stage and the degree of stress of vegetation (Gamon and Surfus, 1999; Gitelson *et al.*, 2002; Sims and Gamon, 2002). In stressed plants, the proportionally stronger decline of green pigments (i.e. Chls) can be used to detect defoliation. The two major pigment indices that have been found to be an indicator of plant stress are photochemical reflectance index (PRI) and Carotenoid Reflectance Index (CRI) (Gitelson *et al.*, 2002). The PRI and CRI were developed as a remotely-sensed indicator of Light Use Efficiency (LUE) (Gamon *et al.*, 1992). They use narrow spectral bands to detect changes in leaf reflectance at 531 nm relative to a reference band that is usually located at around 570 nm and is not affected by changes in short-term stress events.

Carotenoid pigments have multiple functions, but they are generally found in higher concentrations in plant leaves that are either stressed or dead. PRI and CRI have been correlated with plant stress in several field studies at the leaf and ecosystem levels (Pen˜uelas *et al.*, 1995). They provide a quick and non-destructive assessment of leaf physiological properties (Pen˜uelas *et al.*, 1995) and may be used for wide range of species (Gamon *et al.*, 1992). The limitation of these indices for defoliation mapping occurs when they are related to plant water status (Pen˜uelas *et al.*, 1995) especially during wilting of leaves in dry periods. It must be noted that PRI is sensitive to soil background reflectance (Pen˜uelas *et al.*, 1995). However, they may be integrated with other vegetation indices that exploit biophysical variables to provide strategic remote sensing monitoring of defoliation. In mapping defoliation within mopane woodland, it is believed that defoliation will result in low green pigments hence low values of PRI and CRI based on leaf pigments during the peak of defoliation since photosynthesis activities is low or

even nonexistent. Narrow-band spectral reflectance may also provide information on the ratio of carotenoid to Chl for detecting stress effects.

2.7.2 Change Detection

Other studies have approached defoliation in terms of change detection methods (Collins and Woodcock, 1996). Relatively robust change detection methods include image differencing and ratio differencing. Image differencing is probably the most widely applied change detection algorithm for a variety of geographical environments (Coppin *et al.*, 2004). It involves subtracting one date of imagery from a second date that has been precisely registered to the first. With "perfect" data, this would result in a data set in which positive and negative values represent areas of change and zero values represent no change. Nelson (1983) delineated forest canopy changes due to Gypsy Moth defoliation in Pennsylvania more accurately with vegetation index differencing than with any other single band difference or band rationing. Image ratio differencing on the other hand is one of the simplest and quickest change detection methods in insect defoliation monitoring where data are rationalized on a pixel-by-pixel basis. A pixel that has not changed will yield a ratio value of one. Areas of change will have values either higher or lower than one. The major drawback for these two change detection algorithms is that they do not adequately address differences in sun elevation angles or phenological changes between images recorded at different dates (Radeloff *et al.*, 1999). Infact Riordan (1981) criticized the ratio change detection algorithm in combination with an empirical threshold definition as being statistically invalid.

Image transformation techniques are frequently applied to multidate imagery that has been stacked in 2n-dimensional space (where n is the number of input bands per image): principal component analysis (PCA) and tasseled cap (Radeloff *et al.*, 1999). Using multi-date Landsat TM data, Collins and Woodcock (1996) compared Kauth-Thomas and Principal Components transforms with Gramm Schmidt orthogonalization for mapping pest-induced forest mortality in the Lake Tahoe Basin, concluding that the KT transform was most sensitive to changes in vegetation condition. Muchoney and Haack (1994) also examined merged principal components analysis, image differencing, spectral-temporal change classification, and post classification differencing for detecting forest defoliation. Their results indicated that of the entire algorithm employed, defoliation was best determined by image differencing and principal components analysis. The exact nature of the principal components derived from multi-temporal data sets is difficult to ascertain without a thorough examination of the structure of the data and a visual

inspection of the combined images. To avoid drawing faulty conclusions, the analysis should not be applied as a change detection method without a thorough understanding of the study area (Diago *et al.*, 2010). Moreover, the vegetation indices and most of the transformation methods used for monitoring defoliation are however limited by their dependence on the visibility of leaves in image pixels since for defoliation, it is the absence of leaves that determines the severity of the stress (Stone and Coops, 2004).

By using combined registered data sets, or corresponding subsets of bands, collected under similar conditions, researchers have come up with another algorithm for monitoring change detection in forest health known as composite analysis (Coppin *et al.*, 2004). They came up with classes where forest canopy change would be expected to have statistics significantly different from those where no change has occurred, and could be identified as such. The method can incorporate multistage decision logic and is sometimes referred to as "layered spectral / temporal change classification", "multidate clustering", or "spectral change pattern analysis". While this technique necessitates only a single classification, it is a very complex one, in part because of the added dimensionality of two dates of data.

In numerous cases it requires many classes and many often redundant features when no discriminant analysis has preceded the process. It furthermore demands prior knowledge of the logical interrelationships of the classes and should only be used when the researcher is intimately familiar with the study area (Coppin *et al.*, 2004). Burns and Joyce (1981) found the method to produce only change in forest cover per se without providing accurate information on the character of the change. Coppin *et al.* (2004) remarked that, since spectral and temporal features have equal status in the combined data set, they cannot be easily separated in the pattern recognition process. As a consequence, class labeling using this algorithm may be difficult.

A mathematical model that best describes the fit between two multidate images of the same area can be developed through stepwise regression and also use to detect defoliation in forest. The algorithm assumes that a pixel at time is linearly related to the same pixel at a later time in all bands of the electromagnetic spectrum acquired by the sensor. This implies that the spectral properties of a large majority of the pixels have not changed significantly during the time interval (Coppin *et al.*, 2004). The dimension of the residuals is an indicator of where change occurred. The regression technique accounts for differences in mean and variance between pixel values for different dates. Simultaneously, the adverse effects from divergences in atmospheric

conditions and/or sun angles are reduced. The critical part of the method is the definition of threshold values or limiting dimensions for the no-change pixel residuals.Singh (1989), on the other hand, reported the highest change detection accuracy for tropical forest change detection with the regression method.

2.7.3 Statistical Classifiers

Change Vector Analysis (CVA) is a multivariate statistical analysis that has been used extensively to identify changes in forest as a result of insect defoliation between image dates and is widely discussed in the remote sensing literature (Collins and Woodcock, 1996; Townsend *et al.*, 2004). Generally, users rely upon two outputs of CVA to represent the magnitude of change, computed as the absolute geometric difference in the soil brightness (B), vegetation greenness (G), and surface wetness (W, collectively BGW), between dates, and an eight-level classification representing all possible directions of change bounded by BGW all increasing between dates and BGW all decreasing between dates (Allen and Kupfer, 2000). As noted in numerous studies (Allen and Kupfer, 2000; Cohen and Fiorella, 1998; Townsend *et al.*, 2004), the limitation to this approach for mapping defoliation is that it requires the user to identify a threshold level in magnitude change that represents actual change between dates rather than changes within a date. This becomes very difficult to use when defoliation is rapid and can only be applied with imageries of high temporal resolution.

Radeloff *et al.* (1999) developed an approach for monitoring defoliation in terms of the relative proportion of leaves in image pixels using linear spectral mixture analysis (SMA) which can quantify the proportion of each pixel that is occupied by individual image component (Sims *et al.* 2007). These methods have previously been used to measure defoliation in *Pinus radiata* plantations using high-resolution multi-spectral images but methods for calculating image fractions from hyperspectral image data are in their developmental infancy. The application of SMA for the assessment of defoliation offers several advantages over simple regression methods using spectral indices and other transformation methods in that it is capable of detecting vegetation cover at low and fragmented levels, and has the ability to reference a small number of spectrally stable endmembers (vegetation, soil, water, etc.) (Goodwin *et al.*, 2005). The technique decomposes the reflectance of each pixel into the relative contribution of a limited number of surface endmembers making it easy to separate image components (Somers *et al.*, 2010). SMA methods have been used to monitor insect defoliation in broadleaved forest (Goodwin *et al.*, 2005; Somers *et al.*, 2010). However and to date, the full potential of SMA for

forest defoliation assessment has not yet been achieved. Residual error in fraction estimates provided by SMA is often introduced by the natural variability in the conditions of scene components, i.e., soil, plant, etc. inherent in remote sensing data.

Similarly, researchers have used other statistical machine learning algorithm for defoliation monitoring (Krahwinkler and Rossman, 2011; Melgani and Bruzzone, 2004; Nitze *et al.*, 2012). Two of the most common ones are support vector machine (SVM) and random forest (RF) algorithms. Random Forest is a machine learning algorithm that employs a bagging (bootstrap aggregation) operation where a number of trees (*ntree*) are constructed based on a random subset of samples derived from the training data (Breiman, 2001). The trees are independent of each other grown to maximum size based on a bootstrap sample from the training data set without any pruning, and each node is split using the best among a subset of input variables (*mtry*) (Breiman, 2001). The algorithm classifies the data that are not in the trees (out-of-bag or OOB data) and by averaging the OOB error rates from all trees, the random forest algorithm gives an error rate called the OOB classification error for each input variable (Breiman, 2001). On the other hand, SVM is a relatively new non-parametric algorithm for image classification. In its classical implementation, it uses two classes (e.g. presence/absence) of training samples within a multi-dimensional feature space to fit an optimal separating hyperplane (in each dimension, vector component is image gray-level) (Foody and Mathur, 2006). SVM consists of projecting vectors into a high dimensional feature space by means of a kernel trick, then fitting the optimal hyperplane that separates classes using an optimization function. Several kernels are used in the literature (Foody and Mathur, 2006). Previous studies have used both RF and SVM for classification with great success (Adam and Mutanga, 2010; Krahwinkler and Rossman, 2011; Melgani and Bruzzone, 2004; Menze *et al.*, 2009; Nitze *et al.*, 2012; Petropoulos *et al.*, 2012; Rodriguez-Galiano *et al.*, 2012; Statnikov *et al.*, 2008; Waske *et al.*, 2007), however and to the best of our knowledge no study have used the two classifiers for insect defoliation levels classification. Both RF and SVM algorithm are easy to implement as only two parameters (*ntree* and *mtry* for RF) and (*C* and λ for SVM) are needed for optimization (Breiman, 2001; Özçift, 2011).

2.8 Potential of Mapping Mopane Defoliation Using Remote Sensing
The current method used to spatially map the mopane worm defoliation of mopane woodland is by field-based exercises. The effectiveness of this method is questionable because the method is

qualitative, subjective, and dependent on the skill of the surveyor (Coops *et al.*, 2004). However, the ability of remote sensing to successfully detect and map forest health has been given great attention with diverse range of imageries and modeling techniques (Coops *et al.*, 2004; Pontius *et al.*, 2005; Radeloff *et al.*, 1999). Thus remote sensing has the potential to ensure that the detection, mapping and monitoring of mopane worm defoliation is a possible task provided a sound understanding of the progression and patterns of defoliation are known. Knowledge of these defoliation processes allows for the development of algorithms to detect changes in foliar characteristics using remotely sensed data. Digital remote sensing technologies measure the amount of electromagnetic energy reflected from the leaves and canopy of the tree using a number of wavelengths which can range from 350 to 2500 nm. Researchers have used this spectral information, in the form of individual bands, band combinations, and vegetation indices to detect and map forest health (Coops *et al.*, 2004; Pontius *et al.*, 2005). Additionally, remote sensing technology can image large areas and allow for the repetitive monitoring and assessment of tree damage and mortality (Van der Sanden *et al.*, 2006).

A combination of both multispectral and hyperspectral imageries will give more insight for detecting and monitoring the defoliation process within mopane woodland in order to determine the best spatial and spectral resolution to which it can be monitored. In mapping and monitoring mopane defoliation using satellite remote sensing platforms therefore, spatial, spectral and temporal resolution must be of great importance. We therefore suggest the use of Word View-2 and RapidEye imagery for monitoring the defoliation process. The primary strengths of RapidEye and Word View-2 are their high temporal resolution (1.1 days), the presence of windows (red edge) for monitoring defoliation and their ability to image large areas at a relative high spatial resolution (1.84m). The view of these authors is that with the relatively high resolution images of World View-2 and RapidEye, we can not only be able to map but also determine the extent to which spatial, spectral and temporal resolution of satellite imageries play in the detection and monitoring of defoliation from mopane woodland at different stages i.e. undefoliated, partly defoliated and refoliating stages.

Given the success and limitations of the different techniques used in monitoring insect defoliation discussed above, we recommend the combination of different techniques with the aim of determining the best approach to mapping and monitoring defoliation within mopane woodland. For instance the use of vegetation indices will provide the spatio-temporal patterns of

mopane defoliation while more sophisticated image classification techniques such as SVM and RF which have been previously associated with biophysical indicators that are related to defoliation (Radeloff *et al.*, 1999; Somers *et al.*, 2010) will reduce the limitations encountered with the use of vegetation indices.

2.9 Challenges of Remote Sensing In Mapping Mopane Defoliation

Despite all the efforts of applying remote sensing, insect defoliation monitoring has been only moderately successful. Reliable insect defoliation monitoring has often been limited to three classes (e.g., heavy, medium, and light) with accuracies around 70–80%. Low defoliation levels remain difficult to detect. Consequently, the challenge would be to assess the different characteristics and defoliation process within mopane woodland, and then associate each level of observation with different remote sensing data types in order to provide the appropriate level of detail and accuracy for detection and mapping purposes.

Three challenges may make it difficult to monitor mopane defoliation with remote sensing. First, mopane worm-mopane tree interactions are dynamic and periods where defoliation can be detected are often short. For instance, mopane worm defoliation of mopane woodland are bivoltine across most of its distribution with the first defoliation in November to December and the second in February to March, except in more arid areas where it is univoltine. This restricts the time period when defoliation can be detected to about 2 months. However, during the early stages of defoliation (when the worms are feeding on the tree), the canopy of the tree appears green and visually indistinguishable from healthy trees (Ciesla, 2003). Leaves become pale green gradually leading to total defoliation. At this stage, discriminating the defoliated part of the tree using remote sensing is dependent on detecting the little changes that might have occurred in the spectral reflectance of the tree. The subtle changes in the reflectance of defoliated vegetation, when measured by various broad band sensors, are often masked by the high degree of variation in reflectance caused by factors such as varying view geometry, illumination, and canopy density (Lucas *et al.*, 2004). Moreover, because mopane worm is a wasteful feeder, its feeds on the entire leaflets of mopane tree giving the trees their characteristic red brown color (Ditlhogo *et al.*, 1996).

Given these challenges, there are strong possibilities of using high spectral resolution data (hyperspectral) for effective detection, mapping and monitoring early stages of mopane defoliation because the data allow for the detection of detailed features using many narrow bands

which would have been otherwise masked by broad band sensors (Kumar *et al.*, 2001; Mutanga and Skidmore, 2004a; Mutanga *et al.*, 2009).

The second challenge is the presence of other endmembers in mopane woodland. Most mopane woodland contain other tree species that are always present during the mopane worm defoliation making it very difficult to distinguish defoliation of mopane from reflectance of other end members (Vogelmann *et al.*, 1993). Infact, reflectance of heavily defoliated mopane tree may be mistaken for the end members. To resolve this, two things are suggested. Firstly, it will be important to use hyperspectral dataset to characterize the reflectance of end members in order to distinguish them from actual defoliation (Somers *et al.*, 2010). This can then be applied to satellite images. Secondly, in situation where hyperspectral data are not available high resolution multispectral with strategic bands will be accurate for detection and mapping purpose at canopy level and able to distinguish different understoreys using advance classification algorithm such as RF and SVM.

The third challenge may occur when an image at the peak of an outbreak is being analyzed; it is unclear if an effect (changes in chlorosis, nutrient content or tree vitality) or a determining factor of the insect population drives the satellite image classification (Somers *et al.*, 2010). Effects and determining factors can both lead to reasonable classification accuracy peak-outbreak satellite imageries are analyzed. Effects and determining factors may not be important for a forest manager mainly interested in a quick assessment of insect outbreak. However separating the two and being able to identify actual defoliation is crucial for a scientist who may want to study the relationship between effects and defoliation. To achieve this, there may be need to relate the effects with the determining factors. For example, the reflectance of levels of defoliation from image classification algorithms can be correlated with biochemical and biophysical variables at sampling location to detect if there are any relationships with effects and determining factor of mopane defoliation.

2.10 Synthesis and Recommendations for Remote Sensing of Mopane Defoliation

The capabilities of the various sensor systems presently available to the forest health protection specialist are a key factor in the type of sensor selected for a specific application. Aerial sketch mapping, for example, is an excellent tool for background mapping, but subjective and their reliability is difficult to assess hence not suitable for the assessments of forest health. Today's Earth-orbiting satellites and airborne sensors offer the advantage of image acquisition at

regular intervals provided that the targets of interest are not under cloud cover. They also provide a range of spectral sensitivity across the electromagnetic spectrum. Initial satellite sensors have a poor spatial resolution, however, when compared to airborne sensors. This limits their ability to resolve all but the most severe of forest damage signatures. However, remote sensing is a dynamic technology. New and improved methods of data collection, with superior resolution, are continuously becoming available. For instance, high spatial resolution remote sensing for forestry applications has reached an almost mature phase with wide range of applications. Numerous opportunities and challenges such as the robustness of remote sensing data processing and analytical methods remain.

With increasing availability of high resolution remote sensing data like hyperspectral scanners and Digital Multispectral Image (DMSI) in Southern African sub-region which offers a potential source for the effective collection of spatially accurate, consistent, and timely imagery, it is essential to study the impacts of *Imbrasia belina* on mopane woodland. High resolution remote sensing data is capable of achieving higher mapping accuracies by identifying individual crowns (Wulder and Franklin, 2003). This is an important benefit for mapping and monitoring *Imbrasia belina* in mopane woodland as it helps to dissociate insect defoliation from other events, such as climate disturbance and phenology of forest type (Stone and Coops, 2004). Given this development, it is prudent to assess what remotely sensed methods or data sources may have potential for detecting, mapping and monitoring defoliation in mopane woodland. Since defoliation of mopane woodland by mopane worms occurred in the past and will likely re-occur in the future, lessons learned from this research may be applied in future mopane woodland management. As such, a lot of research is still to be done to fully understand the potential of high spatial and spectral resolution data in insect defoliation especially in broadleaf forest defoliation such as mopane woodland.

CHAPTER THREE

Species Discrimination

This chapter is based on:

S. **Adelabu**, O. Mutanga, E. Adam and M. A. Cho, "Exploiting Machine Learning Algorithms for Tree Species Classification in a Semi-Arid woodland Using RapidEye image," *Journal of Applied Remote Sensing*, 7 (1), 073480, 2013.

ABSTRACT

Classification of different tree species in a semi-arid area can be challenging as a result of the change in leaf structure and orientation due to soil moisture constraints. Tree species mapping is however a key parameter for forest management in semi-arid environment. In this study we examined the suitability of RapidEye satellite data for the classification of 5 tree species in mopane woodland of Botswana using machine leaning algorithms with limited training samples.We performed classification using Random Forest (RF) and Support Vector Machines (SVM) based on EnMap box. The overall accuracies for classifying the 5 tree species classes was 88.75% and 85% for both SVM and RF respectively. We also demonstrated that the new red-edge band in the RapidEye sensor has potential for classifying tree species in semi-arid environment when integrated with other standard bands. Similarly, we observed that where there are limited training samples, SVM prefers over RF. Finally, we demonstrated that the two accuracy measures of quantity and allocation disagreement are simpler and more helpful for the vast majority of remote sensing classification process than the Kappa coefficient. Overall, high species classification can be achieved using strategically located RapidEye bands, integrated with advanced processing algorithms.

Keywords: Random Forest, Support Vector Machines, Tree Species Classification, Semi – Arid Environment, Red-Edge

3.1 Introduction

In Southern Africa, natural forest and rangeland resources form an important resource base for food and medicinal products that form part of people's subsistence as well as their economic base and well-being. One such natural forest is the mopane woodland. Mopane trees provide varied products that include construction and fence poles, wood for tools handles, carvings and utensils, firewood, rope, gum, tannin, medicines and resin, green manure, livestock browse and edible caterpillars (commonly referred to as mopane worms) (Mojeremane and Kgathi, 2005). The value of mopane woodland in Botswana alone has been estimated at US $3.3 million per annum, of which approximately 40% goes to producers who are primarily poor rural women (Sebego *et al.*, 2008). Recent studies have however showed that the long term sustainability of the woodland and its resources is under threat (Adelabu *et al.*, 2012; Ditlhogo *et al.*, 1996). Mapping tree species in mopane woodland are extremely important for forest management purposes. Until now, there are no category specific maps of species distribution in mopane woodland. Moreover, forest managers need to understand the species diversity to suggest possible management practices that will enable efficient and sustainable use of the resources emanating from mopane woodland. This will further support scientific knowledge of environmental management practices in Africa and other sites in the world more sensitive to global changes. However, it is nearly impossible to acquire detailed tree species information over large areas purely on the basis of field assessments. Therefore, enhanced methods are required to get explicit information on the tree species composition and distribution patterns.

Remote sensing has been a valuable source of information over the course of past few decades in mapping and monitoring forests (Quackenbush and Ke, 2007). It provides a cost-effective tool to help forest managers better understand forest characteristics, such as forest area, locations, and species, even down to the level of characterizing individual trees. The application of remote sensing in forest management began with manual interpretation of aerial photographs, but is increasingly reliant on new data and methods (Franklin, 2001). Over the last 20 years, the spectral and spatial resolution of satellite data has steadily increased. Medium resolution satellite data such as Landsat and SPOT can obtain regional-scale forest variables (Wulder and Franklin, 2003). Airborne hyperspectral sensors meet the requirements regarding spectral and spatial resolution for tree species mapping. However, due to the high costs and their limited availability, hyperspectral data have gained limited acceptance for operational use. As higher resolution

satellite imageries (e.g. RapidEye) become more available, there is an increasing potential to provide more detailed information. Unlike medium resolution satellite imagery, which provides an aggregated response over a region, individual trees are visible in high spatial resolution imagery. This provides opportunities to differentiate species and individual trees.

RapidEye has a relatively higher spatial resolution (5m) and a new spectral band (red-edge) in addition to the 4 standard bands (Blue, Green, Red, and Near Infrared). The Red Edge band is supposed to discriminate between healthy trees and trees that are impacted by disease and to enhance the separation between different species and age classes. Recent studies have shown that with the inclusion of the red-edge channel, RapidEye has the large capabilities for enhanced species mapping (Schuster *et al.*, 2012; Tigges *et al.*, 2013), however, its applicability in species classification especially in areas of limited training sites is still evolving.

Many classification methods have been used for tree species mapping using remote sensing data. These include maximum likelihood, minimum distance, discriminant analysis and spectral angle mapper classifiers (Carleer and Wolff, 2004; Cho *et al.*, 2010; Lobo, 1997). Maximum likelihood and minimum distance classifier are commonly used supervised classification methods with conventional multispectral data. However, there is a limitation with the application of these classifiers for mapping areas with limited training samples (Cho *et al.*, 2010). Previous studies in semi-arid environment have shown that collecting training samples is a mammoth task because of the terrain and maps produced by vegetation canopy introduces noise to vegetation classification (Ringrose *et al.*, 1989). As a solution to this problem, robust classification methods are essential for mapping species in semi-arid environment.

We assessed two commonly used machine learning classification algorithms namely: random forest (RF) and support vector machines (SVM) to discriminate between the species. Previous studies have shown that the RF or SVM algorithm can be successfully used in species mapping and classification purposes (Adam and Mutanga, 2009; Chan and Paelinckx, 2008; Krahwinkler and Rossman, 2011; Melgani and Bruzzone, 2004). To the best of our knowledge so far, no study has compared the main machine learning algorithms RF and SVM for tree species mapping in a semi-arid environment and in particular, using the new generation satellite sensors.

This study aims at separating *Colophospermum mopane* and its co-existing species in a semi-arid forest using spectral information provided by the RapidEye sensor applying Random Forest and

Support Vector Machines classification algorithms. We focus on the following research questions: (1). Can *Colophospermum mopane* and its co-existing species be separated using the RapidEye image with strategically positioned spectral bands? (2). Do the additional red-edge band of RapidEye improve the classification accuracy compared to the 4 standard bands? (3). Which of the classification algorithms (RF and SVM) is better for detailed tree species classification in areas with limited training samples based on classification accuracies?

3.2 Study Area

The study was undertaken in and around the Palapye/Tswapong axis of Central Region of Botswana (Longitude $27^{o}00$' E- $27^{o}33$' and Latitudes $22^{o}23$' - $22^{o}52$' S). It is situated along the main north-south highway and is about 230 km north of Gaborone, the capital city of Botswana (Sebego *et al.*, 2008). Soils around the area are well developed and variable, hence locally referred to as hardveld region. Temperature and rainfall regimes are highly variable, both temporally and spatially and are both characterized by seasonality in their occurrence (Bhalotra, 1987). The axis covered by this study is described as mopane bushveld because mopane tree is found in all its growth forms and is locally monospecific. However, several other tree species grow in association with mopane trees. Only the dominant tree species as identified by local ecologist have been used in this study (**Table 3.1**): *Grewia bicolar* GB; *Dichrostachys Cinerea* DC; *Acacia erubiscens* AE; *Acacia tortilis* AT. *Grewia bicolar* and *Dichrostachys Cinerea* are deciduous species with short leaves and several flowering periods while *Acacia tortilis* and *Acacia erubiscens* have a dual flowering period during the raining and dry season (Do *et al.*, 2008) . On the other hand, *Colophospermum mopane* is a deciduous slow-growing species, with an erect narrow crown. The leaves are pinnate with two large leaflets that can vary considerably in size on the same tree (Wessels *et al.*, 2007) and within a growing season (Sebego *et al.*, 2008). *C. mopane* drops its leaves in an irregular fashion from the onset of the dry season and is generally leafless from August to October. However, trees may retain their leaves between successive rainy seasons, depending on the amount and distribution of rainfall (Wessels *et al.*, 2007).

Table 3. 1: The number of sample plots, local names and the type code for *Colophospermum mopane* and its co-existing species.

Specie Name	Local Name	Type Code	No. of Plots
Grewia bicolar	Mogwana	GB	53
Dichrostachys Cinerea	Moselesele	DC	51
Acacia erubiscens	Moloto	AE	53
Acacia tortilis	Mosu	AT	54
Colophospermum mopane	Mopane	CM	55

3.3 Field Data Acquisition

The field data acquisition was conducted during the summer month of January 2012 coinciding with the image acquisition date. Field measurements were done following random points that were generated from an existing land cover map of the study area using Hawth's Analysis Tool (HAT) in ArcGIS 9.3 (Sebego and Arnberg, 2002). A handheld *Garmin eTrex30* GPS (± 3m accuaracy) was then used to navigate to the respective points. Thereafter, vegetation polygons in which one or more of these species (*Grewia bicolar* GB; *Dichrostachys Cinerea* DC; *Acacia erubiscens* AE; *Acacia tortilis* AT) co-exist with *Colophospermum mopane* was created around the centered point. The vegetation polygon was defined as covering 20 m × 20 m, where the target species were more homogenous and were representative of more than 80% of the target species in each plot. This method resulted in 51 to 55 vegetation polygons for each target species. Thereafter, the vegetation polygons were then used to create training area and then overlaid on the true colour composite RapidEye image using Environment for Visualizing Images (ENVI) software (ENVI, 2006). The metadata such as the site description (coordinates, altitude and land cover class) and general weather conditions were also recorded.

3.4 Image Acquisition and Data Preparation

High resolution multispectral imagery was acquired over the study area by the RapidEye sensor on 25 January 2012. The RapidEye data set composed of 5-band multispectral imagery with 5 m ground sampling distance (GSD). The five bands include: blue (440 – 510 nm), green (520 – 590 nm), red (630 – 685 nm), red-edge (690 – 730 nm) and near-infrared (760 – 850 nm). For the purpose of machine learning optimization, spectra pixels from RapidEye image (5 m × 5 m) were extracted using Environment for Visualizing Images (ENVI) software (ENVI, 2006). In this study, only pixels that fell completely within the measured polygons were used in the reference dataset so as to minimize variability and exclude mixed pixel effects of tree species (Adam *et al.*,

2012). Before field data was used for analysis, data for each polygon was then averaged to represent one sample.

The imagery over the study area contained 0% cloud cover, with a relatively clear atmosphere. All RapidEye's products are collected by a 12 bit imager. During on-ground processing, radiometric corrections are applied and all image data are scaled up to 16 bit dynamic range (RapidEye, 2011). The scaling is done with a constant factor that converts the (relative) pixel DNs from the sensor into values directly related to absolute radiances. The scaling factor was originally determined pre-launch. Previous experimentation performed by Naughton *et al.* (2011) verified that the image registration was within a single pixel, hence further geometric processing was not applied. The image was atmospherically corrected using the quick atmospheric correction procedure in ENVI 4.7 (ENVI, 2006).

3.5 Machine Learning Classifiers

3.5.1The Random Forest Algorithm (RF) and Support Vector Machine (SVM)

Random forest is a machine learning algorithm that employs a bagging (bootstrap aggregation) operation where a number of trees (*ntree*) are constructed based on a random subset of samples derived from the training data. Each tree is independently grown to maximum size based on a bootstrap sample from the training data set without any pruning, and each node is split using the best among a subset of input variables (*mtry*) (Breiman, 2001). The multiple classification trees then vote by plurality on the correct classification (Breiman, 2001; Lawrence *et al.*, 2006). The ensemble classifies the data that are not in the trees (out-of-bag or OOB data) and by averaging the OOB error rates from all trees, the random forest algorithm gives an error rate called the OOB classification error for each input variable (Breiman, 2001). Therefore, as part of the classification process, the RF algorithm produces a measure of importance of each input variable by comparing how much the OOB error increases when a variable is removed, whilst all others are left unchanged (Benediktsson and Sveinsson, 2004). RF algorithm is easy to implement as only two parameters (*ntree* and *mtry*) need to be optimized (Breiman, 2001; Özçift, 2011). A more detailed discussion on RF can be found in (Breiman, 2001).

SVM, on the other hand was originally introduced as a binary classifier (Vapnik, 1998) and is extensively described by Burges (1998) and Hsu *et al.* (2009). However, typical remote sensing problems usually involve identification of multiple classes (more than two). Adjustments are

made to the simple SVM binary classifier to operate as a multi-class classifier using methods such as one-against-all, one-against-others, and directed acyclic graph (Krahwinkler and Rossman, 2011). In its classical implementation, it uses two classes (e.g. presence/absence) of training samples within a multidimensional feature space to fit an optimal separating hyperplane (in each dimension, vector component is image gray-level). In this way, SVM tries to maximize the margin that is the distance between the closest training samples, or support vectors, and the hyperplane itself. SVM consists of projecting vectors into a high dimensional feature space by means of a kernel trick, then fitting the optimal hyperplane that separates classes using an optimization function. Several kernels are used in the literature. According to Hsu *et al.* (2009) and supported by many other authors (Krahwinkler and Rossman, 2011; Melgani and Bruzzone, 2004; Vapnik, 1998), the Gaussian radial basis function (RBF) has both advantages (i) of being very successful since it works in an infinite dimensional feature space; and (ii) having a single parameter contrary to the other well working kernels (e.g. polynomial). Noise in the data can be accounted for by defining a distance tolerating the data scattering, thus relaxing the decision constraint.

To demonstrate the effectiveness of RF and SVM for mapping *Colophospermum mopane* and its co-existing species, classifications of the 5 species were trained on 70% occurrences and evaluated on the remaining 30%. The same number of pixels were used for the presence class and for the absence class in order to avoid discrepancies due to unbalanced training sets (Hsu *et al.*, 2009; Vapnik, 1998). The EnMap box was used to carry out the RF and SVM analysis. EnMap is an IDL-based tool for classification and regression analysis of remote sensing imagery. It can be fully integrated into the commercially available IDL/ENVI environment. It may also be run as an add-on freely available, license-free and platform-independent processing environment for remote sensing imagery. EnMap includes: generic image files with an ENVI type text header for the image data as well as the continuous or categorical reference data and model outputs; two-step image analysis consisting of separate model parameterization and application; trained models are saved and may be applied several times, e.g., for transfer to other data sets and calculate variable importance and accuracy.

3.5.2 Optimization

Classifier parameters, namely the number of trees and the number of variables to be randomly selected from the available set of variables for the RF and the regularization parameter C and the

kernel parameter λ for the SVM, were selected (Hsu *et al.*, 2009). The goal was to identify optimal parameters so that the classifier could accurately predict unknown data. This method is easy to use, fast and can be more reliable than more advanced iterative techniques that do not always consider parameters as independent.

3.5.3 Classification and Accuracy Assessment

It has become customary in the remote-sensing literature to report the Kappa index of agreement for purposes of accuracy assessment, since Kappa also compares two maps that show a set of categories. However, recent studies have shown limitations of Kappa in that it gives information that is redundant or misleading for practical decision making. Furthermore, Pontius Jr. and Millones (2011) stated that the standard Kappa and its variants are frequently complicated to compute, difficult to understand and unhelpful to interpret. To solve this problem, Pontius Jr. and Millones (2011) recommend that the use of Kappa for purposes of accuracy assessment and map comparison be abandoned, and instead summarize the cross-tabulation matrix with two much simpler summary parameters: quantity disagreement and allocation disagreement.

In this study, accuracy assessments were obtained using the Kappa analysis and the quantity disagreement and allocation disagreement as proposed by Pontius Jr. and Millones (2011). A confusion matrix was constructed so as to compare the true class with the class assigned by the classifier and to calculate the overall accuracy as well as the class accuracies (Ismail *et al.*, 2008). We report on both statistics using the confusion matrix proposed by Pontius Jr. and Millones (2011) and available online at http://www.clarku.edu/~rpontius/.

The two machine learning algorithms were compared on the basis of the accuracy they produce when trained on the same set of training data. Based on the accuracy metrics obtained for each classifier in each accuracy assessments method, a statistical analysis can be performed to test if the difference is significantly equal or different. We compared the confusion matrix yielded by the RF and the SVM classifications by using the McNemar test with a 95% confidence interval. McNemar test is a non-parametric test that is simple to understand and execute. It has been shown to be sensitive and precise in comparing two or more accuracy assessments (Manandhar *et al.*, 2009; Petropoulos *et al.*, 2012). The test is based on chi square (z^2) statistics, computed from two error matrices and given as:

$$(z^2) = (f_{12} - f_{21})^2 / (f_{12} + f_{21})$$

Where f_{12} denotes the number of cases that are wrongly classified by classifier 1 but correctly classified by classifier 2 and f_{21} denotes the number of cases that are correctly classified by classifier 1 and wrongly classified by classifier 2 (Manandhar *et al.*, 2009). The difference in accuracy between a pair of classifications is viewed as being statistically significant at a confidence of 95% if the calculated z-score in McNemar test is larger than 1.96 (Manandhar *et al.*, 2009).

3.6 Results

3.6.1 Optimization

The results of optimizing random forest parameters (*ntree* and *mtry*) at each split has shown that, the setting of *mtry* value of 2 yielded the lowest OOB error rates. When examining the *ntree* parameter, results indicated that there were relatively stable OOB error rates after *ntree* settings value of 5000. Similarly, for the SVM optimization parameters C and λ, the best value was obtained at 0.01 and 10 for C and λ respectively. The results show that the changes in *ntree* and *mtry* for RF and λ and C for SVM influence the accuracy of the classification results produced. Therefore, the settings of *mtry* (2), *ntree* (5000), C (0.01) and λ (10) were used for all further analyses.

3.6.2 Classification Result

The vegetation map of RF and SVM classifications are presented in **Fig 3.1**. From the maps, it is clear that using both methods of classification (RF and SVM), mapping of CM species and its co-existing species yielded good results from RapidEye images. The SVM classification produced higher overall accuracies of 88.75% compared to RF classification of 85%.
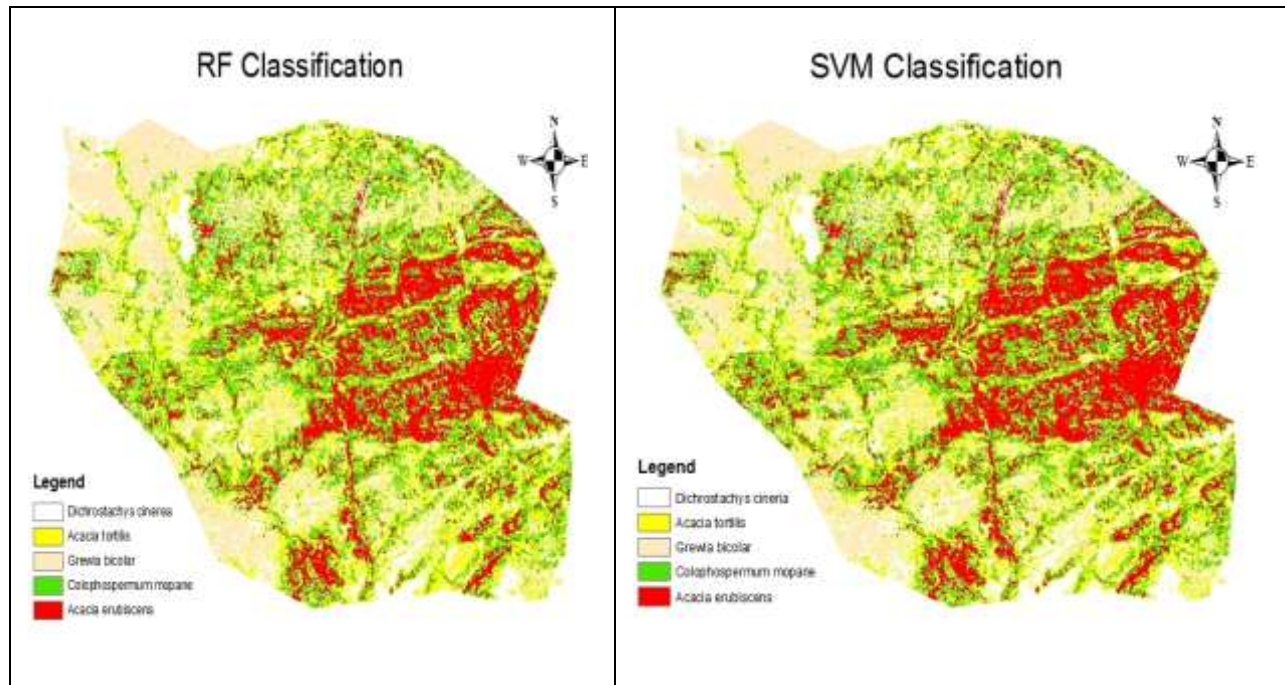
Figure 3. 1: Maps showing the classification of the 5 tree species using RF and SVM machine learning algorithm.

Generally, classification results were best in depicting CM and AE using both methods **(Fig 3.2)**. In fact, stands of CM were accurately classified with 100% class accuracy using both methods. Moreover, we tested the importance of the red-edge band as part of the rapid eye sensor for species classification in a semi- arid environment by running the classification with and without the red-edge band. Results showed that in both methods, when the red-edge band is excluded, the accuracies decreased by 8%. Furthermore, the relative importance of each of the RapidEye's bands (n = 5) in mapping the 5 species was measured using random forest algorithm.
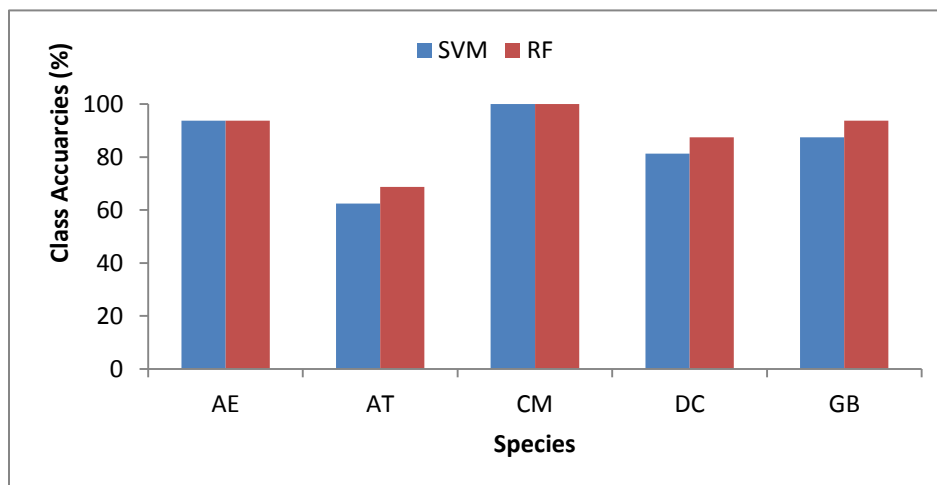
Figure 3. 2: Class accuracies for each species using the RF and SVM classification algorithm.

54

The result further indicated that band 3 (Red Band; 630-690 nm) had the highest mean decreasing accuracy, with only the lowest mean decreasing accuracy obtained with band 1 (Blue; 440-510nm) **(Fig 3.3)**.
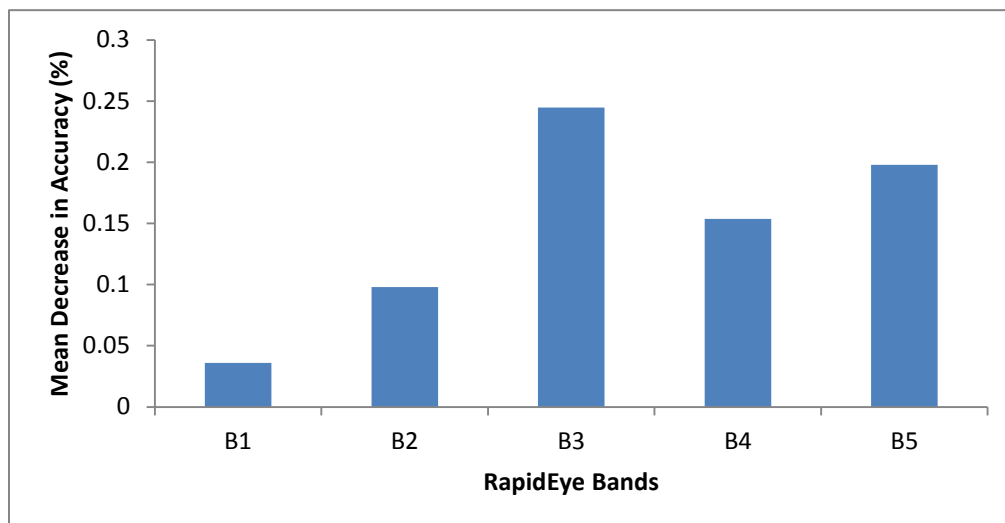
Figure 3. 3: Comparing the relative importance of each RapidEye band in mapping the 5 species using the mean decrease in accuracy. The mean decrease in accuracy was estimated using the random forest algorithm.

Similarly, we also measured the importance of each band in discriminating individual species among the other species **(Fig 3.4)**. Results indicate that band 3 (Red Band; 630-690 nm) and band 5 (NIR; 760-880nm) are more important in separating individual species from the other.
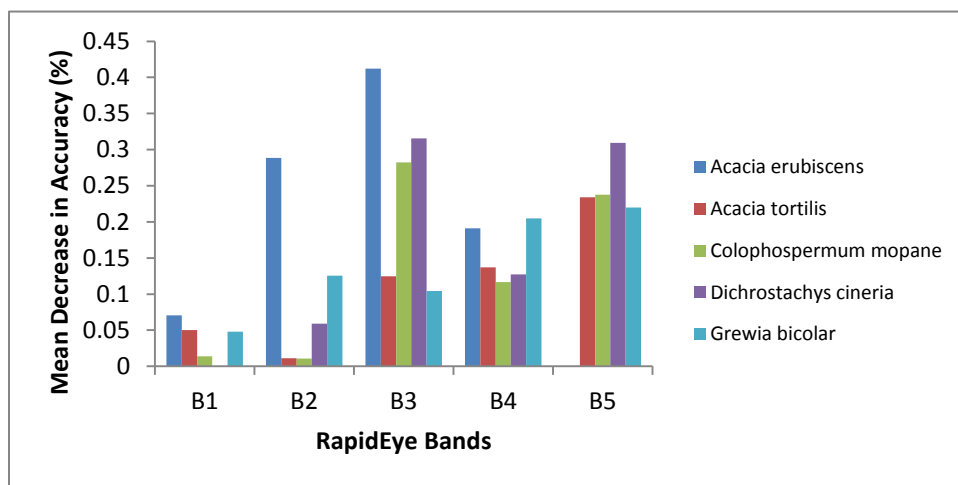
Figure 3. 4: Comparing the relative importance of each RapidEye band in mapping individual specie relative to other species using the mean decrease in accuracy. The mean decrease in accuracy was estimated using the random forest algorithm.

3.6.3 Classification Accuracy and Assesment

The confusion matrix derived from the test dataset for both RF and SVM are presented in **Table 3. 2**. The matrix consists of classification accuracy (CA) and overall accuracy (OA). When test data was used to test the RF classification result, the RF successfully mapped the five species (GB, DC, AE, AT, CM) with an overall accuracy of 85%. On the other hand, when the SVM was used for classification, the confusion matrix yielded an overall accuracy of 88.75%. Similarly, the accuracy assessment results for class pair indicates that *Acacia erubiscens* AE and *Grewia bicolar* GB; appears to be unique amongst the other species based on the highest classification accuracy of 93.75% (AE) and 93.75% & 100% (GB) for the SVM and RF classifications respectively. Hence it is easier to discriminate *Colophospermum mopane* from *Acacia erubiscens* and *Grewia bicolar* GB compared with *Dichrostachys cineria* DC which yielded a class pair accuracy of 82.35% and 76.47% for both the SVM and RF classifications respectively. Overall class accuracy of *Colophospermum mopane* was 84.21% for SVM and 72.73% for RF.

Table 3. 2: Comparison of confusion matrix obtained after the classification of *Colophospermum mopane* and its co-existing species from both the SVM and RF. The confusion matrix includes overall accuracy (OA) and class accuracy.

SVM						RF				
	AE	AT	CM	DC	GB	AE	AT	CM	DC	GB
AE	15	1	0	0	0	15	0	1	0	0
AT	0	11	1	3	1	0	10	3	3	0
CM	0	0	16	0	0	0	0	16	0	0
DC	1	0	1	14	0	1	0	2	13	0
GB	0	0	1	0	15	0	0	1	1	14
CA (%)	93.75	91.67	84.21	82.35	93.75	93.75	90.91	72.73	76.47	100
OA= 88.75%						OA= 85%				

The results of the performance of the two classification assessment (Kappa and Total disagreement) in mapping the five species are presented in **table 3.3** for both the SVM and RF classification methods. The total disagreement is separated into two components: quantity disagreement and allocation disagreement. Quantity disagreement is the amount of pixels of a class in the training data that is different from the quantity of pixels of the same class in the test data while allocation disagreement is location of a class of pixel in the training data that is different from the location of the same class in the test data. Kappa for both SVM (0.86) and RF

(0.81) in this study represents a strong agreement between the training data and the test data. **Table 3.3** also shows that in both methods, the total disagreement as proposed by Pontius Jr. and Millones (2011) (11% and 15% for SVM and RF respectively) is slightly lower than the disagreement from Kappa (14% and 19% for SVM and RF respectively).

Table 3. 3: Comparing Kappa and Total disagreement methods of classification assessments. NB: All calculations were done using the confusion matrix proposed by Pontius Jr. and Millones (2011) and sourced from http://www.clarku.edu/~rpontius/.

Parameters	SVM	RF
Kappa	0.86	0.81
Kappa Total disagreement (%)	14	19
Allocation disagreement (%)	5	9
Quantity disagreement (%)	6	6
Total disagreement (%)	11	15

We also used the confusion matrix to compare the accuracies of the two algorithm (RF and SVM) employed in this study using the McNemar. **Table 3.4** shows that there are no significant differences in the accuracies obtained from both the RF and SVM classifications. The McNemar test z score was less than 1.96 which is required for the two algorithms to be statistically significantly different at 95% confidence level.

Table 3. 4 : Comparison of SVM and RF using McNemar Test

		SVM		
		Correctly Classified	Misclassified	Total
	Correctly Classified	67	4	71
RF	Misclassified	1	8	9
	Total	68	12	80

* McNemar z Score = 1.342

3.7 Discussion

This study highlights the effect of new generational multispectral image (RapidEye) in mapping tree species in semi-arid environment where there are limited training samples. We also set out to compare two machine learning algorithms and two classification accuracy assessments methods for mapping tree species in semi-arid environment.

3.7.1 Spectral Classification of the Tree Species

Classification of different tree species in semi-arid area such as Palapye in Botswana can be challenging because of the change in leaf structure and orientation as a result of limited soil

moisture (Sebego and Arnberg, 2002; Sebego *et al.*, 2008). However, this study has shown that rapid-eye satellite data is highly suitable for classifying tree species in mopane woodland. The achieved class accuracies of the various tree species ranged between 72.73% - 100% for RF and 82.35%-100% for SVM. In both algorithms, the lowest class values were found for *Colophospermum mopane* (CM) (72.73% & 82.35% for RF and SVM respectively). The higher observed misclassifications of CM may be due to spectral overlaps between CM and the other tree species. Similarly, from our observations in the field, a great deal of the identified misclassifications could be explained by the very complex forest structure in the test area. Most of the leaves of the tree species investigated in this study look similar. This characteristic may have led to mixed pixels that may cause misclassifications.

In general, the present studies show an improvement to the works of Sebego *et al.* (2008) which was able to map mopane and its co-existing species with 74% accuracy. There are two reasons that may be responsible for this. Firstly, the spatial resolution of the Landsat TM image used for their study is coarse (30 m) compared with that of the RapidEye (5 m) used in the present study. Secondly, the classification algorithms (RF & SVM) used in the present study has been proved to outperform the one used by previous study (Maximum Likelihood)(Nitze *et al.*, 2012). We therefore conclude that spatial resolution and algorithms plays an important role in classifying tree species in semi-arid environment.

3.7.2 The Role of RapidEye Red-Edge Band in Tree Species Classification

Although the five tree species could be separated accurately by using only the 4 standard bands (i.e. Blue, Green, Red and Near-Infrared), the use of the additional red-edge band led to a significant improvement of classification accuracy. For instance, the classification accuracies increased from 78% and 80.25% to 85% and 88.75% for RF and SVM respectively when the red edge band was added. The positive effect of the red-edge band can be explained by its relationship with the chlorophyll content of vegetation (Gitelson and Merzlyak, 1994). Several studies have reported the red-edge spectrum to be specifically sensitive to vegetation chlorophyll content (Gitelson and Merzlyak, 1994; Mutanga and Skidmore, 2007). Chlorophyll content can be regarded as an additional factor to explain particular sensitivities to the red edge channel. The chlorophyll in green leaves absorbs light for photosynthesis in the red-edge region of the spectrum (Thomas *et al.*, 2008). For this reason, the red-edge region is most useful for detecting the absorption of visible light by the chlorophyll pigments. Moreover, since at any time each of

the tree species will be at different health state, the red-edge region will be efficient in separating the species based on their health status.

However, the present study did not show that the red-edge is the most important band in classifying the tree species. In fact, the red region and the near infrared region outperformed the red-edge band in classifying the species. Other studies that have found the red and NIR region more important for classification of forest species using remote sensing data suggest that the presence of red-edge may only be important when incorporated with other standard bands (Schuster *et al.*, 2012; Tapsall *et al.*, 2010).

Overall, these results are consistent with the previous studies that have found that RapidEye bands, with its high spectral resolution has great potential in terms of classifying and mapping vegetation species (Kim *et al.*, 2011; Krahwinkler and Rossman, 2011; Sah, 2013; Tapsall *et al.*, 2010).

3.7.3 Machine Leaning Algorithm's for Species Classification in Semi-Arid Environment

The Random Forest and Support Vector Machine were compared on their ability to map *Colophospermum Mopane* (CM) and its four major coexisting species. McNemar test shows that there is no statistically significant difference between RF and SVM. Nevertheless, SVM slightly out performed RF by 3.75%. Previous studies have shown that RF and SVM are the best techniques for mapping tree species using high spatial resolution imagery such as RapidEye (Pouteau *et al.*, 2011). These techniques are better than conventional classification methods such as Maximum likelihood, Minimum Distance e.t.c. in that they handle voluminous, highly multi-collinear, multispectral dataset (Redo and Millington, 2011). Moreover, they are very computing time efficient compared to the conventional methods. To our knowledge, SVM and RF have never been compared for mapping tree species in areas with limited training samples like semi-arid environment reminiscent of ours.

Other studies have showed that SVM generally outperforms RF especially when the number of training pixels is small (Foody and Mathur, 2006; Vapnik, 1998). The main reason is most likely the result of the paradigm of SVM based on a small pixel sample (i.e. support vectors) (Foody and Mathur, 2006). Consequently, SVM can be trained with few meaningful pixels and is able to fit limited information. The results from this study shows that in our context, SVM is able to map the tree species with higher accuracy from only small training pixels of each species (presence

and absence) compared to RF. Based on the above, we therefore postulate that SVM should be used for mapping tree species in semi-arid environment with small training pixels.

Furthermore, it is clear from our results that kappa which incorporates an adjustment for random allocation agreement is not a good statistics to determine the level of agreement between the training dataset and test dataset in classification process because kappa has numerous conceptual flaws. Most importantly, it fails to distinguish quantity disagreement from allocation disagreement (Redo and Millington, 2011). Therefore it is far more important to report how much less than perfect the classification is, rather than how much better than random the classification is (Pontius and Millones, 2011; Redo and Millington, 2011). The quantity disagreement and allocation therefore present a better option to kappa.

3.8 Conclusions

This study has demonstrated the effectiveness of using new multispectral sensors with high spatial resolution and specific bands such as RapidEye when classifying tree species in a semi-arid environment. Specifically, the study has highlighted three important conclusions:

1. *Colophospermum mopane* and its co-existing species have a strong potential to be mapped using high resolution data such as RapidEye imagery with high accuracy.

2. The presence of red-edge band in the RapidEye sensor has potential for classifying tree species in semi-arid environment when incorporated with other standard bands.

3. Considering the relative high accuracies, low cost, simplicity and few parameters needed for operation, SVM will be preferred over RF in areas where there are limited training samples.

Acknowledgement

CHAPTER FOUR

Insect Defoliation Levels Discrimination Using Ground based Hyperspectral Dataset

This chapter is based on:

S. **Adelabu**, O. Mutanga, E. Adam and R. Sebego, "Spectral Discrimination of Insect Defoliation levels in Mopane Woodland using Hyperspectral Data ," *IEEE Journal of Selected Topics in Earth Observation and Remote Sensing*. 7 (1), 177-186, 2014.

ABSTRACT

Mopane woodland are a source of valuable resources that contribute substantially to rural economies and nutrition across Southern Africa. However, a number of factors have, of late, brought the sustainability of the mopane woodland resources into question. One of such factors is the difficulty in monitoring of defoliation process within the woodland. In this study we set out to discriminate the levels of change in forest canopy cover detectable after insect defoliation using ground based hyperspectral measurements in mopane woodland. Canopy spectral measurements were taken from three levels of defoliation: Undefoliated (UD), Partly defoliated (PD) and Refoliating plants (R) using ASD FieldSpec Handheld 2^{TM}. A pre-filtering approach (ANOVA) was compared with random forest independent variable selector in selecting the significant wavelengths for classification. Furthermore, a backward feature elimination method was used to select optimal wavelengths for discriminating the different levels of defoliation in mopane woodland. Results show that optimal wavelengths located at 707nm, 710nm, 711nm, 712nm, 713nm, 714nm, 727nm and 1066nm were able to discriminate between the three levels of defoliation. The results further show that there was no significant difference in the overall accuracy of classification when random forest variable selector was used 82.42% (Kappa = 0.64) and the pre-filtering approach (ANOVA) 81.21% (Kappa = 0.68) used before building the classification. Overall, the study clearly demonstrated that the dynamic process of defoliation in mopane woodland can be assessed and detected using hyperspectral dataset and effective algorithm for discrimination.

Keywords: Mopane woodland, defoliation, random forest, ANOVA, hyperspectral

4.1 Introduction

Mopane (*Colophospermum mopane*) woodland are a source of valuable resources that contribute substantially to rural economies and nutrition across Southern Africa (Wessels *et al.*, 2007). Mopane trees provide varied products that include construction and fence poles, wood for tools handle, carvings and utensils, firewood, rope, gum, tannin, medicines and resin, green manure, livestock browse and edible caterpillars (commonly referred to as mopane worms) (Mojeremane and Kgathi, 2005). The value of mopane woodland in South Africa alone has been estimated at £2,850 ha-1 with annual population veldt worth £57m, of which approximately 40% goes to producers who are primarily poor rural women (Styles and Skinner, 2000). However, a number of factors have, of late, brought the sustainability of the mopane woodland resources into question. One of such factors is the difficulty in monitoring defoliation process within the woodland. This has resulted not only in the depletion of woodland resources in most rural areas but also low vitality and productivity of the woodland (Makhado *et al.*, 2012). However, no report has related the depletion of resources to insect defoliation taking place in mopane woodland. Mopane tree defoliation follows the same pattern as other insect defoliators and when the outbreak occurs, about 200 mopane worms feed on a single tree leading to 90% of mopane trees, if not all, left without leaves within a mopane woodland (Ditlhogo *et al.*, 1996). While mopane woodland often recover within a relatively short period after defoliation with little mortality, continuous defoliation may lead to deplorable effect such as decrease in chemical defence resulting in loss of canopies that may be fatal in the long term (Wessels *et al.*, 2007). Defoliation may affect the quantity and quality of subsequent canopy growth and therefore the quality and availability of mopane woodland resources.

Previous studies have demonstrated insect defoliation as a general response of stress associated with biophysical variables (de Beurs and Townsend, 2008; Fraser and Latifovic, 2005). For instance, changes in Leaf Area Index (LAI) and Chlorophyll Content (CC) have been related to variation in photosynthetic activities (Niemann *et al.*, 2012). Both LAI and CC increases with increase in health status of plants (Koike, 1987). Therefore, when a particular tree is stressed due to factors such as insect defoliation, the health status of the tree depreciates. It is important to note that other factors such as drought and wind can also contribute to defoliation and hence changes in biophysical variables. However, it is difficult to spectrally differentiate between insect and other environmental factors induced defoliation. However, mopane worm defoliation is different from drought or wind induced defoliation in that defoliation follows good-healthy

63

growth patterns that would otherwise not occur if there was a severe drought or wind. Moreover, insect defoliation occurs in patches (within the healthy forests) but drought or wind would impact on all vegetation. Furthermore, insect defoliation in mopane woodland mainly occurs in the summer (November-April; rain season) after which defoliated plants come to leaf for the second time. The normal drought or wind induced defoliation for deciduous tree occurs in winter therefore making it easier to detect insect defoliation during summer in Mopane woodland. Nevertheless, we assumed in our study that drought is constant over a wide area and by restricting our measurements to the period immediately after defoliation; we increased the chances of capturing the worm effect.

Since major leaf components (e.g. pigments, water, carbon, and nitrogen) produce distinctive reflectance signals at specific wavelengths; advanced remotely sensing data with high spectral and moderate spatial resolution is needed for effective assessment of vegetation response to stress levels after insect defoliation. In this regard, hyperspectral measurement will be of great benefit to the study of canopy level changes as a result of insect defoliation because of its reliability and relative accessibility (Mutanga *et al.*, 2009; Stone and Coops, 2004). Although multispectral sensors have been found to map insect defoliated area (Fraser and Latifovic, 2005; Kharuk *et al.*, 2007), hyperspectral sensors bring an advantage of spectral details at different levels of defoliation. Moreover, hyperspectral datasets facilitates detailed spectral measurement of reflectance related to biochemical and biophysical attributes of plants, which are associated with its structure, physiology and phenology, and therefore with its health status, a mammoth task that could not be done accurately with multi-spectral sensors (Cho and Skidmore, 2006; Mutanga *et al.*, 2004; Schlerf and Atzberger, 2012). Furthermore, there is mounting evidence that hyperspectral data have the capability, not only to assess defoliation, but to detect early signs of defoliation even before visual symptoms are apparent due to its high spectral bands (Ismail *et al.*, 2008). To date and to the best of our knowledge, no study has been carried out using ground based hyperspectral dataset to discriminate different levels of insect induced defoliation.

However, while hyperspectral data may be good in identifying insect defoliation levels, the process of analyzing it is challenging due to the high dimensionality, over fitting when applying statistical methods, an excessive demand for sufficient field samples, and high cost (Vaiphasa *et al.*, 2007). Therefore, identifying the optimal and powerful wavelengths using variable selection methods without losing any important information is a pre-requisite to hyperspectral remote

sensing application (Adam and Mutanga, 2009). This method is done, not only to reduce the amount of variables to simplify, but also to determine which explanatory variables are most suitable in discriminating. Different statistical techniques have been used to identify the optimal wavelengths such as discriminant analysis, canonical variate analysis, classification trees, support vector machines; and principal component analysis (Adam and Mutanga, 2009; Mutanga *et al.*, 2009).

Recently, random forest (RF) algorithm which was developed by (Breiman, 2001), has been successfully used as a variable selection and classification algorithm for hyperspectral data (Lawrence *et al.*, 2006). Hyperspectral data is likely to be noisy due to factors that include saturation of signal, mislabeling, sensor problems and viewing geometry (DeFries and Chan, 2000). However, random forest algorithm has been found to be more robust with respect to noise than other tree based ensemble methods (Ismail and Mutanga, 2011). The effectiveness of random forest can be explained by its ability to exploit noise and reduce dimensionality in hyperspectral dataset to create a more diverse classifier (Breiman, 2001). Studies in the past have applied pre-filtering procedure for reducing dimensionality of dataset before building the random forest classifier (Ismail and Mutanga, 2011). The interest is to remove irrelevant variables and to score individual variable according to their discriminating power i.e. their ability to separate the variables. While pre-filtering approach may be able to reduce the large dataset before building the random forest classifier for discrimination purpose, no studies have used hyperspectral dataset to compare the accuracy of results obtained when pre-filtering is used and when independent variable selector of random forest is used.

Dynamic detection of levels of defoliation and the areas impacted by the worms are extremely important for mapping and management purposes. This will in turn minimize the potential loss of resources in mopane woodland of Southern Africa in the long run. Up till now, there are no category specific defoliation maps for mopane woodland. Initial studies (Sebego *et al.*, 2008) did not have such knowledge available to them and thus did not discriminate the levels of defoliation within the woodland. Moreover, resource managers need to know the impacts, the vulnerability and the extent of defoliation to suggest possible management practices that will enable efficient and sustainable use of the resources emanating from mopane woodland. This will further support scientific knowledge of environmental management practices and food security in Africa and other sites around the world more sensitive to global changes. Furthermore, the study introduces

the usage of ground based hyperspectral dataset for discriminating the different health levels in insect induced defoliated woodland.

In this study therefore, our main aim is to discriminate the levels of change in forest canopy cover detectable after insect defoliation by ground based hyperspectral measurements in mopane woodland. Spectral discrimination will improve identification of the mopane canopy levels associated with dynamic defoliation processes that may improve our ability to map, monitor and assess the mopane worm defoliation of mopane woodland. Consequently, we set out to investigate the following; (i) whether insect defoliation levels in mopane woodland can be discriminated; (ii) if yes, which region(s) of electromagnetic spectrum can best be used in discriminating the defoliation levels. An additional objective was to examine the effectiveness of a filtering process which precedes the actual construction of random forest in discriminating the levels of insect defoliation in mopane woodland.

4.2 Materials and Methods

4.2.1 Study Area

The study was conducted in the eastern part of Central District of Botswana (27°E and 27°33`E and 22$^{\circ}$23`S and 22°52`S). The selection of this area was based on two major reasons: firstly, it is an area with typical pockets of mopane woodland where the species is found in all its various growth forms (tree, tall shrubs and short shrubs). Secondly, depletion of the worms has been reported in the area over the past few years with heavy defoliation occurring (Sebego *et al.*, 2008). Intensive field work was conducted with the help of locals and ecological experts to identify the areas that are associated with constant insect defoliation in the study area.

4.2.2 Identifying the Defoliation Levels and Field Reflectance Measurements

During the reconnaissance study, we identified three primary categories of canopy impact ranging from non-impacted undefoliated plants (UD) to partly defoliated plants (PD) and finally refoliating plants after severe defoliation (R) at different sites of the study area (**Table 4.1**). The time series of defoliation in mopane woodland is represented by its canopy impacts. While the non-impacted undefoliated canopies (UD) represents time before defoliation, the partly defoliated (PD) and refoliating (R) canopies represent during and after defoliation respectively. In the healthy non-impacted mopane canopies, leaves are distinctive, consisting of two large triangular leathery leaflets, sometimes likened to butterfly wings. Leaf area index   and
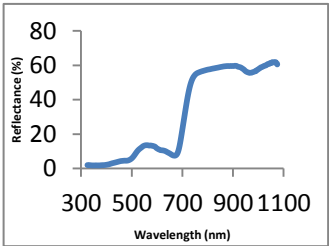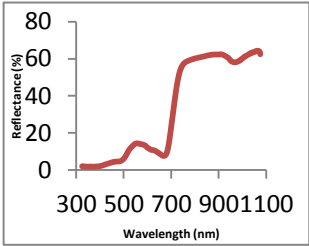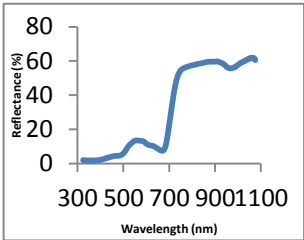
chlorophyll content are usually high in UD state whereas and although the PD canopies appears green, they may be visually indistinguishable from UD canopies except for the part eaten by the worms (Ciesla, 2003). Leaves become pale green gradually in PD state leading to total defoliation and hence lower leaf area index and chlorophyll content (Adelabu *et al.*, 2012). However, after total defoliation, new shoots (refoliating canopy) spring out usually after 2 weeks. Leaflets are initially bright red-brown and glossy. The subtle visual difference in the canopy colour levels of UD and PD challenged us to consider more detailed and specific spectral-based tools for monitoring and assessing the three levels of mopane worm defoliation.

To explore the possibilities of using hyperspectral data to detect changes in canopy status of mopane after insect defoliation therefore, field reflectance measurements were recorded using the Analytical Spectral Devices (ASD) FieldSpec Handheld 2[TM], (ASD, 2005) to collect reflectance data in natural stands of Mopane at canopy level as recommended by (Barry *et al.*, 2008). The ASD spectrometer wavelength ranges from 325 to 1075nm with a sampling interval of 1nm. It has a spectral resolution of < 3nm at 700nm and wavelength accuracy of ±1nm (ASD, 2005).

A total of 20-25 reflectance measurements for each of the three categories namely: Healthy Green Undefoliated (UD), Partly Defoliated (PD) and Refoliated (R) were taken at canopy level in 55 different plots for each state demarcated after careful observations of the study area by ecological experts on 19th January, 2012 to determine areas that are used to insect defoliation (**Table 4.1**). This was done in full midday sun with spectral reflectance collected across wide wavelengths ranging from 325-1075 nm for each canopy at 1-nm-wide narrowband sensor. Canopy reflectance was measured using a pole erected approximately 2m above the canopies with a 25° field-of-view connected to the hyperspectral spectro-radiometer. To eliminate the influence of branches, below-canopy vegetation or soil as a result reflectance of eaten parts, canopies of the PD were clipped and spectrometer measurements were taken on plane surface for each of the PD plots. For the purpose of this study, PD state was described as leaves in which 40-60% of the canopies of the tree were eaten by the leaves. All the spectrometer measurements were normalized using a standard white panel after every 5-10 spectra measurements. The metadata such as the site description (coordinates, altitude and land cover class) and general weather conditions were also recorded alongside the spectral measurements (Milton *et al.*, 2009).

These spectral measurements from each plot (n = 20–25) were then averaged to represent the spectral reflectance of each category in the different plots demarcated.

Table 4. 1 : The number of sample plots, total number of spectral measurements collected and the mean reflectance for different defoliation levels in mopane woodland

Defoliation levels	Codes	Images	Number of Plots	Number of Measurements	Mean Reflectance
Healthy Green Undefoliated	UD		55	1265	
Partly Defoliated	PD		55	1250	
Refoliated	R		55	1320	

4.2.3 Data Analysis

4.2.3.1 Random Forest Algorithm

Random Forest algorithm is a commonly used and accurate non-parametric technique developed by (Breiman, 2001) that can handle both categorical and continuous independent variables

(Horning, 2010). Random forest is based on model aggregation that can be used for both classification and regression problems (Benediktsson and Sveinsson, 2004; Hudak *et al.*, 2012).The algorithm consists of a set of random decision trees, each tree contributing to the final classification outcome. Through this algorithm, the variables of a dataset can be ranked and the most important ones can be identified to explain the outcome of interest.

In the random forest algorithm, each tree is grown on a separate training set that is a bootstrap replicate of the original data (Breiman, 2001). The training data is sampled to create an in-bag partition to construct the tree (2/3 of the training data), and a smaller out-of-bag (OOB) partition (1/3 of the training data set) to validate the performance of each constructed tree (Özçift, 2011). The multiple trees then vote by majority on correct classification. To demonstrate the effectiveness of random forest algorithm for discriminating insect defoliation in mopane woodland, we first carried out a classification process that grows random forest on the hyperspectral data with ANOVA as a filter and then compared it with classification built without pre-filtering features **(Fig 4.1)**.

4.2.3.2 Variables Selection: Filter Approach

In this study, a one-way analysis of variance (ANOVA) was used as a baseline filter approach. This was done to check if pre-filtering approach could improve the accuracy of random forest method in the selection of variables, which was used as input for classification. Furthermore, ANOVA analysis allowed us to examine whether there was statistical difference between the mean reflectance of the different defoliation levels at every spectral wavelength. For the purposes of this study, the null hypothesis was that there was no significant difference in spectral reflectance of different defoliation classes at every spectral wavelength; H0: $\mu1 = \mu2 = \mu3$ versus the alternative hypothesis that there was a difference; vis: H1: $\mu1 \neq \mu2 \neq \mu3$. In this case, $\mu1$, $\mu2$ and $\mu3$ are the mean spectral reflectance for UD, PD and R at every spectral wavelength, respectively. Wavelengths with no statistical significance were discarded while the significant wavelengths were used as input variables into the relevant classification algorithm. The distribution of spectral responses at every spectral wavelength was assumed to be normal, consistent with the Central limit theorem (Horning, 2010). We tested our hypothesis at every wavelength between 325 and 1075 nm (a total of 751 wavelengths) with 95% confidence limits ($\alpha = 0.05$).

4.2.3.3 Variable Importance Selection Using the Random Forest Algorithm

The classification mode of random forest builds multiple trees based on random bootstrapped samples of the training data. Consequently, random forest calculates three variable importance measures, namely, the number of times each variable is selected, the Gini importance and the permutation accuracy importance measure (Hudak *et al.*, 2012). The Gini importance of random forest allows for explicit feature elimination but may not be optimally adapted to spectral data due to the topology of its constituent classification trees which are based on orthogonal splits in feature space (Menze *et al.*, 2009). In order to obtain high accuracy in classification, previous studies have suggested that two parameters of random forest need to be optimized (Mutanga *et al.*, 2004). These parameters are the number of trees (*ntree*) growth and the number of variables (*mtry*) used in each tree split (Benediktsson and Sveinsson, 2004). A large number of *ntree*s can be developed using default value of *mtry* which is usually the square root of the number of variables (Özçift, 2011) .

For the purpose of our study, we adopted the permutation accuracy importance measure using the OOB method to calculate the importance of a specific predictor variable (in our case wavelengths n=751) in discriminating between the different levels of mopane defoliation. Therefore, the importance of each variable used in this study is inversely proportional to the accuracy of the classification when that variable (wavelength) was permuted randomly (Cohen and Fiorella, 1998; Milton *et al.*, 2009). The importance of each wavelength was then estimated by

(a)   randomly permuting the mean reflectance values of each wavelength for the OOB samples and then passing down the new OOB to each tree for new predictions;

(b)   averaging the difference in misclassification rate between the original and the new OOB data over all the trees that are grown and;

(c)   using the average as a ranking index for identifying the measure of importance of the wavelengths with relatively large importance in the classification process (Rodriguez-Galiano *et al.*, 2012).

The random forest library developed in R statistical software (R Development Core Team 2008) was used to implement the RF algorithm.

4.2.3.4 Optimal Variable Selection

The limitation of variable selection methods is in their inability to choose the most important variables for classification. While ANOVA and random forest can provide a measure of variables importance, they do not automatically choose the optimal number of variables that yield the best classification accuracy (Liaw and Wiener, 2002). The challenge is therefore to select the optimum number of wavelengths that can yield the smallest misclassification error rate. In this regard, a backward feature elimination method (BFE) integrated with random forest algorithm as part of the evaluation process was implemented (Adam and Mutanga, 2010). The method starts by building multiple random forests after which it iteratively discards the variables with the smallest variable importance. At each iteration, the method selects the best *mtry* and *ntree* while the least promising variable is eliminated and the OOB error calculated. The smallest subset of variables with the lowest OOB error is then selected. To select the optimal wavelengths for discriminating the different defoliation levels in mopane woodland, we input the entire important wavelengths derived from both ANOVA and the variable selector of random forest (n=120) using the BFE model. Two wavelengths were eliminated at each iteration from the model and the error was calculated using the OOB error estimate method.
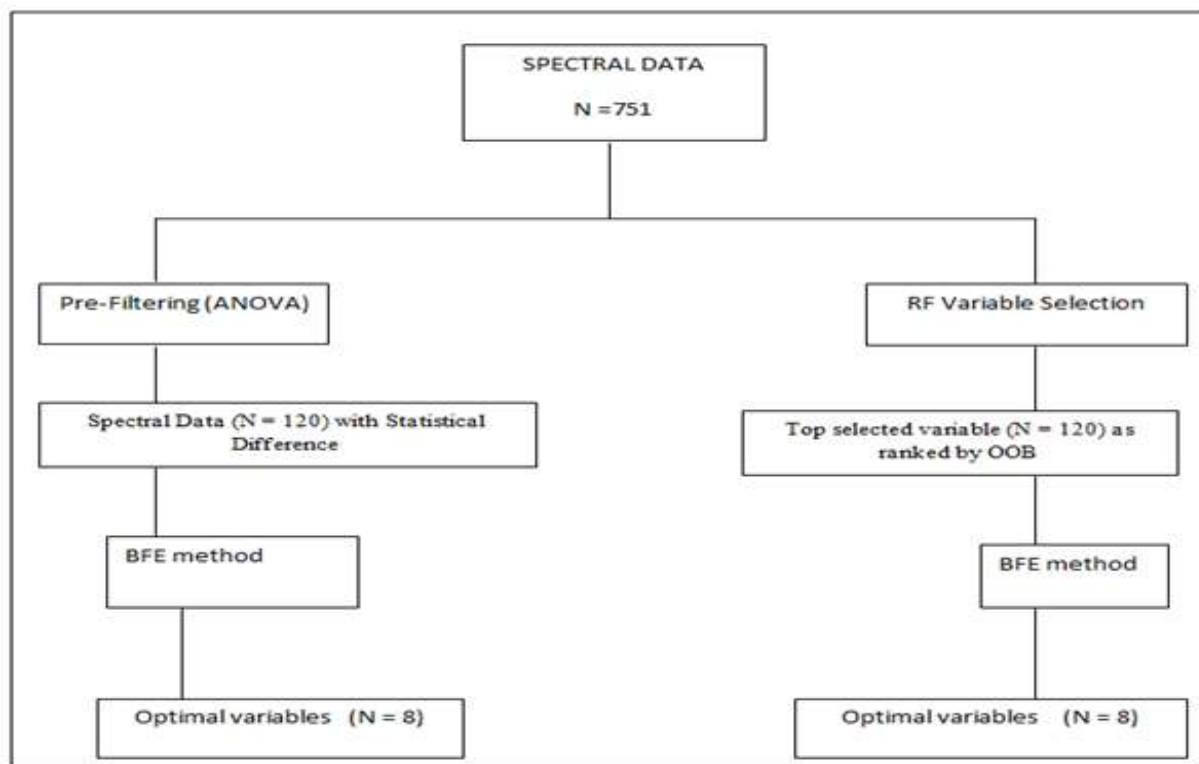
Figure 4. 1 : Summary of Data Analysis

4.2.3.5 Accuracy Assessment

For any algorithm to be adjudged well, its prediction performance has to be tested. Previous studies have shown that in random forests, there is no need for cross-validation or a separate test to get an unbiased estimate of test error as this can be done using the internally built OOB error estimate (Adam and Mutanga, 2009; Lawrence *et al.*, 2006). In this study the OOB error was used to estimate the classification accuracy and for choosing the optimal number of wavelengths that yielded the smallest error rate. A confusion matrix was constructed to compare the true class with the class assigned by the classifier and to calculate the overall accuracy as well as the producer's and user's accuracies. The producer's accuracy was computed by splitting the number of correctly classified trees in each crown condition class by the number of data sets used for that class (column total in the confusion matrix). User's accuracy was calculated by dividing the number of correctly classified trees by the total number of trees that were classified in that crown condition class (row total in the confusion matrix)(Ismail *et al.*, 2008). In addition, Kappa analysis, a discrete multivariate technique was used in accuracy assessment. Kappa analysis yielded Khat statistic (an estimate of Kappa), which was calculated to determine if one error matrix is significantly different from another (Cohen and Fiorella, 1998).

4.3 Results

4.3.1 Selecting Significant Wavelengths Using Filter Approach (ANOVA)

In this study, we set out to examine whether there was statistical difference between the mean reflectance of the different insect defoliation levels in mopane woodland at every spectral wavelength. Result from ANOVA for the three class combinations (PD vs. UD vs. R) for reflectance spectra is shown in **Fig 4.2**. The shaded areas indicate the spectral wavelength locations where the class pairs exhibit statistically significant differences ($p < 0.05$) in spectral response.

The conclusions from the ANOVA test are that the mean reflectances between the partly defoliated, undefoliated and refoliated stages of mopane woodland are significantly different in many measured wavelengths (n=120) in the visible and infrared regions of electromagnetic spectrum. The reflectance spectra data sets show significant differences in two spectral regions, the red edge and the near infrared regions of the electromagnetic spectrum.

Table 4.2 shows the frequency of the significant wavelengths grouped into the three spectral domains which is widely employed for vegetation discrimination from hyperspectral dataset (Kumar *et al.*, 2001). The table shows that there are no statistically significant wavelengths located in the visible (blue and Green & red) region for discriminating the three different defoliation levels. However, the table also show that Red-Edge (N = 65) and near-infrared (N = 55) are the most important regions where defoliation levels has the most statistically significant wavelengths. To achieve our objective of investigating the effect of pre-filtering approach on random forest, the subset of significant wavelengths (N = 120) obtained from ANOVA was used as input for classification. The result from this is also expected to narrow the number of wavelength(s) for discriminating the insect defoliation levels.
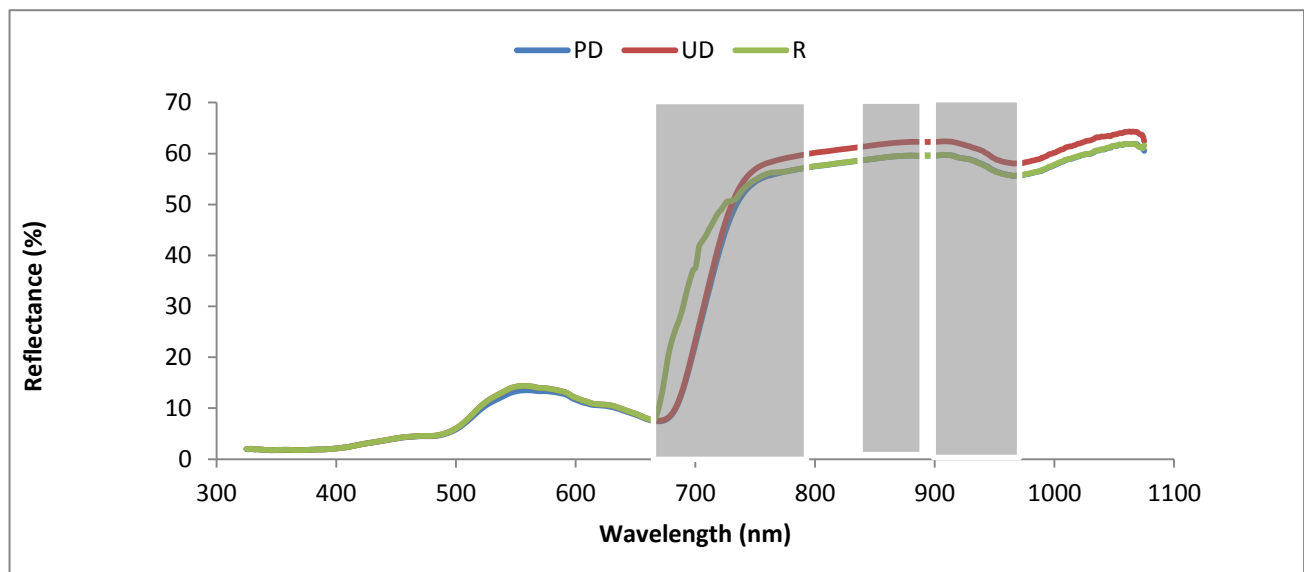
Figure 4. 2 : ANOVA results for the partly defoliated, undefoliated and refoliated stages using their mean spectra: The Grey shades indicate the electromagnetic region where there were significant differences (α =0.05).

Table 4. 2: Frequency of significant wavelengths for discriminating the defoliation levels grouped into the three spectral domains defined by (Kumar *et al.*, 2001).

Wavelength Region (nm)	Description	Wavelength no.	Significant Wavelength	
			No	%
325-670	Visible	346	0	0
680-750	Red Edge	71	65	91.54
780-1075	Near- Infrared	295	55	18.64

4.3.2 Measuring Variable Importance Using Random Forest

The OOB method in random forest algorithm was applied to measure the relative importance of each of the entire wavelengths (N = 751) in discriminating the three defoliation levels. These variables (wavelengths) yielded an OOB error rate of 26.72%. The mean decrease in accuracy was calculated and used to rank the wavelengths **(Fig 4.3)**. The results clearly indicated that the top 120 significant wavelengths with the highest mean decrease in accuracy are located predominantly in the red edge (N= 70) and near-infrared (N=35). Of importance is also the ability of the OOB method in random forest to separate noisy data as indicated in the negative values in figure 4.3. However unlike the ANOVA results, the random forest variable selector was able to select a pocket of wavelengths with high mean decrease in accuracy in the visible region (n=15) as part of the top 120 **(Fig 4.4)**. When the 120 top ranked wavelengths were selected, the OOB error rate was 19.39%. The top 120 wavelengths were then selected and used as input for classifying the defoliation levels to correspond to the number of significant wavelengths obtained from ANOVA.
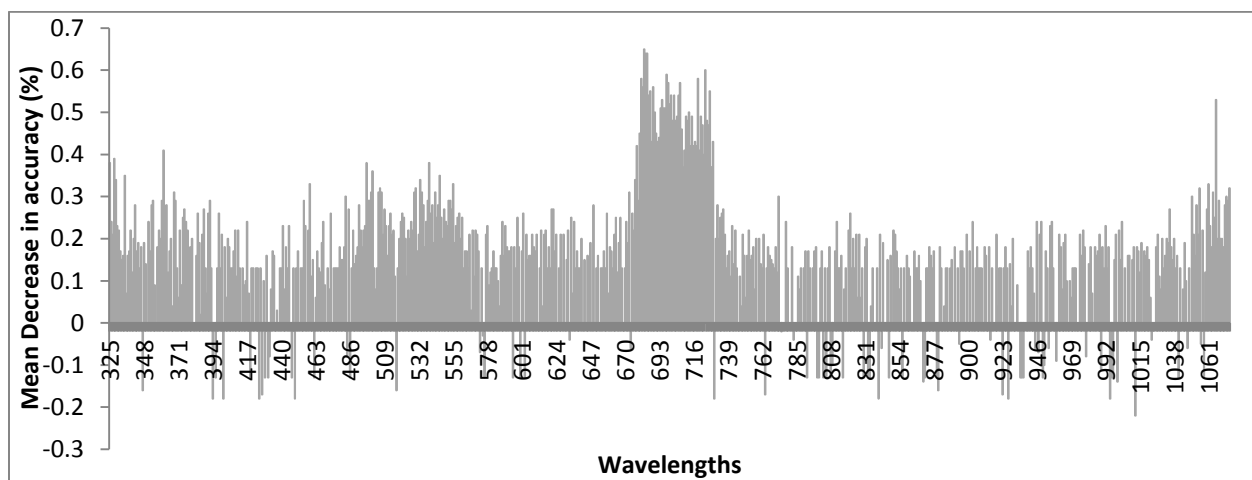
Figure 4. 3: Variables importance as selected by the random forest algorithm for 751 wavelengths. The important wavelengths are those with the highest mean decrease accuracy.
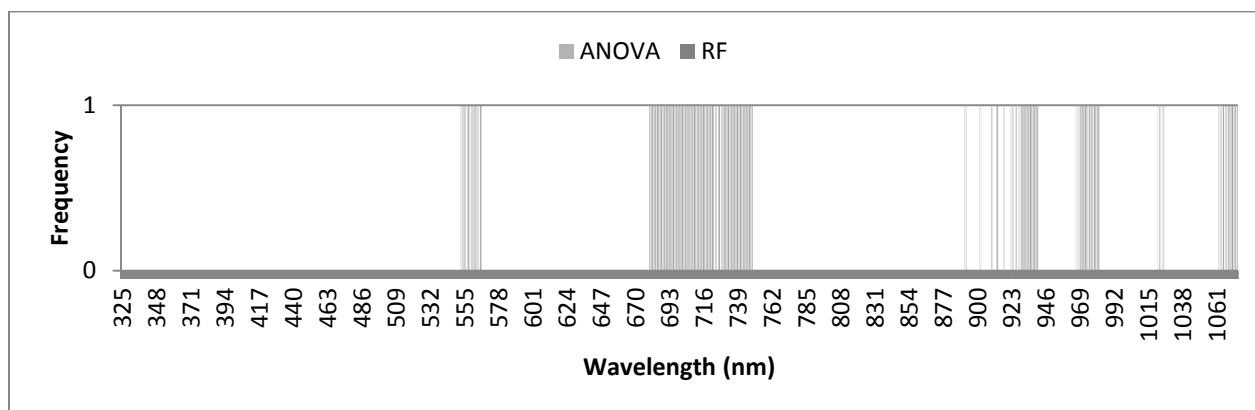
Figure 4. 4: The selection of significant wavelengths by the analysis of variance (ANOVA) and the random forest method where 0 indicate no significant wavelength selected and 1 significant wavelength selected.

4.3.3 Backward Feature Elimination

The BFE method was performed for the significant wavelengths obtained from both ANOVA and random forest (N = 120) to select the optimal numbers of wavelength in classifying the three levels of defoliation (UD, PD, and R). The optimal wavelengths with the smallest OOB error are shown in **Fig 4.5**. The results of the BFE show that in both cases (ANOVA and random forest), 8 optimal number of wavelengths yielded the lowest OOB error of 18.79% and 17.53% respectively. The results of the backward feature elimination process further indicate that the lowest misclassification error rate as determined by ANOVA was located at 711nm, 712nm, 713nm, 714nm, 710nm, 727nm, 1066nm and 707nm while the random forest method's lowest misclassification error rate was located 713nm, 712nm, 719nm, 705nm, 707nm, 727nm, 711nm and 1066nm (the ranking is based on the importance measures). It is interesting to note that the ANOVA and the random forest methods of selection follow a similar trend (**Figure 4. 5**). The top 8 wavelengths were then used as input variables into the final random forest model to classify the defoliation levels in mopane woodland.

ANOVA	Random Forest
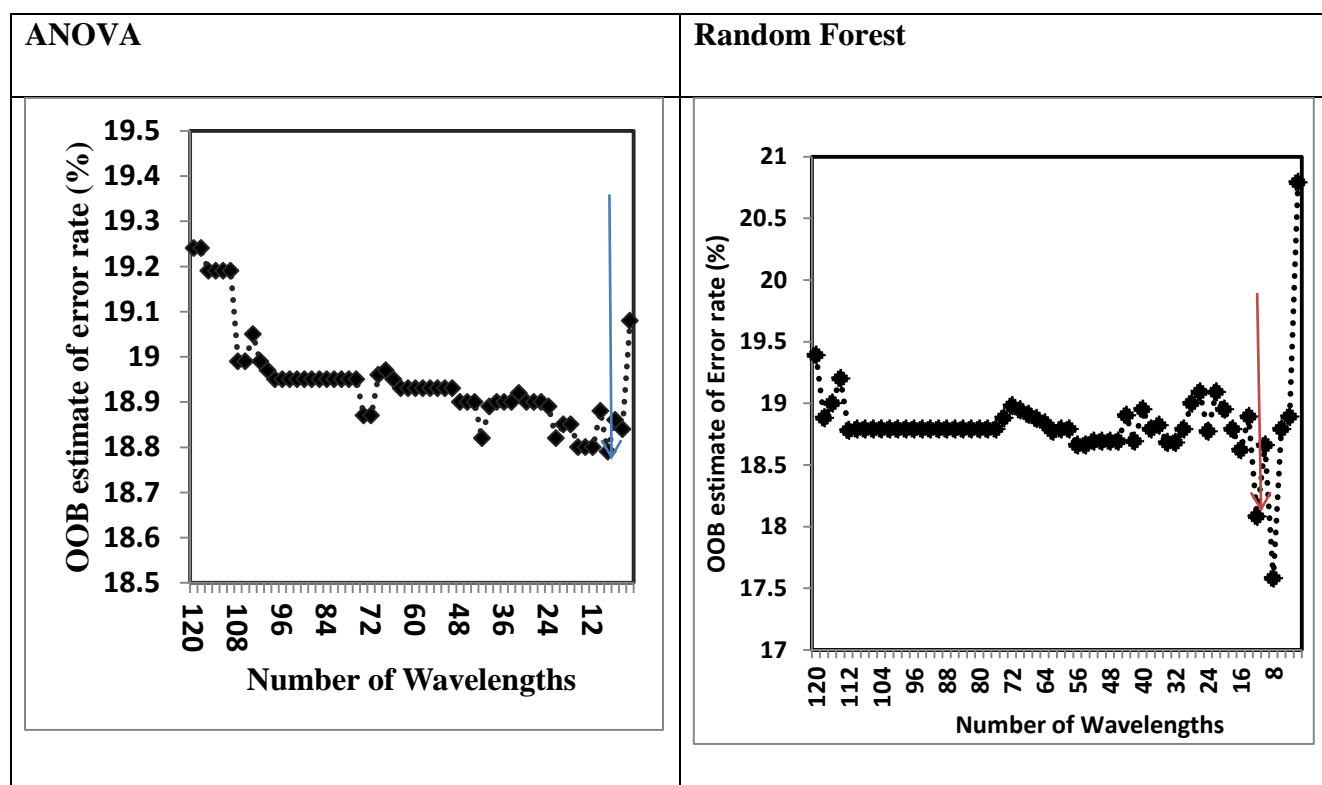
Figure 4. 5: The backward feature elimination method for identifying the optimal wavelengths of discriminating defolaition levels using the significants wavelengths derived from ANOVA and Random forest. The best number of wavelengths with the lowest error rate is shown by the arrows.

4.3.4 Classification and Accuracy Assessment

The optimal subset of wavelengths (N =8) derived from both ANOVA and random forest was used as input variables in the RF classifier to discriminate between the three different defoliation levels in mopane woodland. **Table 4.3** illustrates the confusion matrix derived from the OOB error estimation for the significant wavelengths derived from ANOVA and random forest. This matrix includes overall accuracy (OA), Kappa, user's accuracy (UA) and producer's accuracy (PA). When all the significant wavelengths from both methods (ANOVA and random forest (N = 120) were used to test the classification accuracy, the random forest algorithm successfully distinguished the three different levels of defoliation (PD, UD and R) with an overall accuracy of 79% and 80.61% respectively. On the other hand, when utilizing the top 8 wavelengths for both methods, ANOVA and random forest methods yielded an overall accuracy of 81.28% (k = 0.64) and 82.42% (k= 0.68) respectively as determined by the OOB. Similarly, the producer's accuracy for discriminating between PD, UD, R levels was higher (77.27% and 78.26%) when the top 8 wavelengths selected by both ANOVA (711nm, 712nm, 713nm, 714nm, 710nm, 727nm, 1066nm and 707nm) and random forest (713nm, 712nm, 719nm, 705nm, 707nm,

76

727nm, 711nm and 1066nm) methods compared to the 64.71% and 76.74% obtained when all the wavelengths (N=120) were used. Additionally, we used the OOB estimate of error rate to compare the performance of the ANOVA and random forest methods of selecting optimal wavelengths for discriminating the three defoliation levels. The result of performance assessments for both methods using different subsets of wavelengths compared with the misclassification error obtained when the whole dataset was used are shown in **Fig 4.6**. It is clear, over a range of the three subsets of wavelengths presented; the overall misclassification error rates obtained by the random forest algorithm were much lower than the misclassification error rates obtained by the pre-filtering method of ANOVA and the whole dataset. It is interesting that the use of the top 8 wavelengths yielded the lowest misclassification error rate for both the ANOVA (18.72 %) and random forest (17.58 %) and that the highest misclassification error (25%) was obtained using the whole wavelengths (n=751).

Table 4. 3: Comparison of confusion matrix obtained after the classification of the three levels of defoliation (PD, UD, R) using the optimal wavelengths for discriminating. The confusion matrix includes overall accuracy (OA), Kappa, class error (CE), producer's accuracy (PA) and user's accuracy (UA)

				ANOVA								**RF**			
	PD	**UD**	**R**	**Total**	**UA (%)**	**PA (%)**	**CE (%)**		**PD**	**UD**	**R**	**Total**	**UA (%)**	**PA (%)**	**CE (%)**
PD	34	21	0	55	77.27	61.82	38.18	**PD**	36	19	0	55	78.26	65.45	34.55
UD	10	45	0	55	68.18	81.82	18.18	**UD**	10	45	0	55	70.31	81.82	18.18
R	0	0	55	55	100	100	0	**R**	0	0	55	55	100	100	0
Total	44	66	55	165				**Total**	46	64	55	165			
CE (%)	22.73	32	0					**CE (%)**	21.7	29.7	0				
Kappa = 0.64, OA = 81.21%								**Kappa = 0.68, , OA = 82.42%**							

Figure 4. 6: Comparison between the performance of the ANOVA and the random forest using different subsets of wavelengths selected by BFE. The misclassification error rate was estimated using the OOB estimate of error rate at N= 751, 120, and 8.

4.4 Discussion

4.4.1 Discriminating the Defoliation Levels Using ANOVA, Random Forest Variable Selector and BFE

Previous studies have investigated the ability of ground based hyperspectral measurement to discriminate the damage caused by insects on vegetation (Coops *et al.*, 2004; Ismail *et al.*, 2008). However, to the best of our knowledge, no studies have reported spectral discrimination of levels of defoliation in mopane woodland. The present study is the first to report on the levels of defoliation namely undefoliated, partly defoliated and refoliating levels in mopane woodland with results indicating that hyperspectral reflectance data measured in the field can successfully discriminate the varying levels of canopy impact. Results from the two methods (ANOVA and random forest) used in this study show that there is a significant difference between the mean reflectance for the three levels with a large number of significant wavelengths located in the visible and near-infrared regions of the electromagnetic spectrum. More specifically, results show that in both methods, at least 60% of significant wavelengths are located in region between 680nm to 750nm (n= 65 for ANOVA and n=70 for random forest) while the other wavelengths are majorly located in the near infrared (n=55 for ANOVA and n=35 for random forest) and green (n=15 for random forest).

The ability of random forest to select a pocket of wavelength (n=15) at the green region of electromagnetic spectrum can be explained by its effectiveness in exploiting noise and reducing dimensionality of hyperspectral data when compared to ANOVA(Benediktsson and Sveinsson, 2004; Breiman, 2001). As mentioned in (Carter and Miller, 1994), green and red edge wavelengths are sensitive to the chlorophyll content. Studies have also shown that blue shift in canopy reflectance at the red region of the electromagnetic spectrum are generally indicators of stress in vegetation (Carter, 1994; Carter and Knapp, 2001). The locations of the significant wavelengths in this region confirm that worms affect the chlorophyll content of mopane leaves. Although chlorophyll content was not measured in this study, it is plausible to say that the prevalence of significant wavelengths in the red edge region is due to the effects of chlorophyll. It is therefore assumed that both the chlorophyll concentrations will vary with the different levels of insect defoliation in mopane woodland. This hypothesis is still speculative and would therefore have to be further tested using foliar biochemistry analysis. While it is the pigments that control the spectral responses of leaves in the visible wavelengths, the infrared region is sensitive to the cellular structure and water content of leaves (Coops *et al.*, 2004).The significant difference in the near infrared region results obtained in this study may be an indicator of water stress suffered by the mopane leaves. Thus the results suggest that future research may benefit from correlation analyses between hyperspectral data and the mopane leaves CC, LAI and cell structure.

Furthermore, one of our objectives was to find the wavelengths with strongest discriminating power between the defoliation levels. However, both ANOVA and the random forest variable selection methods were unable to automatically choose the optimal number of variables that could yield the lowest error rate. Hence, the BFE method used in this study provided the optimal numbers of important variables that offer the lowest misclassification error rate in both methods. Although no single wavelength was capable of total separability, spectral separability of all the levels is possible when using eight wavelengths combination. It is worthy of note that in both methods, the majority of wavelengths (n=7 in both methods) that produce the best combination of wavelengths with highest discriminatory power are located in the red edge region of electromagnetic spectrum. This suggests and confirms the results of (Carter and Knapp, 2001) and our earlier results that the red edge is an important region for discriminating defoliation levels as a result of insect defoliation. The successful use of BFE for the discrimination of the different level of insect defoliation with only a few wavelengths (N=8) confirmed its utility as a

good variable selection method (Lawrence *et al.*, 2006). This result confirms the assertions of previous studies (Adam *et al.*, 2012) that have reported that the BFE has been applied in remote sensing image classifications with much better performance. Overall, the result shows the excellent performance of the BFE method in dimensionality and noise reduction without sacrificing significant spectral information. The results therefore encourage more investigation into the effectiveness of using airborne and satellite hyperspectral and multispectral sensors for mapping insect defoliation in mopane woodland.

4.4.2 Effectiveness of Pre-Filtering Process on Random Forest in Discriminating Defoliation Levels

The result from the pre-filtering process (ANOVA) employed in this study did not significantly differ from the result from the random variable selector. For instance, the classification accuracy yielded 79% (ANOVA) and 80.61% (RF variable selector) with a KHAT value of 0.70 and 0.71 when selecting the top 120 important wavelengths for discrimination. However, these accuracies were relatively higher than when the whole dataset (n=751) was used for classification (75%). This may suggest the effectiveness of filtering process is considerable before building the random forest classifier. As the whole dataset has a very large number of variables but it is expected that only very few are important, filtering increases the probability to capture informative variables. As a consequence, growing a few trees on small sized subsets results in higher accuracy values compared with the forest grown on the original dataset. The results are comparable with Lawrence *et al.* (2006) who found that using full hyperspectral dataset in classification produced lower overall accuracy than when the significant wavelengths were selected using a filter. Moreover, the results indicate that, in the model-based analysis, the increase of hyperspectral bands could lead to a decrease in the classification accuracy because the noise in the data propagates through the classification model (Benediktsson and Sveinsson, 2004).

The results from the BFE further indicate that when the optimal selected wavelengths (n=8) was used for classification, it produces an overall accuracy of 81.28% and 82.42% and KHAT value of 0.72 and 0.74 for ANOVA and RF respectively. This further shows that there is no significant difference in the overall accuracy of both method of filtering. Although previous studies have showed that pre-filtering is an important task before building random forest classifier for discrimination (Adam *et al.*, 2011; Ismail and Mutanga, 2011), our results suggest that filtering

process itself can be done using the random forest variable selector with good classification accuracy rather than using a different filter like ANOVA to prevent time wastage and inconsistency in the selection process.

 In summary, this study clearly demonstrated that the dynamic process of defoliation from healthy undefoliated canopies through partly defoliated and then refoliating canopies after insect defoliation can be assessed and detected using hyperspectral dataset and effective algorithm for discrimination. We therefore postulate that remotely sensed area-wide hyperspectral and multispectral imagery could be effectively applied for early detection and spatial distribution of multiple health levels of vegetation after insect defoliation if the image data covers the spectral ranges as described above.

4.5 Conclusions

In this study, we examined the possibility of using ground based hyperspectral data to discriminate between three levels of mopane worm defoliation severity (undefoliated, partly defoliated and refoliating coded by UD, PD and R respectively). A better understanding has been gained about specific regions of the electromagnetic spectrum that offer the useful information content for discriminating insect defoliation in mopane woodland. The study has shown that there is significant difference between the mean reflectance for all the three levels of defoliation with a large number of significant wavelengths located in the red edge and near infrared region of electromagnetic spectrum. Furthermore, the important pre-requisite (i.e. optimal wavelengths selection) for the potential upscaling of results to either airborne or spaceborne platform was established. Although no single wavelength was able to discriminate between all the defoliation levels, wavelengths located at 707nm, 711nm, 712nm, 713nm, 727nm and 1066nm show the greatest potential for discrimination. We further established the effectiveness of the random forest variable selector over the pre-filtering approach in the classification process. Overall, the result from this study provides evidence that encourages the capability of hyperspectral data for discriminating insect defoliation levels in mopane woodland.

Future work will include further improvement of hyperspectral signal analysis through the correlation analyses between hyperspectral data and the mopane leaves chlorophyll content and structure based on those wavelengths that are determined to be most significant. In this regard, we expect that the result from this study to be useful for early detection, mapping and

management of mopane defoliation as a result of mopane woodland thereby reducing the depletion of resources derived from the woodland. Furthermore, the potential for assimilation of the optimal bands obtained in this study could help in developing future land-use-land-cover mapping sensors.

CHAPTER FIVE

Classification of Insect Defoliation Levels Using New Generation Multispectral Imagery

This chapter is based on:

ABSTRACT

The prospect of regular assessments of insect defoliation using remote sensing technologies has increased in recent years through advances in the understanding of the spectral reflectance properties of vegetation. The aim of the present study was to evaluate the ability of the red edge channel of RapidEye imagery to discriminate different levels of insect defoliation in an African savanna. Random Forest and Support vector machine classification algorithms were applied using different sets of spectral analysis involving the red edge band. Results show that the integration of information from red edge increases classification accuracy of insect defoliation levels in all analysis performed in the study. For instance, when all the 5 bands of RapidEye imagery were used for classification, the overall accuracies increases about 19% and 21% for SVM and RF, respectively, as opposed to when the red edge channel was excluded. We also found that the Normalized Difference Red-Edge index yielded a better accuracy than Normalized Difference Vegetation Index. We conclude that the red-edge channel of relatively affordable and readily available high-resolution multispectral satellite data such as RapidEye has the potential to considerably improve insect defoliation classification especially in sub-Saharan Africa where data availability is limited.

Keywords: defoliation, red-edge, RapidEye, support vector machine

5.1 Introduction

The prospect of regular assessments of insect defoliation using remote sensing technologies has increased in recent years through advances in the understanding of the spectral reflectance properties of vegetation (Coops *et al.*, 2004; Pietrzykowski *et al.*, 2006; Stone and Coops, 2004). Recent evidence suggests that new damaging insect defoliators are appearing at an increasing rate which could affect the future sustainability of forest industries in developing world such as Africa (Ditlhogo *et al.*, 1996). In Southern Africa, canopy defoliation caused by insects represents a major disorder in many forest ecosystems. Damage from insect affects the forest growth and disturbs its productivity and value over large areas. Moreover, the economic losses due to insect attacks and failure to account for insect defoliation can result in large errors in long-term predictions of carbon sequestration (de Beurs and Townsend, 2008; Kovacs *et al.*, 2005). However, literature on insect defoliation classification in sub-Sahara-Africa using remote

sensing is still rudimentary. This could be as a result of lack of understanding of the defoliators, techniques, and unavailability of right sensors for classification among other.

The remote sensing approach in assessing and monitoring insect defoliation has focused on relating differences in spectral response to chlorosis (yellowing), foliage reddening, or foliage reduction over time (Franklin, 2001; Hall *et al.*, 2003a). However, high technical potential for improvement of the accuracy of insect defoliation classification results are still needed (Schuster *et al.*, 2012). Thus, developments are still ongoing for improving the capabilities of satellite, sensors and classification algorithms for insect defoliation classification. With regards to resolutions of remote sensing sensors, recent studies have shown that hyperspectral data could adequately be used for insect defoliation classification regarding spectral and spatial resolution (Adelabu *et al.*, 2013; Carter and Knapp, 2001; Lawrence and Labus, 2003). Nevertheless, owing to its high costs and limited availability, hyperspectral data have gained limited acceptance for operational use. Development has therefore shifted towards the introduction of additional bands in the broad-band multispectral sensors, in particular the red-edge spectrum (Schuster *et al.*, 2012).

Several researchers have associated the changes in chlorophyll concentration to the shift in the red edge which is the inflection point that occurs in the rapid transition between red and near infra-red reflectance (Gitelson *et al.*, 1996). This changes have been related to plant stress, forest decline, and leaf development (Thomas *et al.*, 2008; Zarco-Tejada *et al.*, 2000). Majority of researches on the red-edge region of electromagnetic spectrum evaluate chlorophyll content or the physiological status of plants in view of identifying vegetation stress (Gitelson *et al.*, 2006; Mutanga and Skidmore, 2007). Similarly, few studies have analyzed the relationship between chlorophyll content and vegetation indices, especially relations between wavelength of the red edge and reflectance in the red region of the spectrum (Gitelson *et al.*, 1996). The most common vegetation spectral index, Normalized Difference Vegetation Index (NDVI), Tucker (1979) has been found to be useful for defoliation mapping because it has been related to changes in chlorophyll concentration and estimation of photosynthetic activity (Carter and Knapp, 2001; Gupta *et al.*, 2003). Studies have however shown that NDVI becomes saturated over high biomass (Mutanga and Skidmore, 2004b) and are very sensitive to canopy background brightness (Huete *et al.*, 2002). However, the NDVI-RE (red-edge adaptation of NDVI) have recently been successfully tested in relations to canopy cover (Schuster *et al.*, 2012; Souza *et al.*,

2010), it is therefore necessary to investigate it as a potential substitute to the classic NDVI for classifying insect defoliation.

Previous studies have inferred red-edge from mainly reflectance taking from field spectrometry, introduction of narrow band analysis has however brought development to imaging spectroscopy especially the airborne sensors (Gupta *et al.*, 2003; Schuster *et al.*, 2012; Smith *et al.*, 2004). While the Medium Resolution Imaging Spectrometer (MERIS) sensor operated as payload on ENVISAT was the first to provide discontinuous red edge spectral bands (Dash and Curran, 2007), the data has not been very popular in African environment due to the cost involved. Other companies such as German RapidEye since early 2009 have developed high-resolution multispectral satellite operationally providing the red edge spectrum. Previous study by Schuster *et al.* (2012) has tested the red-edge channel for improving land use classifications with great results, however no study has tested the ability of the red edge of multispectral imagery for classifying insect defoliations. The good results of Schuster *et al.* (2012) prompted the present study. Consequently, the aim of the present study is to evaluate the benefit of the red-edge band of rapid eye imagery in discriminating different levels of insect defoliation.

5.2 Study Area and Field Data Collection

The study was conducted in mopane woodland in the eastern part of Central District of Botswana (27°E and 27°33`E and 22$^{\circ}$23`S and 22°52`S) (**Fig 5.1**). The area was chosen because heavy defoliation has been reported in the area over the past few years (Sebego *et al.*, 2008). Intensive field work was conducted with the help of locals and ecological experts to identify areas that are associated with constant insect (mopane worm) defoliation. Visual observation shows that the time series of defoliation in mopane woodland are in terms of its canopy impacts. Three primary categories of canopy impact ranging from non-impacted undefoliated plants (UD) to partly defoliated plants (PD) and finally refoliating plants after severe defoliation (R) at different sites of the study area were observed (**Fig 5.2**). While UD represents time before defoliation, PD and R represent during and after defoliation respectively. In the healthy non-impacted Mopane canopies (UD), leaves are distinctive, consisting of two large triangular leathery leaflets, sometimes likened to butterfly wings. Although the PD canopies appear green, they may be visually indistinguishable from UD canopies except for the part eaten by the worms. Leaves become pale green gradually in PD state leading to total defoliation. However, after total

defoliation takes place, new shoots (refoliating canopy) also spring out usually 2 weeks after. Leaflets are initially bright red-brown and glossy.

It is important to note that while other factors such as drought and wind can also contribute to defoliation and hence changes in biophysical variables, it is difficult to spectrally differentiate between insect and other environmental factors induced defoliation. However, mopane worm defoliation is different from drought or wind induced defoliation in that defoliation follows good-healthy growth patterns that would otherwise not occur if there was a severe drought or wind. Moreover, insect defoliation occurs in patches (within the healthy forests) but drought or wind would impact on all vegetation. Furthermore, insect defoliation in mopane woodland mainly occurs in the summer (November-April; rain season) after which defoliated plants come to leaf for the second time. The normal drought or wind induced defoliation for deciduous tree occurs in winter therefore making it easier to detect insect defoliation during summer in mopane woodland. Nevertheless, we assumed that in our study drought is constant over a wide area and by restricting our measurements to the period immediately after defoliation, we increased the chances of capturing the worm effect.

The field data acquisition was conducted during the summer month of January 2012, coinciding with the image acquisition date. Field measurements were done following random points that were generated from an existing land cover map of the study area using Hawth's Analysis Tool (HAT) in ArcGIS 9.3 (Sebego and Arnberg, 2002). A handheld *Garmin eTrex30* GPS (± 3m accuaracy) was then used to navigate to the respective points. Thereafter, defoliation plots were created around the centered point in which 80% of the trees are said to belong to a particular defoliation level considering the total tree cover per plot. To eliminate the influence of branches, below-canopy vegetation or soil as a result reflectance of eaten parts, canopies of the PD were clipped and spectrometer measurements were taken on plane surface for each of the PD plots. For the purpose of this study, PD state was described as leaves in which 40-60% of the canopies of the tree were eaten by the leaves.

The defoliation plot was defined as covering 20 m × 20 m resulting in 80-84 defoliation plots for each defoliation level. Thereafter, the defoliation plots were then used to create training area and then overlaid on the true colour composite RapidEye image for extraction of the pixels spectra (5 m × 5 m) using Environment for Visualizing Images (ENVI) software (ENVI, 2006). The

metadata such as the site description (coordinates, altitude, and land cover class) and general weather conditions were also recorded.
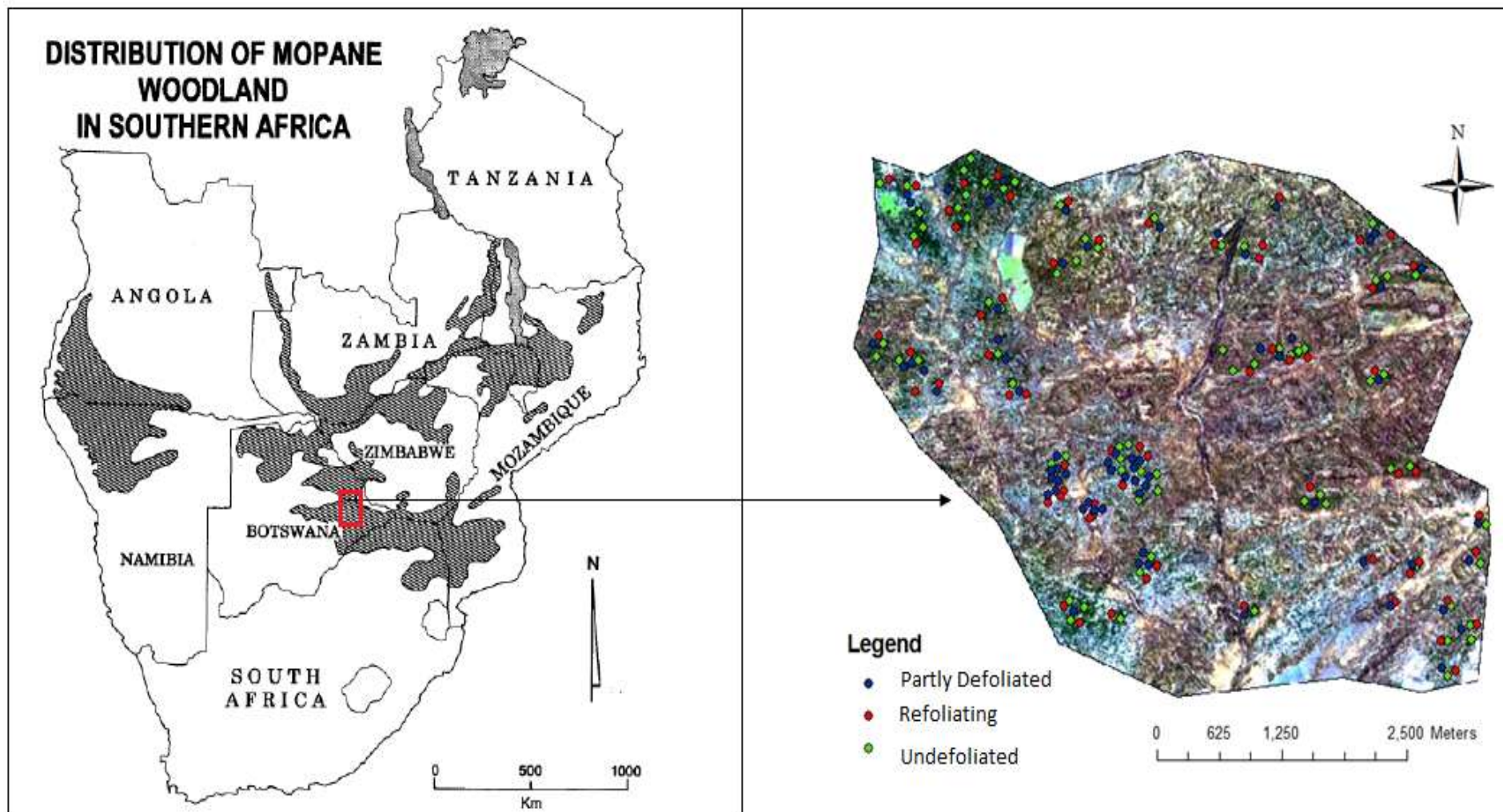
Figure 5. 1: Distribution of Mopane woodland in South Eastern Botswana and the Blue, Green, and Red Band Combination of RapidEye Image of the study area

Defoliation levels	Codes	Images
Healthy GreenUndefoliated	UD	
Partly Defoliated	PD	
Refoliating	R	

Figure 5. 2 : Images for Different Defoliation Levels in Mopane Woodland

5.3 Satellite Remotely Sensed Data

RapidEye image used in this study was acquired on 25 January 2012. The RapidEye data set composed of 5-band multispectral imagery with 5 m ground sampling distance (GSD). The five bands include: blue (440 – 510 nm), green (520 – 590 nm), red (630 – 685 nm), red-edge (690 – 730 nm) and, near-infrared (760 – 850 nm). The imagery over the study area contained 0% cloud cover, with a relatively clear atmosphere. Previous experiments performed by Naughton *et al.* (2011) led to the assumption that the image registration was within a single pixel, hence no further geometric processing was applied. The image was atmospherically corrected using the quick atmospheric correction procedure in ENVI 4.7 (ENVI, 2006).

5.4 Spectral Analysis

For analyzing the extracted spectral data, four different sets of spectral features were used as classification input. This include the original five bands of RapidEye image, four bands of the RapidEye image excluding the red-edge band, NDVI derived from band 3 and band 5 of the image and, the red-edge adaptation of NDVI derived from band 4 and band 5 of the RapidEye image. The result from each of these classification, were put to comparison as described below:

Analysis 1: All bands vs. 4 bands (excluding red-edge)

Analysis 2: NDVI vs. NDVI-RE

Analysis 3: All bands + NDVI vs. All bands + NDVI-RE

Similarly, combination of all input variables i.e. all bands, NDVI and NDVI-RE were used for classification to test the sensitivity of each input variable to the classification result.

Two different machine learning algorithms: support vector machine (SVM) and random forest were used to test the strength of the results obtained from this study. SVM is a relatively new non-parametric algorithm for image classification. In its classical implementation, it uses two classes (e.g. presence/absence) of training samples within a multi-dimensional feature space to fit an optimal separating hyperplane (in each dimension, vector component is image gray-level) (Foody and Mathur, 2006). In this way, SVM tries to maximize the margin, which is the distance between the closest training samples, or support vectors, and the hyperplane itself. SVM consists of projecting vectors into a high dimensional feature space by means of a kernel trick, then fitting the optimal hyperplane that separates classes using an optimization function. Several kernels are used in the literature (Foody and Mathur, 2006).

On the other hand, Random Forest is a machine learning algorithm that employs a bagging (bootstrap aggregation) operation where a number of trees (*ntree*) are constructed based on a random subset of samples derived from the training data (Breiman, 2001). Each tree is independently grown to maximum size based on a bootstrap sample from the training data set without any pruning, and each node is split using the best among a subset of input variables (*mtry*) (Breiman, 2001). The multiple classification trees then vote by plurality on the correct classification (Breiman, 2001; Lawrence *et al.*, 2006). The ensemble classifies the data that are not in the trees (out-of-bag or OOB data) and by averaging the OOB error rates from all trees, the random forest algorithm gives an error rate called the OOB classification error for each input variable (Breiman, 2001). Therefore, as part of the classification process, the RF algorithm produces a measure of importance of each input variable by comparing how much the OOB error increases when a variable is removed, whilst all others are left unchanged (Benediktsson and Sveinsson, 2004). Both RF and SVM algorithm is easy to implement as only two parameters (*ntree* and *mtry* for RF) and (C and λ for SVM) are needed to be optimized (Breiman, 2001; Özçift, 2011). In this study, 70% of the dataset for each of the defoliation level was used for

training while the other 30% was used for validation in both algorithms. All analysis was done in R statistical software (R Development Core Team 2008).

5.5 Accuracy Assessment

In describing the results of this study, we performed accuracy assessment for each class using the quantity disagreement and allocation disagreement as proposed by Pontius Jr. and Millones (2011). A confusion matrix was constructed so as to compare the true class with the class assigned by the classifier and to calculate the overall accuracy as well as the class accuracies. In order to evaluate and discuss red-edge effects using the class and overall classification accuracies for the respective comparative feature sets of analysis 1, 2, 3, and 4, McNemar test with a 95% confidence interval was used. McNemar test is a non-parametric test that has been shown to be sensitive and precise in comparing two or more accuracy assessments (Manandhar $et\ al.$, 2009; Petropoulos $et\ al.$, 2012). The test is based on chi square (z^2) statistics, computed from two error matrics and given as:

$$(z^2) = (f_{12} - f_{21})^2 / (f_{12} + f_{21})$$

Where f_{12} denotes the number of cases that are wrongly classified by classifier 1 but correctly classified by classifier 2 and f_{21} denotes the number of cases that are correctly classified by classifier 1 and wrongly classified by classifier 2 (Manandhar $et\ al.$, 2009). The difference in accuracy between a pair of classifications is viewed as being statistically significant at a confidence of 95% if the calculated z-score in McNemar test is larger than 1.96 (Manandhar $et\ al.$, 2009).

5.6 Results

5.6.1 Classification using the RapidEye Five Bands Vs. Four Bands (Excluding Red-Edge)

Results from the first analysis (analysis 1) of both classification algorithms (Support Vector Machine and Random Forest) indicate that the overall accuracy were significantly higher when all the RapidEye 5 bands were used as opposed to when 4 bands were used, excluding the red-edge (**Table 5.1**). For Instance, the overall accuracy reaches 80% (SVM) and 79% (RF) when the red-edge channel is included as input features, while the overall accuracy was 61% (SVM) and

58% (RF) when the red edge channel was excluded. Hence, the overall accuracies increase about 19% and 21% for SVM and RF, respectively, resulting in an increase of 20% when averaged for both algorithms. However, as **Table 5.1** indicates, the sensitivity to the red-edge depends on the defoliation level being investigated. Comparing the two classification (E1 and E2) using McNemar test, it was observed that there was significant difference (McNemar z score = 2.416 and 2.612 for RF and SVM respectively) between the matrix produced from the classifications in both algorithms. With regards to class specifics, differences were observed in terms of accuracies especially for refoliating levels. For instance, the improvement for PD and UD ranges only from 4% -8% whereas the R level improved by 44-56%.

Table 5. 1 : Class accuracies and percentage deviations for Analysis 1 based on SVMand RF

	RF			SVM		
	AB	AB-RE	Dev. (%)	AB	AB-RE	Dev. (%)
PD	56	52	4	68	60	8
R	96	40	56	88	44	44
UD	84	80	4	84	80	4
Overall Accuracy (%)	79	58	21	80	61	19

  * AB = All bands, AB-RE= All bands excluding Red-edge

5.6.2 Classification using NDVI vs. NDVI-RE

Overall accuracies for the feature stacks investigated in Analysis 2, where the NDVI classification was compared with NDVI-RE classification are reported in **Table 5.2**. Again, both algorithms showed a rise in classification accuracy when the NDVI-RE was used instead of the NDVI. The result shows an increase in overall accuracy deviation of 23% and 32% for both RF and SVM, respectively. With respect to class-specific, outstanding accuracy increases were again observed for refoliating defoliation level as compared with other classes. For instance, the accuracy improved by 40% and 52% for RF and SVM respectively while the PD level increased by 20% and 16% and UD 36% and 10%. Similarly, the same trends as Analysis 1 were observed for the comparison of the two classifications. Relatively higher significant differences were observed for the two classifications as compared to Analysis 1 (McNemar z score = 4.32 and 3.56 for RF and SVM, respectively). It must also be noted that the NDVI-RE shows the highest accuracy values (84% and 83% for RF and SVM, respectively) in all the classification conducted in this study.

Table 5. 2: Class accuracies and percentage deviations for Analysis 2 based on SVM and RF

	RF			SVM		
	NDVI	NDVI-RE	Dev. (%)	NDVI	NDVI-RE	Dev. (%)
PD	52	72	20	60	76	16
R	52	92	40	44	96	52
UD	48	84	36	80	70	10
Overall Accuracy (%)	61	84	23	51	83	32

5.6.3 Classification using NDVI vs. NDVI-RE as Additional Features

Results from Analysis 3 (using NDVI or NDVI-RE as additional features) shows that the overall accuracy did improve when the NDVI-RE was added to the 5 bands of RapidEye image over the addition of NDVI in both algorithms. For example, there was an increase of 17% and 10% in the overall accuracy for RF and SVM, respectively (**Table 5.3**). Although, when comparing the effect of red-edge, Analysis 3 shows a slight increase in overall accuracy (80% and 81% for RF and SVM, respectively) over Analysis 1 (79% and 80% for RF and SVM, respectively), it was still lower than Analysis 2 (84% and 83% for RF and SVM, respectively). Similarly and just like the previous 2 analysis (1&2), significant difference was observed when the NDVI-RE was used as an additional feature as opposed to NDVI (McNemar z score = 2.52 and 2.38 for RF and SVM, respectively). Regarding class-specific results, **Table 5.3** illustrates that percent deviations remained higher for R (24% for RF and 18% for SVM) as compared to PD (12% for RF and 16% for SVM) and UD (16% for RF and 11% for SVM) as observed in the previous analysis.

Table 5. 3: Class accuracies and percentage deviations for Analysis 3 based on SVM and RF

	RF			SVM		
	AB+NDVI	AB+ NDVI-RE	Dev. (%)	AB+NDVI	AB+ NDVI-RE	Dev. (%)
PD	52	64	12	56	72	16
R	72	96	24	70	88	18
UD	64	80	16	73	84	11
Overall Accuracy (%)	63	80	17	71	81	10

* AB = All bands

5.6.4 Sensitivity Analysis using All Input Variables

Classification from the combination of all the inputs (RapidEye bands plus NDVI plus NDVI-RE) did not improve the accuracies over previous analysis (SVM: 68% and RF: 72%). However,

we further examined the importance of the red edge parameters used in this study (red-edge band and NDVI-RE) by ranking their mean decrease in accuracy based on RF classification. Results from the analysis further confirmed that NDVI-RE and red-edge band of RapidEye image outperformed the other input variable (NDVI, Blue, Green, Red, and Near Infrared bands of RapidEye image) in classifying the defoliation levels (**Fig 5.3**).
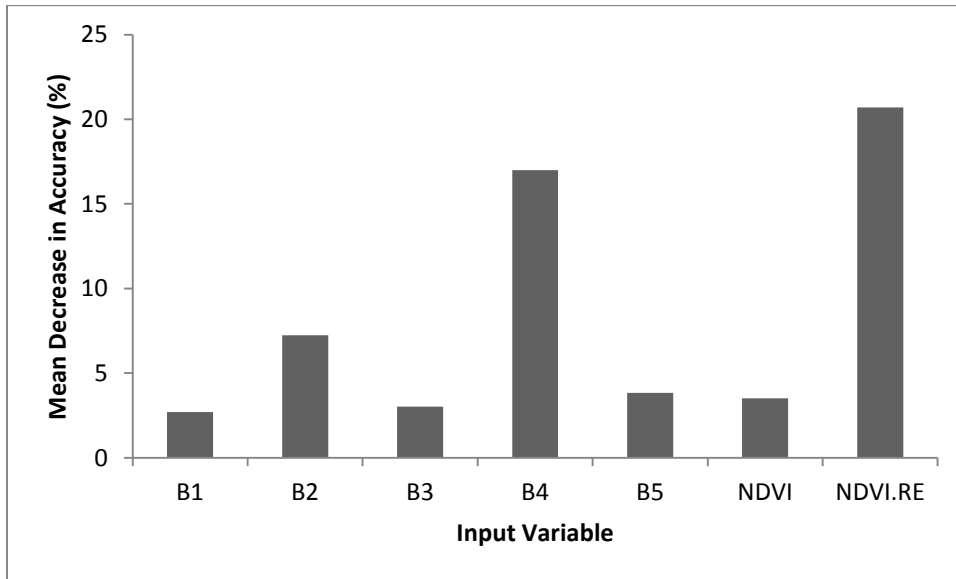
Figure 5. 3: Ranking of 5 bands of RapidEye image, NDVI, and NDVI-RE using mean decrease accuracy for classifying levels of defoliation.

5.7 Discussion

One of the primary challenges in classifying insect defoliation levels using multispectral data is in the ability to identify significant differences in health levels visually and to spectrally characterize these differences. Previous studies have investigated the ability of red-edge channels from hyperspectral measurement to discriminate insect defoliation levels in vegetation with reasonable success (Adelabu *et al.*, 2013; Coops *et al.*, 2003; Lawrence and Labus, 2003). However, the cost and availability of hyperspectral data in Sub-Saharan Africa is still a challenge. The solution therefore lies in identifying cheap and readily available sensors with strategic bands that can identify the defoliation even when symptoms are not visible to the human eye (Adelabu *et al.*, 2012; Zarco-Tejada *et al.*, 2000). This study therefore, presents valuable evidence of application and potential of the red edge channel of RapidEye multispectral image information to accurately classify insect defoliation levels (Partly defoliated, undefoliated

and refoliating). The subtle visual difference in the canopy colour levels of UD and PD challenged us to consider more detailed and specific spectral-bands for classifying the three levels of mopane worm defoliation.

Results from this study have shown the potential of the red-edge channel of RapidEye image in effectively dealing with the spectral variability that exists between the insect defoliation levels. For instance, when all the 5 bands of RapidEye imagery were used for classification, the overall accuracies, increases about 19% and 21% for SVM and RF, respectively as opposed to when the red edge channel was excluded. Furthermore, the results of the McNemar test over all the analyses shows that there are significant differences when the red-edge feature is included as opposed to when it is excluded.

The results of this study are comparable to the results obtained in previous defoliation classification using hyperspectral datasets (Adelabu *et al.*, 2013; Carter and Knapp, 2001; Coops *et al.*, 2003; Lawrence and Labus, 2003; Somers *et al.*, 2010). For example Lawrence and Labus (2003) used sub-canopy spatial resolution hyperspectral imagery for differentiating Douglas-fir trees attacked by the Douglas-fir beetle. They found the region between 700 and 750 nm proved to be the best for discriminating the affected and unaffected trees. Recently, ground-based hyperspectral data was used for insect defoliation levels classification and obtained an overall accuracy of 82.42% observing that 7 of the 8 most important bands were found in the red-edge region (Adelabu *et al.*, 2013). The study done by Carter and Knapp (2001) observed that the optical response to stress near 700 nm, as well as corresponding changes in reflectance that occur in leaves can be explained by insect induced defoliation. Furthermore, Coops *et al.* (2003) classified three levels of Dothistroma needle blight defoliation with an accuracy of over 70% using Hyperspectral airborne remote sensing imagery (CASI-2). The authors identified the red-edge region of the CASI image for classifying the tree levels of defoliation. Finally, Somers *et al.* (2010) classified different levels of defoliation using EO-1 Hyperion data and they also observed that the red-edge region is highly important for defoliation from different source. What is therefore unique in this study, is that the red edge-edge was derived from a multispectral sensors compared to all the above mentioned studies that used hyperspectral data.

Previous studies have explained that plants under defoliation display a decrease in canopy reflectance in the lower portion of the near infrared, a reduced absorption in the chlorophyll

active region, and subsequently a shift in the red edge (Adelabu *et al.*, 2014; Cho and Skidmore, 2006). For instance, (Schuster *et al.*, 2012) observed that when conducting land-use classification, classes with vegetation tend to increase accuracies with inclusion of red-edge channel. Although chlorophyll content was not measured in this study, it is plausible to say that wavelengths near 700 nm are the most sensitive to small changes in chlorophyll concentration (Carter and Knapp, 2001; Gitelson *et al.*, 2006; Schuster *et al.*, 2012). Sims and Gamon (2002), for example, found that reflectance near 705 nm was sensitive to chlorophyll changes irrespective of leaf structure and developmental stage.

With regards to the vegetation indices, the present study shows that NDVI-RE is highly capable of discriminating the health levels of insect defoliation. Result showed an increase in overall accuracy of 23% and 32% for both RF and SVM, respectively when the NDVI-RE was used as compared to NDVI. Similarly, there was an increase of 17% and 10% in the overall accuracy for RF and SVM, respectively when the NDVI-RE was added as an additional feature rather than NDVI. The sensitivity analysis also showed that NDVI-RE and red edge band performed better than the other inputs. This is because red edge indices are constructed with bands sensitive to the chlorophyll content and internal structure of the leaf, and therefore have proven to be closely related to foliage biomass quantity, growth and developmental stage and health status of the plant (Gitelson and Merzlyak, 1994; Sims and Gamon, 2002) .

5.8 Conclusions

Based on our results, the application of a satellite platform that provides red-edge band information appears promising for the development of insect defoliation early warning system and health level classification for forest management. The focus of much of this work was to use relatively rapid and low cost new generation multispectral images for the classification of forest health levels induced by insect defoliation as part of large scale monitoring by forest management agencies (Bennett and Tkacz, 2008). The result from the present study is critical for early warning, detection, and monitoring of insect defoliated forest at large scale level in sub-Sahara Africa even with minimal cost. Furthermore, the result can be used for the development of models that can analyse the relationship between spectral reflectance values and biophysical variables such as Nitrogen and to ultimately predict future health levels of forest induced by

insect defoliation. Although, the study did not aim at choosing the best classifier (RF or SVM) for insect defoliation classification, it is however recommended that the strength and dependability of these classifiers should be carried out for validation purpose.

Acknowledgements

CHAPTER SIX

Testing the Reliability and Stability of RF Classifier

This chapter is based on:

S. **Adelabu**, O. Mutanga, and E. Adam, (In Review), "Testing the reliability and stability of the internal accuracy assessment of Random Forest when compared with independent dataset" *International Journal of Remote Sensing*.

ABSTRACT

In the present study, the strength and reliability of internal accuracy estimate built in random forest (RF) ensemble classifier was evaluated. Specifically, we compared the reliability of the internal validation methods of RF with independent datasets of different splitting options for defoliation classification. Furthermore, we set out to statistically validate the best independent split option for image classification using RF and multispectral imagery (RapidEye). Results show that the internal accuracy measure yields comparable results with those derived from an independent test data set. More important, it was observed that the errors produced by the internal validation methods of RF were relatively stable as statistically shown by the lower confidence interval obtained as compared to the independent test data. Results also showed that the 70%-30% split option had the lowest mean standard errors (0.2351) and hence highest accuracy when compared to the other split options. The study confirms the reliability and stability of the internal bootstrapping estimate of accuracy built within the Random forest algorithm.

Keywords: accuracy, random forest, confidence level, validation

6.1 Introduction

A bootstrap aggregation algorithm, random forest (RF) (Breiman, 2001) was built to hold multiple classification trees that are constructed based on random subset of samples derived from the training data (Adam *et al.*, 2009) (Further description of RF is provided in the methodology section). However, there have been debates in remote sensing on whether to use the RF internal validation technique only or to use an independent test data set for assessing the accuracy of the RF predictive model. Previous studies have shown that in other to evaluate the prediction of an algorithm, the use of a large independent test dataset that was not used initially for training is recommended (Adam and Mutanga, 2010; Congalton and Green, 1999). However, when the data are limited some types of cross-validation techniques are usually carried out (Hawkins *et al.*, 2003). In the RF algorithm, the out of bag (OOB) estimate of error is considered to be such a type of cross-validation technique (Breiman, 2001). The internal validation proponents argue around the assertion that the internal validation in RF is relatively stable and reliable. While previous studies such as Lawrence *et al.* (2006), Adam *et al.* (2011), Breiman and Cutler (2012) and Waske *et al.* (2007) have supported this claim, others such as Bylander (2002), Ruiz-Gazen and Villa (2007), Statnikov *et al.* (2008), Strobl and Zeileis (2008) and Menze *et al.* (2009) have contested the validity of this claim. For instance, Lawrence *et al.* (2006) used hyperspectral imagery to map invasive species and found out that internal validating assessment techniques of RF and independent dataset which was used for independent accuracy assessments, were within on average 3% apart, and most estimates were less than 1% apart.

Furthermore, in an experiment to support the reliability of the internal error of RF for accuracy assessment, Mutanga and Adam (2011) compared independent test dataset with the internal accuracy assessment of RF in discriminating papyrus and its co-existing species. They found that there was no significant difference between the independent test error and the internal error estimates. Similar results were obtained by Breiman and Cutler (2012) where they observed the same trend in the independent test errors and RF internal error estimates for classification. Nevertheless, those that opposed the reliability of the internal error estimates based their arguments on RF internal validation technique to overestimate generalization error. Most importantly and as observed by Ruiz-Gazen and Villa (2007) and substantiated by Strobl and Zeileis (2008) and Mitchell (2011), RF internal validation is not stable when performed severally even with constant *ntree* and *mtry*.

There are several drawbacks with regards to those who support and oppose the reliability of the RF internal validation for remote sensing classification. For instance, the opponents of its reliability have only used dataset other than remote sensing in advocating their claims. Whereas, the proponents of this claim that have used remote sensing data was only done by splitting the initial dataset into two subsets: training dataset (used for classification) and testing dataset (used for independent validation). The question that arises as a result of this will therefore be, (i) will the internal validation in RF be stable and reliable when using remote sensing data? and (ii) will the error produced by the internal validation in RF be the same when compared with independent dataset derived from different splitting options?. Furthermore, there is no agreement on the percentage of dataset to be used for training and test data set to independently validate the internal validation in RF for remote sensing classification. Liu *et al.* (2013) observed that the splitting quality is the performance bottleneck of RF construction. In this regard and for better image classification accuracies, it is therefore necessary to focus more effort on the splitting strategies for RF construction. Therefore, the present study aims at examining the reliability and stability of RF internal validation with independent dataset when split over different options for insect defoliation mapping using multispectral RapidEye Image.

6.2 Methodology

6.2.1 Overview
The process of achieving the objectives set out in this study involves three main parts: pre-processing of data, classification and statistical analysis. **Fig 6.1** describes the connection between these components and shows the way the three different validation techniques were employed.
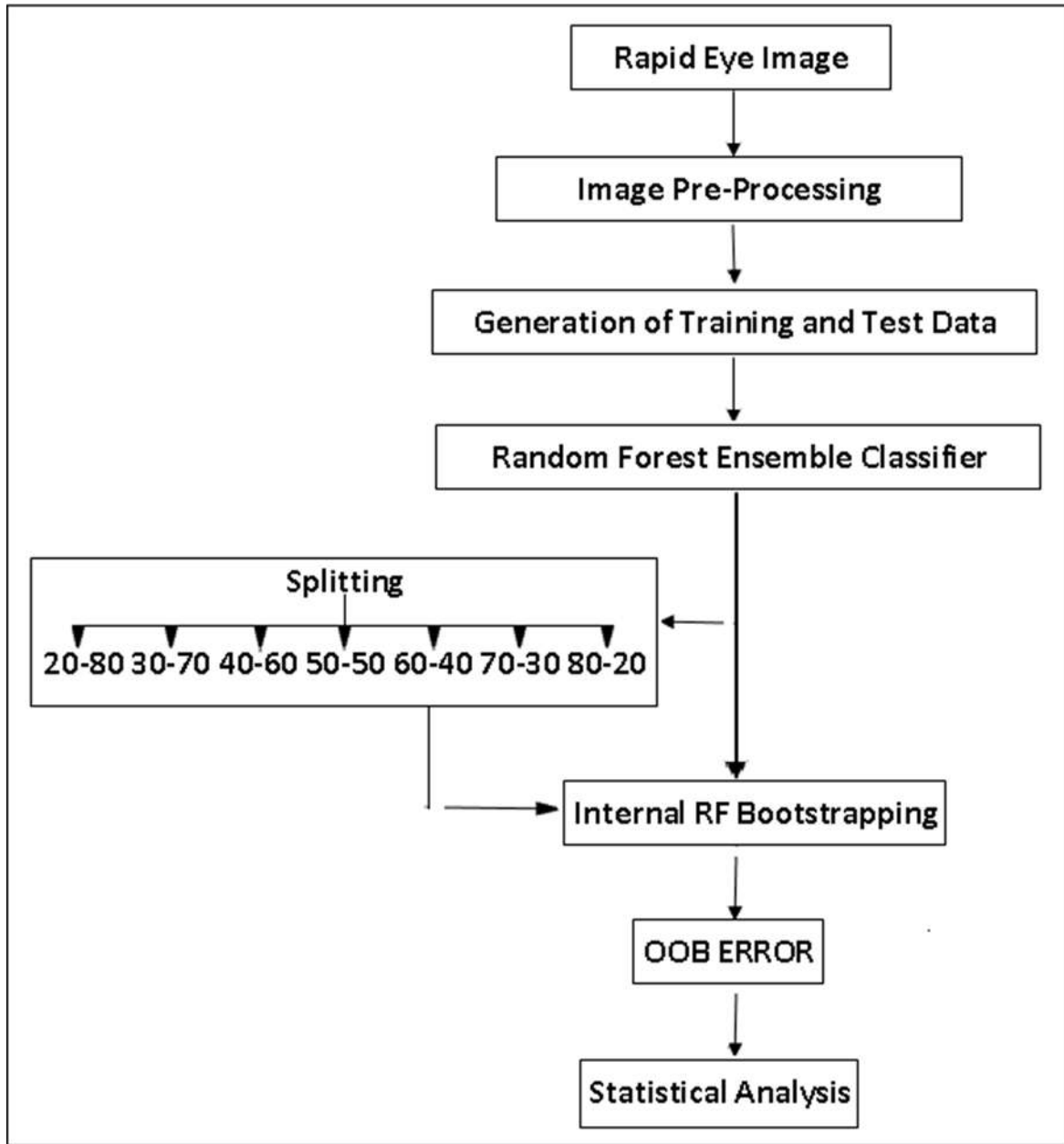
Figure 6. 1: Schematic overview of the validation techniques and the processed used in this study.

6.2.2 Remotely Sensed Data Acquisition and Pre-Processing

The RapidEye imagery used in this study was acquired on 25 January 2012. The RapidEye data set is composed of 5-band multispectral imagery with 5 m ground sampling distance (GSD). The five bands include: blue (440 – 510 nm), green (520 – 590 nm), red (630 – 685 nm), red-edge (690 – 730 nm) and near-infrared (760 – 850 nm). The imagery over the study area contained 0% cloud cover, with a relatively clear atmosphere. Previous experiments performed by Naughton *et al.* (2011) verified that the image registration was within a single pixel, hence no further

geometric processing was applied. The image was atmospherically corrected using the quick atmospheric correction procedure in ENVI 4.7 (ENVI, 2006). The field data acquisition was conducted during the summer month of January 2012, coinciding with the image acquisition date. Three different defoliation levels namely: Undefoliated (UD), Partly defoliated (PD), and Refoliating (R) were identified within the study area in Palapye, Botswana. The area was chosen because heavy defoliation has been reported in the area over the past few years (Sebego *et al.*, 2008). Intensive field work was conducted with the help of ecological experts to identify areas that are associated with constant insect (mopane worm) defoliation. Visual observation shows that the time series of defoliation in mopane woodland are in terms of its canopy impacts. While UD represents time before defoliation, PD and R represent during and after defoliation respectively. In the healthy non-impacted Mopane canopies (UD), leaves are distinctive, consisting of two large triangular leathery leaflets, sometimes likened to butterfly wings. Although the PD canopies appear green, they may be visually indistinguishable from UD canopies except for the part eaten by the worms. Leaves become pale green gradually in PD state leading to total defoliation. However, after total defoliation takes place, new shoots (refoliating canopy) also spring out usually after 2 weeks.

It is important to note that while other factors such as drought and wind can also contribute to defoliation, it is difficult to spectrally differentiate between insect and other environmental factors induced defoliation. However, Mopane worm defoliation is different from drought or wind induced defoliation in that defoliation follows good-healthy growth patterns that would otherwise not occur if there was a severe drought or wind. Moreover, insect defoliation occurs in patches (within the healthy forests) but drought or wind would impact on all vegetation. Furthermore, insect defoliation in Mopane woodland mainly occurs in the summer (November-April; rain season) after which defoliated plants come to leaf for the second time. The normal drought or wind induced defoliation for deciduous trees occurs in winter therefore making it easier to detect insect defoliation during summer in Mopane woodland. Nevertheless, we assumed that in our study drought is constant over a wide area and by restricting our measurements to the period immediately after defoliation, we increased the chances of capturing the worm effect.

An experiment was designed using 255 sample random points that were generated from an existing land cover map of the study area using Hawth's Analysis Tool (HAT) in ArcGIS 9.3

resulting in 85 sample points for each defoliation levels (Sebego and Arnberg, 2002). A handheld *Garmin eTrex30* GPS (± 3m accuracy) was then used to navigate to the 85 field measurements for each defoliation levels. Thereafter, defoliation plots were created around the centred point in which 80% of the trees are said to belong to a particular defoliation level. The defoliation plot was defined as covering 20 m × 20 m resulting in 84 defoliation plots for each defoliation level. Thereafter, the defoliation plots were then overlaid on the true colour composite RapidEye image for extraction of the pixels spectra (5 m × 5 m) using Environment for Visualizing Images (ENVI) software (ENVI, 2006). The metadata such as the site description (coordinates, altitude and land cover class) and general weather conditions were also recorded.

6.2.3 Classification

6.2.3.1 Random Forest Classifier

Random forest (RF) machine learning algorithm was employed in this study to classify the three levels of defoliation (UD, PD, R). RF is a machine learning algorithm that employs a bagging (bootstrap aggregation) operation where a number of trees (*ntree*) are constructed based on a random subset of samples derived from the training data. Each tree is independently grown to maximum size based on a bootstrap sample from the training data set without any pruning, and each node is split using the best among a subset of input variables (*mtry*) (Breiman, 2001). The multiple classification trees then vote by plurality on the correct classification (Adelabu *et al.*, 2014; Lawrence *et al.*, 2006). In the forest building process, when bootstrap sample set is drawn by sampling with replacement for each tree, about $1/3^{rd}$ of original instances are left out. This set of instances is called OOB (out-of-bag) data. Each tree has its own OOB data set which is used for error estimation of individual tree in the forest and by averaging the OOB error rates from all trees, the random forest algorithm gives an error rate called the OOB classification error for each input variable (Breiman, 2001).

According to Kulkarni and Sinha (2013) the OOB error (OOBE) of Random Forest is given as,

$$OOBE = P_{x,y} (mg(X,Y)) < 0$$

Where mg(X,Y) is Margin function. The Margin function measures the extent to which the average number of votes at (X, Y) for the right class exceeds the average vote for any other class and it is directly proportional to confidence (P) in the classification. Where X is the predictor vector and Y is the classification.

In the present study, two different sets of data were input into the RF model. Firstly we used the whole dataset (n = 255) for different defoliation classes into the RF classifier with the aim of allowing RF to classify the defoliation levels and produce internal validation results (OOB). The second set of input is the different split options in which a percentage of the whole data was used for classification while the other subset was set aside for independent validation. Each of the splitting options was iterated 100 times while the internal bootstrapping of RF was also repeated 100 times to ensure that all observations were selected in the validation sub-sample. For each iteration, accuracy was assessed using standard error matrices (OOB error rate). The goal of this study was not to produce the best classification results, but rather to evaluate the reliability of internal validation and to better assess classification error and bias within the sample set while applying RF classification algorithm. Consequently, the OOB error rate of the test data of the split options was only compared to the OOB error rate obtained when the whole dataset was used. A brief description of the applied strategies is given in **Table 6.1**. Further description of RF can be found in (Breiman, 2001) and (Horning, 2010). All analysis was done in R statistical software (R Development Core Team 2008).

Table 6. 1: Overview of the size of training and test sample subsets for the validation techniques used in this study.

Methods	Options	Training Sample size	Testing Sample Size	Repetition
	20-80	17	67	100
	30-70	25	59	100
	40-60	34	50	100
Split Sample	50-50	42	42	100
	60-40	50	34	100
	70-30	59	25	100
	80-20	67	17	100
Internal RF bootstrapping		*	*	100

* Internal RF bootstrapping split its Training and Test dataset randomly

6.2.4 Statistical Analysis

For each option and in all the techniques used in this study, standard error rates were recorded after the classification. The mean and the confidence level at 95% significance level were calculated and recorded. The confidence level was the preferred statistical measure of differentiating between the errors produced from each of the technique used because it has been proved to be the most conventional way to assess the degree of certainty in data analysis

(Mutanga and Skidmore, 2004a). A one way analysis of variance (ANOVA) and Turkey post-hoc tests were executed between errors produced from the split options and internal bootstrapping of RF. This was done to ascertain if the errors produced from the classification by the split options and internal validation in RF are comparable.

6.3 Results

The summary statistics of the standard errors from the classification iterations of each validation techniques are given in **Table 6.2**. Results, shows that while the lowest standard error rates (0.098) was observed in the 80% - 20% split option, the mean error rates after repeating the classification 100 times was significantly higher (0.275) and the standard deviation was even the highest (0.105) of all the split options used in this study.

Table 6. 2: Descriptive statistics of standard errors produced after classification using the validation techniques.

Method	Option (%)	Minimum	Maximum	Mean	Std. deviation
	20-80	0.101	0.467	0.321	0.100
	30-70	0.103	0.487	0.208	0.082
	40-60	0.132	0.454	0.261	0.078
Split Sample	50-50	0.135	0.437	0.259	0.077
	60-40	0.129	0.406	0.248	0.074
	70-30	0.132	0.355	0.235	0.048
	80-20	0.098	0.496	0.275	0.105
Internal RF bootstrapping		0.207	0.257	0.232	0.010

The conclusion from the ANOVA test for the split options is that they are significantly different ($p < 0.001$). However, One-way ANOVA only shows that there is significant difference between the split options but it does not show which pairs are different. We therefore, used a Turkey post-hoc test to determine if there are any differences between the different split options (**Table 6.3**).

Table 6. 3: P values of the Turkey post Hoc test performed to compare the errors produced from different splitting options

	20-80	30-70	40-60	50-50	60-40	70-30	80-20	RF Validating
20-80	1	**< 0.0001**	**0.002**	**0.005**	**< 0.0001**	**< 0.0001**	**0.034**	**< 0.0001**
30-70	**< 0.0001**	1	**0.000**	**< 0.0001**	**0.011**	**0.034**	**< 0.0001**	**< 0.0001**
40-60	**0.002**	**0.000**	1	1.000	0.975	0.877	**0.010**	0.806
50-50	**0.005**	**< 0.0001**	1.000	1	0.902	0.722	**0.010**	0.877
60-40	**< 0.0001**	**0.011**	0.975	0.902	1	1.000	**0.006**	0.902
70-30	**< 0.0001**	**0.034**	0.877	0.722	1.000	1	**0.036**	0.975
80-20	**0.034**	**< 0.0001**	**0.010**	**0.010**	**0.006**	**0.036**	1	**0.002**
RF Validating	**< 0.0001**	**< 0.0001**	0.806	0.877	0.902	0.975	**0.002**	1

* Highlighted values are significantly different.

Results from **Table 6.3** showed that from the P values of the Turkey post-hoc test, there were significant differences between 20% -80% and other split options. The same trend was observed from the 30% -70% and 80% -20% split options. There was however no significant difference at 95% significance level between and among 40% -60%, 50% -50%, 60% -40% and 70% -30% split options indicating that any of these could be used for validating classification. However since we still determined to choose the best from the sample options, we derived the confidence level at 95% significance level as shown in **Fig 6.2**.

Holdout 20-80 — Mean = 0.3244, CL = 0.0236

Holdout 30-70 — Mean = 0.2132, CL = 0.0197

Holdout 40-60 — Mean = 0.2606, CL = 0.0153

Holdout 50-50 — Mean = 0.2594, CL = 0.0151

Holdout 60-40 — Mean = 0.2479, CL = 0.01447

Holdout 70-30 — Mean = 0.2351, CL = 0.0094
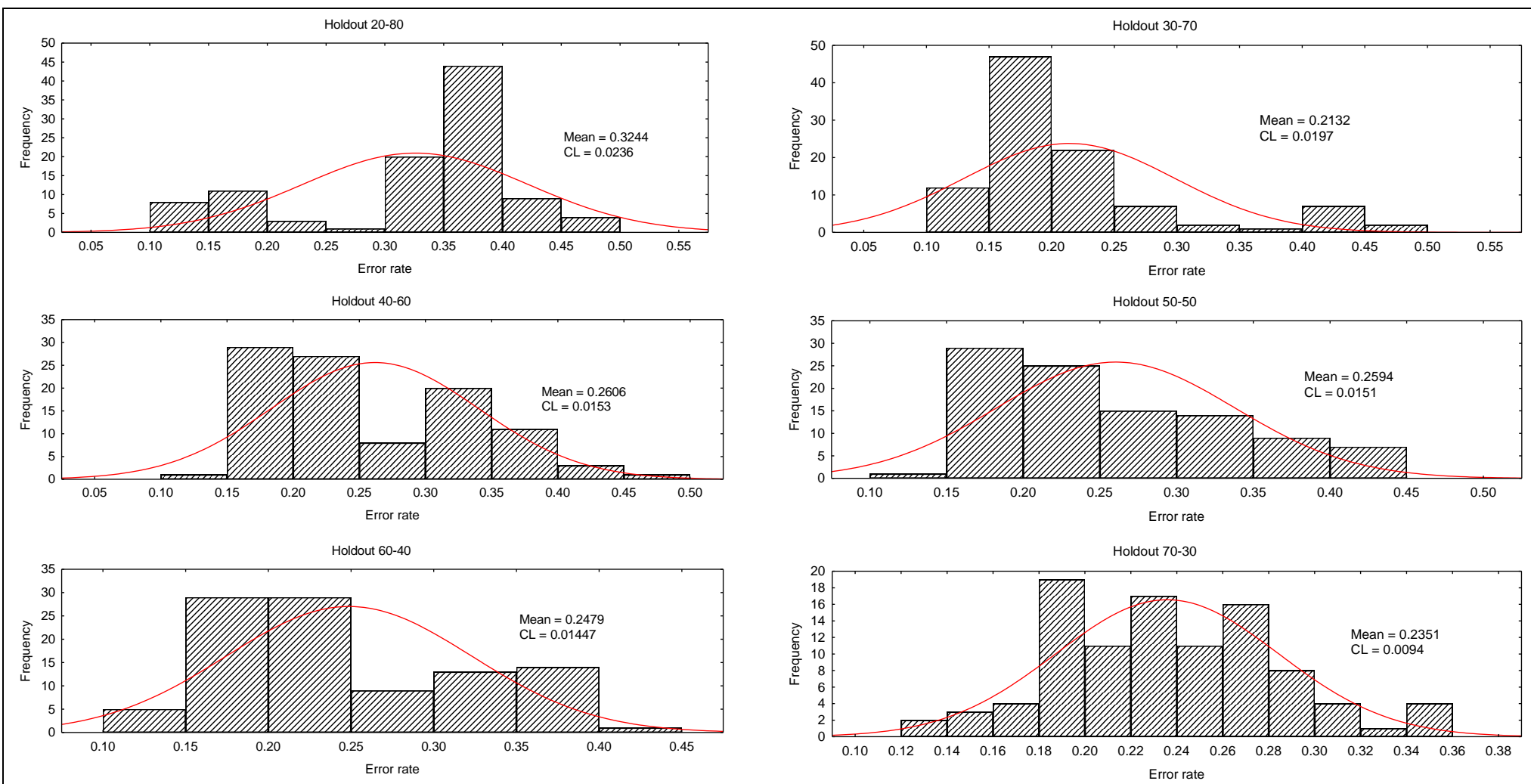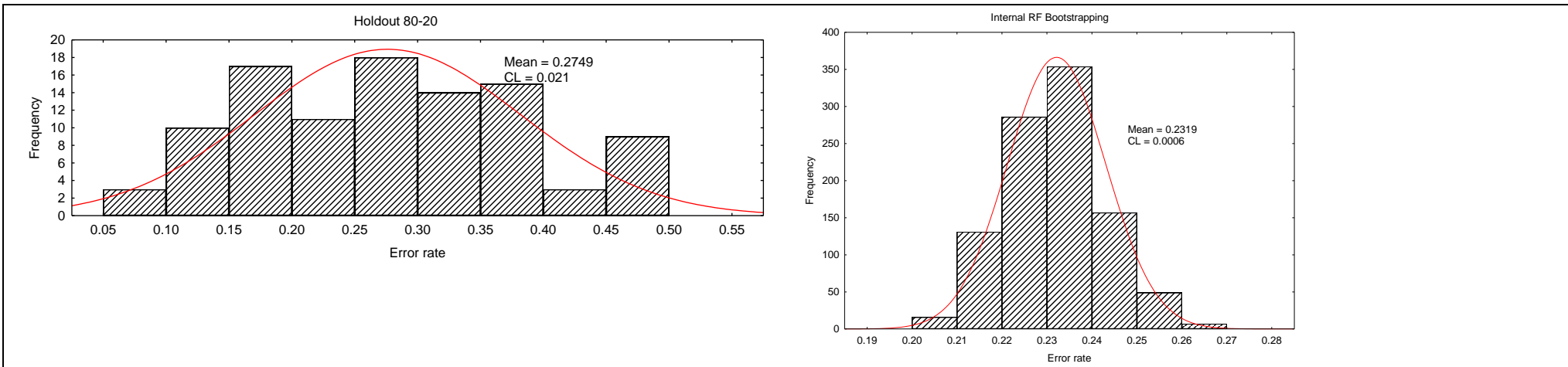
Figure 6. 2: Histogram showing the frequency of standard error produced after classification for different holdout options

A comparison of the confidence level (CL) from all the splitting options showed that a much narrower distribution (CL = 0.0094) was observed from the standard errors produced from 70% - 30% holdout option than other options. Moreover, the 70% -30% has the lowest mean standard errors (0.2351) and hence highest accuracy when compared with the other three (40% -60%, 50% -50%, 60% -40%). Therefore, we conclude that the 70% -30% split option gives better and more robust output. We compared this holdout option (70% -30%) with the internal bootstrapping of RF with the aim of testing the reliability of the latter and choosing the best for validating defoliation levels classification (**Table 6.4**).

Table 6. 4: P values of the Turkey post Hoc test performed to compare the errors produced from between the best splitting option and internal bootstrapping of RF

	70-30	internal bootstrapping of RF
70-30	1	0.086
Internal RF bootstrapping	0.086	1

* Highlighted values are significantly different.

Results from **Table 6.4** showed no significant difference (95% significance level) between internal bootstrapping of RF and the independent 70% -30% split option. This indicate that the internal bootstrapping of RF performed just as the best independent split option with even lower CL (0.0006) and mean standard errors (0.2319).

6.4 Discussion

Collection of ground reference data after insect defoliation for classification is an expensive and complex task especially in semi-arid African environment thereby making allocation of training and test samples for classification a mammoth task. Moreover, most of the remote sensing data in semi-arid area exhibit large heterogeneities due to different factors such as time of collection, soil types, sensors or viewing/illumination geometrics. Therefore validation techniques are important for generalization and independent validity of remote sensing classification (Richter *et al.*, 2012).

The present study aimed at testing the reliability of the bootstrapping estimates of accuracy of RF with respect to insect defoliation levels classification. Additionally, we sought to choose the best split option for independently validating remote sensing classification of defoliation levels.

The results of the present study indicate that extremes split i.e. low training - high test and high training - low test data is not a reliable accuracy assessment measure. While previous studies have randomly preferred the 60%-40% and 70%-30% split options (Adam and Mutanga, 2010; Adelabu *et al.*, 2014), the reasons behind this have not yet been discussed. Our studies have shown that when the number of training samples is relatively small with respect to the number of test samples, the standard error produced becomes larger and hence accuracy becomes lower. Moreover, when the classification was repeated 100 times, the classification errors of the extremes splits were not stable as can be observed from their confidence level at 95% significance level. The reason for this is because of the well-known problem of dimensionality resulting in the risk of over-fitting of the training data that can lead to poor generalization capabilities of the classifier (Dieterle, 2003; Richter *et al.*, 2012). From all holdout options tested in this study, 70% -30% split option yielded the lowest standard error rates. The study has shown that, similar to other studies (Rodriguez-Galiano *et al.*, 2012; Stumpf and Kerle, 2011) a the use of split lower than 60% -40% or greater than 70% -30% for defoliation classification is not recommended.

The evaluation of the stability and reliability of the internal estimates of accuracy of RF in this study supported the assertion that these estimates are relatively stable and reliable. Of the two validation techniques tested in this study, the internal estimates of accuracy produced the lowest mean error (0.2319) and hence the highest accuracy. Furthermore, internal estimates of accuracy gave the lowest CL indicating that the errors were relative stable. This supports the assertion that, with internal estimates of accuracy in RF, it is not necessary to have a separate accuracy assessment data set.

As observed in this study, the internal estimate of accuracy of RF was able to provide reasonable lower error rates using small datasets for classification in this study. The internally generated out-of-bag accuracy assessments were shown to be reliable, potentially obviating the need to collect separate assessment data. The implementation of RF in the R statistics package (as the randomForest package) makes it available to analysts free of charge. We believe, based on our relatively lower error rates obtained, ease of use, low cost, and possibly no need for independent accuracy assessment, it is worth considering RF for remote sensing defoliation classification problems in the future. In this aspect, our result is consistent with that obtained by Lawrence *et*

al. (2006) who found that the accuracy assessment using OOB is nearly identical to an independent accuracy assessment.

6.5 Conclusions

The following conclusions were obtained in this study:

1. The use of split lower than 60% -40% or greater than 70% -30% for defoliation classification yield unstable and therefore unreliable results.
2. As observed from this study, the 70%-30% independent split option is recommended for defoliation classification using RF machine learning algorithm
3. The internal bootstrapping estimates of accuracy in RF are relatively stable and hence reliable
4. With the combination of efficiency and accuracy, along with very useful analytical tools, the Random Forest classifier should be considered very desirable for defoliation levels classification of remote sensing especially in areas with limited sample size.

Although, the results from this study requires further investigation for other defoliators and environments, the study intends to stimulate the remote sensing community (especially those involved in defoliation detection, classification and monitoring) towards agreeing on the exploitation of validation techniques for mapping defoliation. A general agreement on this will help to better compare results of defoliation classification from different studies and other parts of the world. Furthermore, there is urgent need for testing the validation of these techniques in other terrain as increasing measuring and modeling activities carried out in different areas are becoming apparent and in view of the upcoming satellite earth observations missions.

CHAPTER SEVEN

Estimating Biophysical Variable of Insect Defoliated Canopies Using Radiative Transfer Models

This chapter is based on:

S. **Adelabu**, O. Mutanga, M.A. Cho, (In Preparation), "The use of radiative transfer models to estimate LAI of insect defoliated canopies".

ABSTRACT

In the present study, LAI was estimated for different defoliation levels based on data simulated from RapidEye image. We sought to examine if LAI could be used for discriminating different levels of insect defoliation. The PROSAILH radiative transfer model was inverted with canopy spectral reflectance extracted from RapidEye imagery by means of look-up-table (LUT). Additionally, estimated LAI was compared with bands of RapidEye and selected indices. Results show that the estimated LAI was in the range of those reported in the literature. The NDVI-RE index was the most strongly correlated with the estimated LAI as compared with other variables (RapidEye bands and NDVI). Our results further confirm the potential of model inversion for estimating vegetation biophysical parameters of relative homogenous forest using new generation multispectral imagery.

7.1 Introduction

The alliance between remote sensing and biophysical indicators is valuable for studies on detecting, mapping and monitoring defoliation process in forests. Among the many biophysical indicators, leaf area index (LAI) is of prime importance because it is a vital variable for explaining canopy primary production and can be used to infer processes such as photosynthesis, transpiration, evapotranspiration and estimate net primary production (NPP) of terrestrial ecosystems (Yao *et al.*, 2008). LAI is defined as one-half the total surface area of leaves per unit ground area (Adelabu *et al.*, 2012). Relationship between biophysical characteristics such as LAI of canopies and reflectance has previously been studied with great success (Asner, 1998; Atzberger, 2004; Richter *et al.*, 2012; Ustin *et al.*, 2009). Estimating LAI from remote sensing is therefore derived from the knowledge of this relationship (empirical) or radiative transfer (RT) inversion model (Rullan-Silva *et al.*, 2013).

The empirical method involves the use of statistical analysis to derive the correlation between observed LAI and canopy reflectance. However, empirical methods are usually reliant on the site, the type of sampling used and remote sensing sensor employed (Colombo *et al.*, 2003; Meroni *et al.*, 2004). Nevertheless, scientists have carried out studies in the last decade to advance the estimation of LAI from the relationship between observed LAI and canopy reflectance using different remote sensing datasets and techniques. This included the development of vegetation indices (Darvishzadeh *et al.*, 2009), angle indices (Khanna *et al.*, 2007) and red edge position (Cho *et al.*, 2008). Yet, estimating LAI using the empirical method still face a lot of challenges with regards to the collection of datasets needed for model adjustment resulting in high cost and rigorous use of labour (Richter *et al.*, 2011).

However, the RT model approach assumes that the changes in canopy reflectance and soil background could be accurately described using physical laws through the interaction and transfer of radiation inside the canopy. This assertion offers the reason behind the relationship between canopy biochemical and biophysical relationship and the canopy reflectance (Houborg *et al.*, 2007). For RT models, LAI is usually an input variable for describing the relationship between canopy and solar light (Bacour *et al.*, 2006). Nevertheless, in order to use RT models for estimating LAI, their inversion is highly recommended (Kimes *et al.*, 2000). To this end, researchers have developed various techniques such as artificial neural networks (Schlerf and

Atzberger, 2006; Weiss and Baret, 1999), support vector machines regression (Durbha *et al.*, 2007), numerical optimization methods (Jacquemoud *et al.*, 1995; Meroni *et al.*, 2004) and look up table (LUT) (Combal *et al.*, 2002; Weiss *et al.*, 2000) approaches for inverting the RT models. Several studies in the past have used RT inversion models for different remote sensing applications with great success (Asner *et al.*, 1998; Houborg *et al.*, 2007; Jacquemoud *et al.*, 2009; Kimes *et al.*, 2000).

Nevertheless, some shortcomings of RT inversion models have been reported in the literature. For instance, Richter *et al.* (2011) observed that RT inversion model requires extensive parameterization as well as high computational demand. The authors further observed that some RT inversion models may be too simplistic to handle composite canopies especially those in the semi-arid areas which are often affected by soil background. Combal *et al.* (2002) further observed that when performing RT inversion model, the combination of input parameters may produce similar reflectance values, causing high variation and uncertainties in the derived characteristics of the vegetation. This challenge could however be solved by creating some control on the inversion process and by having initial understanding of the model parameters (Combal *et al.*, 2002). Recently, however, attention has shifted to the use of RT inversion approach on simulated data as data collection becomes more and more difficult to collect especially in developing world (Weiss *et al.*, 2000).

Although, advances have been made in the use of hyperspectral (Adelabu *et al.*, 2014) and multispectral (Adelabu *et al.*, 2013) dataset for discriminating insect defoliation levels, the knowledge of biophysical variables such as LAI may help to understand the influence of the insects on the canopy at each of the defoliation levels. Furthermore, better understanding of the relationship of LAI with canopy reflectance at different defoliation levels could help resource managers to know the impacts, the vulnerability and the extent of defoliation and hence suggest possible management practices that will enable efficient and sustainable use of the resources emanating from woodland in Africa. This will further support scientific knowledge of environmental management practices and food security in Africa and other sites around the world sensitive to global changes.

Therefore, the aim of the present study is to estimate LAI of canopies at different insect defoliation level using a well-established canopy reflectance models, parameterized to represent

a wide range of canopy characteristics. Specifically, we set out to use simulated data derived from RapidEye image to estimate LAI by inverting the canopy radiative transfer model PROSAIL H (Jacquemoud *et al.*, 2009) at different defoliation levels. This allowed us to synthetically create artificial LAI measurements that otherwise would have been difficult and expensive to obtain under experimental or field conditions (Cho *et al.*, 2008). Additionally, we established the relationship between RapidEye bands, some selected vegetation indices and the estimated LAI. This allowed us to understand the role of each of the bands and indices in estimating LAI at different defoliation levels.

7.2 Materials and Methods

7.2.1 Study Area and Field Campaign

The study is located in and around the Palapye/Tswapong axis of Central Region of Botswana (Longitude 27°00' E- 27°33' and Latitudes 22°23' - 22°52' S). It is situated along the main north-south highway and is about 230 km north of Gaborone, the capital city of Botswana (Sebego *et al.*, 2008). The study area is covered mainly by mopane (*Colophospermum mopane*) bushveld because mopane tree is found in all its growth forms and is locally monospecific. The value of mopane woodland in Botswana alone has been estimated at US $3.3 million per annum, of which approximately 40% goes to producers who are primarily poor rural women (Sebego *et al.*, 2008). However, a number of factors have of late, brought the sustainability of the mopane woodland resources into question. One of such factors is the difficulty in monitoring defoliation process within the woodland. The canopies of *Colophospermum mopane* are usually defoliated by edible caterpillars popularly called mopane worms (Adelabu *et al.*, 2012). This has resulted not only in the depletion of woodland resources in most rural areas but also low vitality and productivity of the woodland (Makhado *et al.*, 2012). The depletion of the worms has been reported in the study area over the past few years with heavy defoliation occurring (Sebego *et al.*, 2008).

Three primary categories of canopy impact ranging from non-impacted undefoliated plants (UD) to partly defoliated plants (PD) and finally refoliating plants after severe defoliation (R) at different sites of the study area were identified during the field campaign. The UD represents time before defoliation, PD and R represent during and after defoliation respectively. In the

healthy non-impacted Mopane canopies (UD), leaves are distinctive, consisting of two large triangular leathery leaflets, sometimes likened to butterfly wings (Adelabu *et al.*, 2014)..

Field campaign was done during the January 2012 defoliation window. Random Points that were generated from an existing land cover map of the study area using the Hawth's Analysis Tool (HAT) in ArcGIS 9.3 (Sebego and Arnberg, 2002). Thereafter, the respective points were navigated to using a handheld *Garmin eTrex30* GPS. Defoliation plots were created around the centered point in which 80% of the trees are said to belong to a particular defoliation level. The defoliation plot was defined as covering 20 m × 20 m resulting in 80-84 defoliation plots for each defoliation level. The defoliation plots were then used to create training area and then overlaid on the true colour composite RapidEye image for extraction of the pixels spectra (5 m × 5 m) using Environment for Visualizing Images (ENVI) software (ENVI, 2006). The metadata such as the site description (coordinates, altitude and land cover class) and general weather conditions were also recorded.

7.2.2 Remote Sensing Imagery: Acquisition and Processing

High multispectral RapidEye image employed in this study was acquired on 25 January 2012. RapidEye images come with 5 multispectral bands with 5 m ground sampling distance (GSD). The five bands include: blue (440 – 510 nm), green (520 – 590 nm), red (630 – 685 nm), red-edge (690 – 730 nm) and near-infrared (760 – 850 nm). The imagery over the study area contained 0% cloud cover, with a relatively clear atmosphere. Results of previous research by Naughton *et al.* (2011) verified that the image registration was within a single pixel, hence no further geometric processing was applied. The image was atmospherically corrected using the quick atmospheric correction procedure in ENVI 4.7 (ENVI, 2006). Two popular vegetation indices: Normalized difference vegetation index (NDVI) derived from band 3 and band 5 and the red-edge adaptation of NDVI derived from band 4 and band 5 of the RapidEye image were used in this study. Because of the difference in transmittance and reflectance of infrared and near-infrared energy between photosynthetic and non-photosynthetic plant material, there has been a great deal of interest in relating LAI to various spectral measurements that could be gathered by earth orbiting satellites. One of the more promising ratio-based indices is the normalized difference vegetation index (NDVI).

NDVI (Tucker (1979)) has previously been shown to be sensitive to LAI for defoliated areas and the relationships between LAI and satellite-based measurements such as NDVI have been used to drive productivity models applied over large domains (Carter and Knapp, 2001; Gupta *et al.*, 2003). It has however been noted that, NDVI is usually influenced by soil status, and this effect is especially greatest in areas with low vegetation cover such as semi-arid area (Fan *et al.*, 2007). Therefore as an alternative to NDVI, we tested the relationship between the NDVI-RE and LAI with the aim of checking if either could be used for estimating LAI in an effective, fast and nondestructive way in semi-arid woodland.

7.2.3 The PROSAIL H Radiative Transfer model

The combination of PROSPECT (Jacquemoud and Baret, 1990) and SAILH (Kuusk, 1991; Verhoef, 1984) transfer model were used to simulate synthetic reflectance spectra extracted from the RapidEye imagery at different defoliation levels. The combination (PROSAILH) have been widely used and validated, produces realistic results of bidirectional reflectance spectra for different canopies as reported by several researches (Cho *et al.*, 2008; Jacquemoud *et al.*, 2009). For this reason it was preferred over other radiative transfer models where the canopy is explained in a complex way. The PROSPECT model calculates reflectance and hemispherical transmittance as a function of the leaf structural parameter N (unitless); the leaf chlorophyll content a+ b LCC (μgcm^{-2}); the dry matter leaf dry matter content, Cm (gcm^{-2}) and the equivalent water thickness Cw (cm) (Jacquemoud *et al.*, 2009). The transmittance and reflectance estimated in PROSAIL are input into the SAILH model.

The SAILH model is based on the turbid assumption and describes the canopy structure in a fairly simple way (Kuusk, 1991). In this study we used SAILH model to simulate bi-directional canopy reflectance (p) because it requires few input variables and also has a predictive power similar to more elaborated reflectance models (Jacquemoud *et al.*, 2009). According to Cho *et al.* (2008), SAILH assumes the canopy to be a homogenous semi-finite medium with Lambertian leaves characterized by their reflectance and transmittance spectra (pleaf, tleaf). Apart from the reflectance and transmittance, SAILH requires eight input parameters to produce canopy bi-directional reflectance. These are hotspot parameter (HSP; unitless); average leaf angle (ALA; degrees), leaf area index (LAI; unitless); sensor viewing angle ($\varnothing_0$; degrees); azimuth angle ($\varnothing_a$; degrees); fraction of diffuse incoming solar radiation (skyl); soil background reflectance (rsl) and

solar zenith angle ($\emptyset_z$). To account for changes in soil brightness, we employed soil brightness parameter (scale) as described by Atzberger (2004). However, as observed by Cho *et al.* (2008), the soil brightness is not the only parameter to be considered for RT models, the changes in the spectral shape due to variations in the chemical composition of the soil (e.g. soil carbon) needs to be considered too. Therefore when the two models are combined, 12 input parameters are to be specified. Four of the 12 parameters were fixed (sun zenith angle, azimuth angle sensor viewing angle and fraction of incoming solar radiation). The other eight parameters were randomly generated (**Table 7.1**). For all simulations, a nadir looking sensor was assumed ($\emptyset_0 = 0^o$). The fraction of diffuse illumination (skyl) was fixed to 0.1 for each of the bands of RapidEye imagery. A solar zenith angle ($\emptyset_z$) of 45° was assumed.

The PROSAILH model was inverted using a look-up table (LUT). Previous studies have shown that LUT is the simplest method of solving the inversion of RTM model because they permit a global search while showing less unexpected behavior when the spectra characteristics of the targets are not well represented by the modeled spectra (Darvishzadeh *et al.*, 2008; Schlerf and Atzberger, 2006). In this study, 3000 parameter combinations were randomly used to build LUT and hence used in forward estimation of LAI from the PROSAILH model. In the present study, the input parameters as described above and reflectance values extracted from each defoliation level served as input variables into the PROSAILH model for estimating LAI at each of the defoliation levels.

Table 7. 1: Range and Distribution of Input parameters used to establish simulated canopy reflectance for use in LUT.

Model parameter	Abbreviation	Units	**Range	Distribution
Leaf Area Index	LAI	Unitless	0-10	Uniform
Average Leaf Angle	ALA	Degrees	30-80	Uniform
Hot spot parameter	HSP	Unitless	0.1 ± 0.01	Normal
Leaf Chlorophyll content	LCC	μgcm^{-2}	20-80	Uniform
Leaf dry matter content *	C_m	gcm^{-2}	$1.25xC_w$	Uniform
Leaf water content	C_w	Cm	0.004-0.044	Uniform
Leaf structure parameter	N	Unitless	2 ± 0.2	Normal
Soil brightness	scale	Unitless	1 ± 0.14	Normal
Carbon content	Cc	gcm^{-3}	0-6	Uniform

* Combal *et al.* (2002) proposal that C_m is varied proportional to C_w was used. ** In the case were distribution is normal, range indicates mean ± standard deviation.

7.2.4 Data Analysis

The estimated LAI was tested for significant difference between the different defoliation levels using a one way ANOVA and Turkey Post Hoc. This was done to ascertain if the estimated LAI from the defoliation levels are comparable. Furthermore, we tested the interaction between the estimated LAI with NDVI and NDVI-RE as well as each of the 5 bands using Pearson correlation co-efficient which is widely used as a measure of the degree of linear dependence between two variables. This was done to ascertain the relationship between the bands, NDVI or NDVI-RE and LAI.

7.3 Results and Discussion

Table 7.2 shows the summary statistics of the estimated LAI at different defoliation levels. Results shows that mean values (PD; 4.62, UD; 5.12, R; 4.14) are within the LAI values obtained by other simulation studies (Colombo *et al.*, 2003; Weiss *et al.*, 2000). We further observed that the ranges of values are within those reported for different levels of severity of defoliated canopies in other part of the world (Asner *et al.*, 2003a). The current study shows that there was 19% decrease in the LAI between the UD and R levels. As a proxy for canopy density, changes in LAI as a result of stresses such as insect defoliation will be significance in that it determines the productivity of forest. Moreover, it is known that a plant is under stress when there is a

change in the health condition of the plant foliage (Rullan-Silva *et al.*, 2013). Hence, the healthier the vegetation, the higher the vigour and hence the higher the LAI (Adelabu *et al.*, 2012). Hence, a forest that is highly defoliated is expected to have low LAI as observed in this study.

Table 7. 2: Descriptive statistics of the estimated LAI from the Inverted PROSAIL H RT model

Defoliation Levels	Mean	Maximum	Minimum	Std. Dev.
PD	4.622328	5.020619	4.233532	0.140263
UD	5.123606	5.969437	4.494172	0.164477
R	4.147604	4.647977	3.433718	0.170675

Further analysis was however conducted to determine if there were any significant differences between the LAI at different defoliation levels. It was concluded that highly significantly different ($p < 0.001$) exist between the LAI estimated at different defoliation levels. However, this could not help in determine if there exist significant difference between one individual class and another. Consequently, we used a Turkey post-hoc test to determine if there are any differences between the different defoliation levels (**Table 7. 3**).

Table 7. 3: P values of the Turkey post Hoc test performed to determine if there are any significant differences in the LAI estimates in between the different defoliation classes.

	PD	UD	R
PD	1	<0.001	<0.001
UD	<0.001	1	<0.0001
R	<0.001	<0.0001	1

PD = Partly Defoliated, UD = Undefoliated, R = Refoliating

The results from **Table 7.3** confirmed the earlier results from **Table 7. 2** that significant differences exist between the classes in terms of LAI estimates. It was however observed from **Table 7. 3** that it was easier to discriminate the undefoliated (UD) canopies from the refoliating canopies (R) as compared to partly defoliated (PD) using the LAI estimates. Previous study by Adelabu *et al.* (2014) and Adelabu *et al.* (2013) were able to discriminate the defoliation levels at the same study area using hyperspectral and multispectral datasets respectively. However in both studies, they were unable to show if there are any significant differences between the defoliation classes i.e. to what extent a class differs from the other. Collantes *et al.* (1999) observed that

knowledge of how each levels of defoliation differs from each other could actually provide the background to which monitoring and management could be done.

Although LAI was not measured in the field in this study, we compared our LAI derived from simulated data with those in the literature. For instance, Veenendaal *et al.* (2008) found out that the LAI for mopane trees range between $3.5m^2m^{-2}$(short) and $5.6m^2m^{-2}$ (tall). Similar results for mopane specie were obtained by Mantlana (2008) and Greenberg *et al.* (2003). For other species, Weiskittel and Maguire (2007) using destructive sampling method observed that the mean canopy LAI of Douglas fir in Oregon USA was $5.45m^2m^{-2}$. Similarly, Wythers *et al.* (2003) used process oriented model to estimate LAI in a heterogeneous deciduous woody forest in Harvard forest. Their result shows that LAI ranges from $4.3m^2m^{-2}$ to $4.7m^2m^{-2}$ depending on the species. Moreover, Asner *et al.* (2003b) observed that LAI for deciduous trees range between $3.3m^2m^{-2}$ to $5.6m^2m^{-2}$. With regards to estimation of LAI from remote sensing data, previous studies were also similar to the present study. Gonsamo (2011) used CIR colour image to estimate LAI with an average value of $4.5m^2m^{-2}$. Similarly, Asner *et al.* (2003b) analyzed collections of various researches on the use of MODIS LAI and observed that for deciduous forest, the mean LAI $5.5m^2m^{-2}$. Fassnacht *et al.* (1997) observed that for a combination of pine, oak and aspen tree species in a Central Wisconsin forest, the LAI estimated from Landsat TM ranges between $2.5m^2m^{-2}$ and $5.0m^2m^{-2}$.

Other studies have estimated LAI using Lidar (Lange and Solberg, 2008), and Hyperspectral imagery (Lange and Solberg, 2008; Richter *et al.*, 2011) with similar or little difference to the present result. The limitations of these previous studies are that attentions were focused on estimating LAI of groups or individual canopies rather than the health status of the canopies. Only few studies have related reflectance with observed LAI with the aim of evaluating different health levels in forest. For instance, Hall *et al.* (2003b) used Landsat 7 ETM+ to discriminate different levels of insect defoliation in aspen forest and found a very strong relationship (r = 0.84) between estimated and observed LAI. Some other studies used simulated data from hyperspectral imagery to estimate LAI (Darvishzadeh *et al.*, 2009; Darvishzadeh *et al.*, 2008). This may be expensive and not readily available for researchers in sub-Saharan Africa. Our present study has shown that radiative transfer model such as PROSAIL with an implemented LUT algorithm can be used for estimating LAI from relatively cheap and readily available

RapidEye especially in areas where it may be difficult to validate LAI with ground based measurement.

Furthermore, a Pearson's correlation coefficient was run in order to evaluate the relationship between the RapidEye bands, derived vegetation indices and the estimated LAI **(Fig 7.1)**. The correlation results show that RapidEye bands and the vegetation indices are positively correlated with the estimated LAI for the different defoliation levels although; the level of relationship differs from one variable to another. For instance, the NDVI-RE yielded the strongest relationship (Pearson correlation coefficient of 0.76) while the red band (band 2) yielded a weak relationship (Pearson correlation coefficient of 0.05). The result is in line with other studies that have conducted the relationship between LAI and reflectance from different regions of electromagnetic spectrum. For instance, Vina *et al.* (2011) observed a strong relationship between red-edge indices and LAI. Furthermore, Jensen (2005) observed that under stress induced condition, changes in LAI in the visible region is firstly noted near 700nm wavelength (red edge region) where there is blue shift of the red region. This observation therefore, highlights the importance of the red-edge region in estimating LAI as a proxy for detecting insect defoliation in forest environments.

A — Correlation coff. = 0.24

B — Correlation Coff. = 0.53

C — Correlation Coeff. = 0.05

D — Correlation Coeff. = 0.56
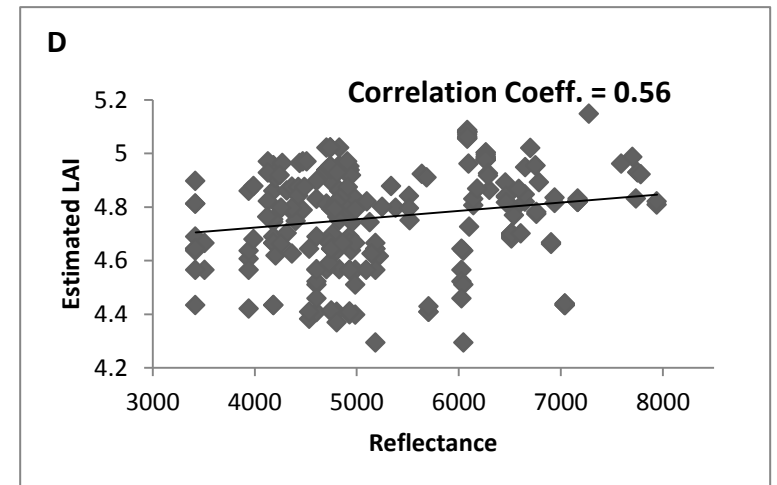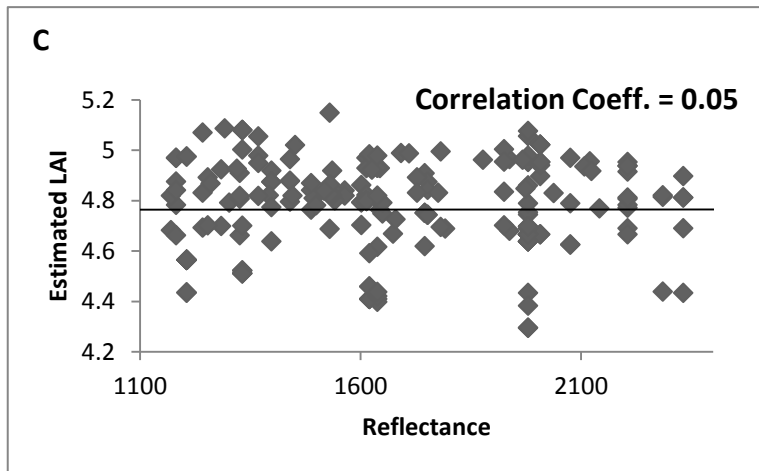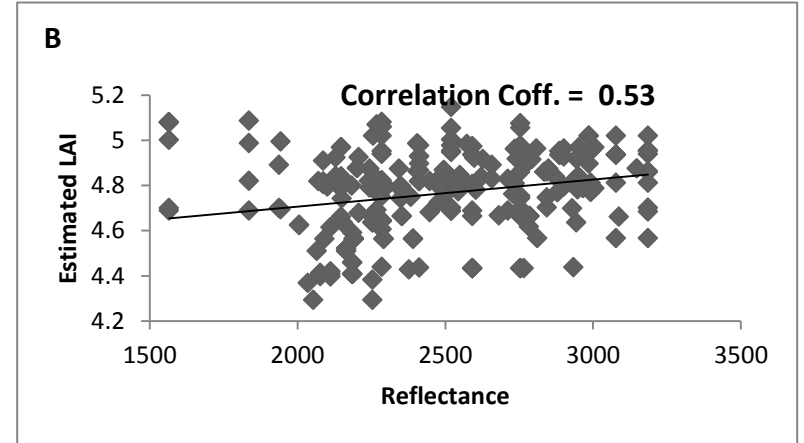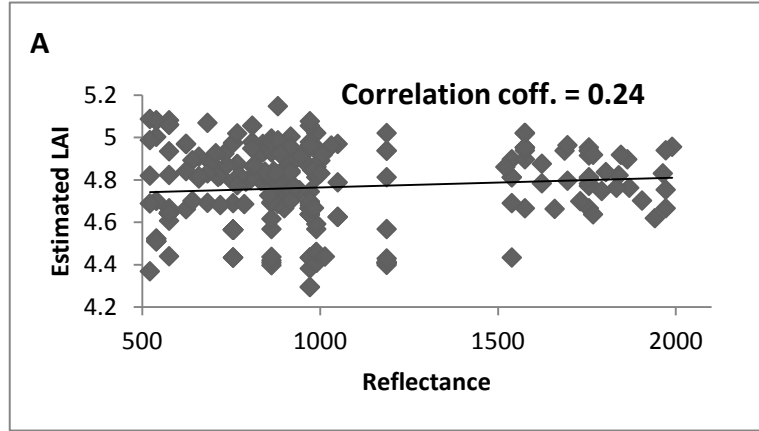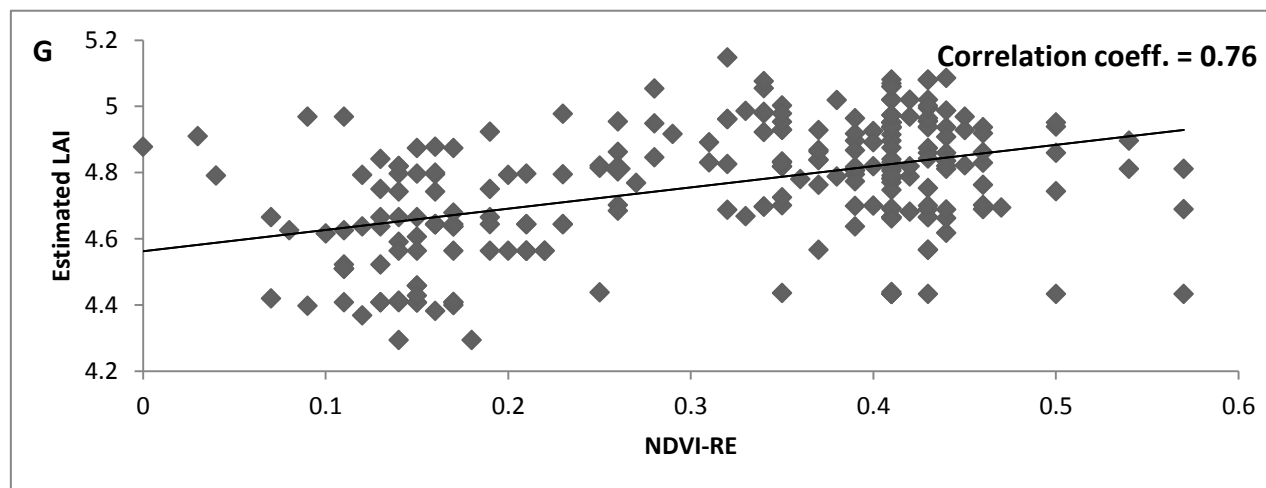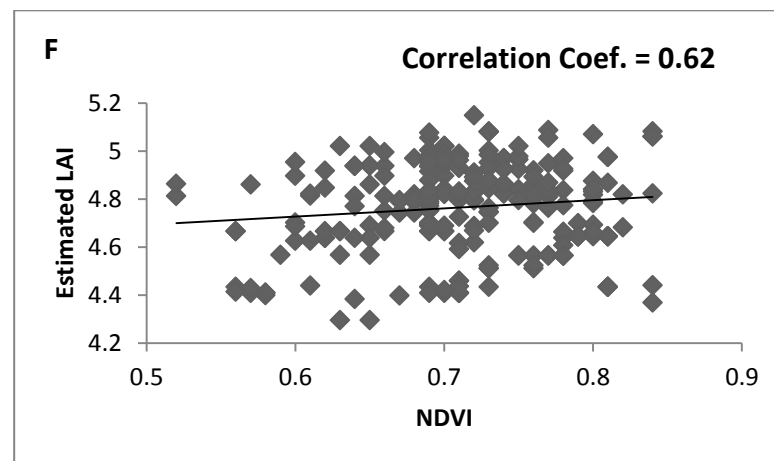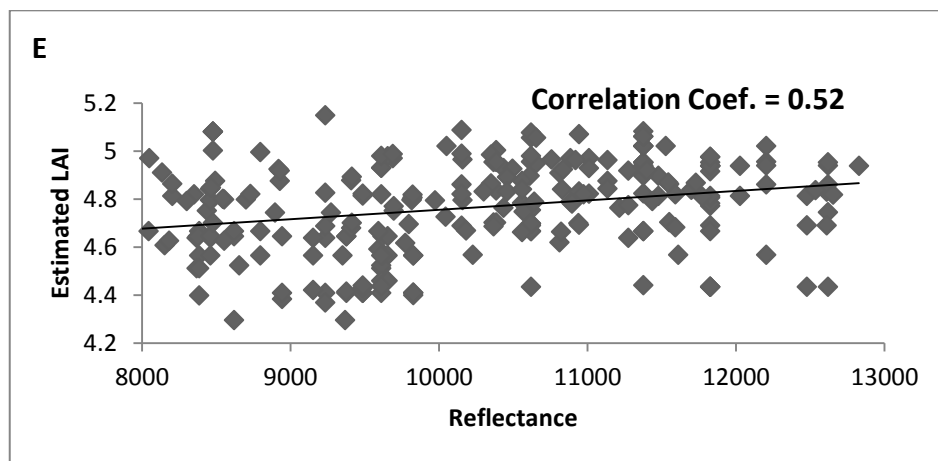
Figure 7. 1: Pearson correlation coefficient of LAI and different bands of RapidEye and selected vegetation indices, A=Band 1, B=Band 2, C=Band 3, D=Band 4, E=Band 5, F=NDVI, G=NDVI-RE

.

7.4 Conclusions

Leaf area index was estimated for different defoliation levels based on simulated data generated from RapidEye image. The results show that the estimated LAI was in the range of those reported in the literature. Similarly, the study observed that NDVI-RE index was strongly correlated with the estimated LAI when compared with other variables (RapidEye bands and NDVI). Although, result from the present study indicates the possibility of estimating LAI using the inverted radiative transfer model and RapidEye imagery extracted reflectance in an homogeneous forest canopies, it will be important to test the validity of this approach in heterogeneous forest canopies. As observed by Richter *et al.* (2011), heterogeneous forests may need higher complexity of models such as 3-dimensional radiative transfer modeling. Regarding future use of PROSAIL H for estimating LAI of insect defoliated levels using new generational multispectral data, we recommend that prior knowledge of the biophysical variable in the specific forest should be known. This will help to reduce the loads of parameterization in RTM models., addressing the relationship between change in LAI at each defoliation levels will no doubt increase the potential for early and effective operational monitoring of insect defoliation patterns.

CHAPTER EIGHT

SYNTHESIS AND RECOMMENDATION

Remote Sensing of insect defoliation in mopane woodland: A synthesis

8.1 Introduction

What benefit is remote sensing for insect defoliation in mopane woodland? Defoliation in mopane woodland is currently viewed as one of the serious impacts in the depletion of the vitality and quality of the resources emanating from the woodland (Adelabu *et al.*, 2012; Ditlhogo *et al.*, 1996). However, while the depletion of worms derived from mopane woodland have been reported in different areas, none of these depletions have been attributed to the impacts of the worms on the vitality and productivity of their host. Defoliation process as a result of the worms if not well managed can in the long run lead to the extinction of the tree and hence the worms within the region. In an effort to minimize the potential loss of mopane worms in mopane woodland of Southern Africa therefore, an integrated management strategy is needed combining detection, mapping and monitoring methods.

However, detecting and monitoring defoliation levels in mopane woodland is challenging using conventional methods as it is very expensive and labour intensive because of time required for sampling, and most importantly the need to collect data immediately before, during and after an extreme event (de Beurs and Townsend, 2008). Hence, the need for methods that will bring into consideration the financial implications, real time detection and advanced techniques for monitoring insect defoliation. Remote sensing however brings an advantage of being able to meet data requirements, and has proven to be a cost-effective alternative to ground data acquisition.

Nevertheless, three challenges may make it difficult to monitor mopane defoliation with remote sensing. First, mopane worm-mopane tree interactions are dynamic and periods where defoliation can be detected are often short (Hrabar *et al.*, 2009b). Secondly, most mopane woodland contain other tree species that are always present during the mopane worm defoliation making it very difficult to distinguish defoliation of mopane from reflectance of other end members (Vogelmann *et al.*, 1993). Finally, a challenge may occur when an image at the peak of an outbreak is being analyzed; it is unclear if an effect (changes in chlorosis, nutrient content or tree vitality) or a determining factor of the insect population drives the satellite image classification (Somers *et al.*, 2010). The solution to these challenges will therefore be to develop cutting edge techniques that can focus on detecting and discriminating levels of insect defoliation in mopane woodland.

Hence, the objectives of this study were:

1. Review the different approaches including the remote sensing platforms and techniques that have been used for assessing insect defoliation and its implications for detecting and monitoring mopane worm defoliation of mopane woodland highlighting their strengths and weakness.

2. Discriminate *Colophospermum mopane* and its co-existing species in a semi-arid forest using spectral information provided by the RapidEye sensor applying advanced classification algorithms.

3. Discriminate the levels of change in forest canopy cover detectable after insect defoliation based on ground based hyperspectral measurements in mopane woodland.

4. Explore and evaluate the benefit of the RapidEye red edge channel for discriminating different levels of insect defoliation in mopane woodland.

5. Test the reliability and robustness of the internal accuracy estimate in random forest (RF) ensemble classifier in discriminating different levels of insect defoliation in mopane woodland.

6. Quantify the impact of insect defoliation on the leaf area index of mopane canopies by estimating leaf area index at different defoliation levels using radiative transfer model.

8.2 Spectral discrimination of Mopane from its Co-existing Species

In order to detect levels of defoliation in mopane woodland, it was necessary to discriminate mopane tree species from other co-existing tree species. The study evaluated the ability of relatively cheap and readily available new multispectral RapidEye imagery for discriminating mopane tree species and its co-existing species using machine leaning algorithms with limited training samples (Chapter 3). Four dominant tree species that co-exist with mopane species namely: *Grewia bicolar* GB; *Dichrostachys cinerea* DC; *Acacia erubiscens* AE; *Acacia tortilis* AT were discriminated from each other using support vector machine (SVM) and random forest (RF). Specifically, the following questions were addressed

➢ Can *Colophospermum mopane* (CM) and its co-existing species be separated using the RapidEye image with strategically positioned spectral bands?

➢ Does the additional red-edge band of RapidEye improve the classification accuracy significantly compared to the 4 standard bands?

➢ Which of the classification algorithms (RF and SVM) is better for detailed tree species classification in areas with limited training samples based on classification accuracies?

The vegetation map of RF and SVM classifications are presented in **Fig 8.1**. From the maps, it is very clear that using both methods of classification (RF and SVM), CM species could be discriminated from its co-existing species using RapidEye images with overall accuracies of 88.75% and 85% for SVM and RF classifications respectively. With regards to class pairs, the study observed that it is easier to discriminate *Colophospermum mopane* from *Acacia erubiscens* and *Grewia bicolar* GB compared with *Dichrostachys cinerea* DC and *Acacia tortilis* AT (**Table 8.1**).
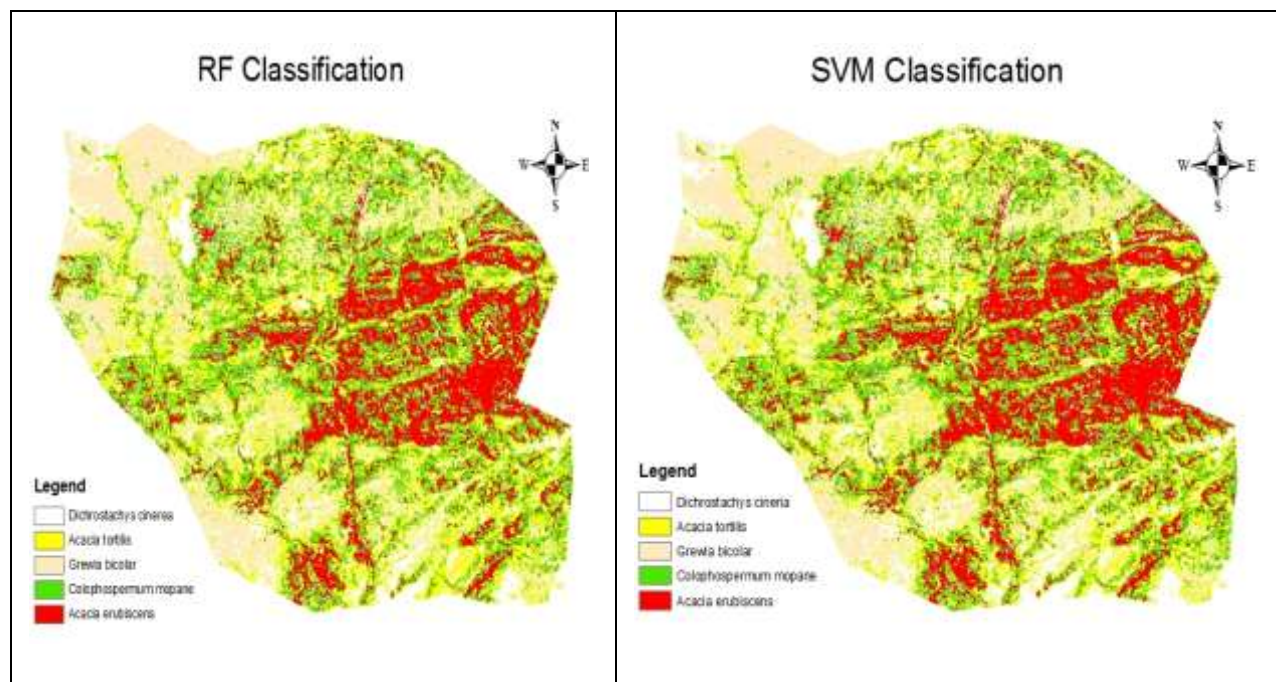
Figure 8. 1: Maps showing the classification of the 5 tree species using RF and SVM machine learning algorithm.

Table 8. 1: Comparison of confusion matrix obtained after the classification of *Colophospermum mopane* and its co-existing species from both the SVM and RF. The confusion matrix includes overall accuracy (OA) and class accuracy.

	SVM					RF				
	AE	AT	CM	DC	GB	AE	AT	CM	DC	GB
AE	15	1	0	0	0	15	0	1	0	0
AT	0	11	1	3	1	0	10	3	3	0
CM	0	0	16	0	0	0	0	16	0	0
DC	1	0	1	14	0	1	0	2	13	0
GB	0	0	1	0	15	0	0	1	1	14
CA (%)	93.75	91.67	84.21	82.35	93.75	93.75	90.91	72.73	76.47	100
OA= 88.75%						**OA= 85%**				

The study also demonstrated that the new red-edge band in the RapidEye sensor has potential for classifying tree species in semi-arid environment when integrated with other standard bands with results showing in both methods, when the red-edge band is excluded, the accuracies decreased by ± 8%. Similarly, we observed that where there are limited training samples, SVM performs better than RF. The results from this study provide the basis for future powerful algorithms that can be used to discriminate among tree species especially in semi-arid environment.

8.3 Hyperspectral Discrimination of Defoliation Levels in Mopane Woodland

Having been able to discriminate mopane species from its co-existing species, the capability of ground based hyperspectral dataset to discriminate levels of insect defoliation in mopane trees was evaluated in this study. Hyperspectral dataset brings an advantage of spectral details at different levels of defoliation. Moreover, hyperspectral dataset facilitates detailed spectral measurement of reflectance related to biochemical and biophysical attributes of plants, which are associated with its structure, physiology and phenology, and therefore with its health status, a mammoth task that could not be done accurately with multi-spectral sensors (Cho and Skidmore, 2006; Mutanga *et al.*, 2004; Schlerf and Atzberger, 2012). Furthermore, there is mounting evidence that hyperspectral data has the capability, not only to assess defoliation, but to detect early signs of defoliation even before visual symptoms are apparent due to its high spectral bands (Ismail *et al.*, 2008). To date and to the best of our knowledge, no study has been carried out using ground based hyperspectral dataset to discriminate different levels of insect induced defoliation. To this end, canopy spectral measurements were taken from three levels of defoliation identified in the field: Undefoliated (UD), Partly defoliated (PD) and Refoliating plants (R) using ASD FieldSpec Handheld 2^{TM} (Chapter 4).

A pre-filtering approach (ANOVA) was compared with random forest independent variable selector in selecting the significant wavelengths for classification. Furthermore, a backward feature elimination method was used to select optimal wavelengths for discriminating the different levels of defoliation in mopane woodland. Results show that optimal wavelengths located at 707nm, 710nm, 711nm, 712nm, 713nm, 714nm, 727nm and 1066nm were able to discriminate between the three levels of defoliation. The results further show that there was no significant difference in the overall accuracy of classification when random forest variable selector was used 82.42% (Kappa = 0.64) and the pre-filtering approach (ANOVA) 81.21% (Kappa = 0.68) used before building the classification (**Table 8.2**). However, the use of

hyperspectral data comes with its own difficulties in terms of cost, availability, processing, high dimensionality and limited availability especially in sub-Sahara Africa (Mansour, 2013). Since, the results of the study indicates the usefulness of the red-edge region of electromagnetic spectrum in discriminating the defoliation levels, the potential use of advanced multispectral remote sensing sensor with the red-edge band was therefore sought for discriminating the defoliation levels.

Table 8. 2: Comparison of confusion matrix obtained after the classification of the three levels of defoliation (PD, UD, R) using the optimal wavelengths for discriminating. The confusion matrix includes overall accuracy (OA), Kappa, class error (CE), producer's accuracy (PA) and user's accuracy (UA)

		ANOVA								RF					
	PD	UD	R	Total	UA (%)	PA (%)	CE (%)		PD	UD	R	Total	UA (%)	PA (%)	CE (%)
PD	34	21	0	55	77.27	61.82	38.18	**PD**	36	19	0	55	78.26	65.45	34.55
UD	10	45	0	55	68.18	81.82	18.18	**UD**	10	45	0	55	70.31	81.82	18.18
R	0	0	55	55	100	100	0	**R**	0	0	55	55	100	100	0
Total	44	66	55	165				**Total**	46	64	55	165			
CE (%)	22.73	32	0					**CE (%)**	21.7	29.7	0				

Kappa = 0.64, OA = 81.21%   **Kappa = 0.68, , OA = 82.42%**

8.4 Evaluating the Capability Of RapidEye High Resolution Imagery in Classifying the Insect Defoliation Levels

The use of multispectral dataset for classifying insect defoliation levels has faced great challenges especially in their ability to visualize and spectrally characterize changes in the canopy of leaves before, during and after an attack. While hyperspectral dataset could be able to address these challenges (Adelabu *et al.*, 2014; Coops *et al.*, 2003; Lawrence *et al.*, 2006), the cost and availability of hyperspectral data in sub-Saharan Africa is still a challenge. The solution therefore lies in identifying cheap and readily available sensors with strategic bands that can identify the defoliation even when symptoms are not visible to the human eye (Adelabu *et al.*,

2012; Zarco-Tejada *et al.*, 2000). In this thesis, valuable evidence of application and potential of the red edge channel of RapidEye multispectral image information to accurately classify insect defoliation levels (Partly defoliated, undefoliated and refoliating) are presented (Chapter 5). The subtle visual difference in the canopy colour levels of UD and PD challenged us to consider more detailed and specific spectral-bands for classifying the three levels of mopane worm defoliation. Two machine learning algorithm namely: Random Forest and Support vector were applied using different sets of spectral feature input, including and excluding the red edge band.

Result showed that the incorporation of red edge increases classification accuracy of insect defoliation levels in all analysis performed in the study. For instance, when all the 5 bands of RapidEye imagery were used for classification, the overall accuracies increases about 19% and 21% for SVM and RF respectively as opposed to when the red edge channel was excluded (**Table 8.3**). The study also found out that the Normalized Difference Red-Edge index (NDVI-RE) yielded a better accuracy (84% and 83% for RF and SVM respectively) than Normalized Difference Vegetation Index (NDVI) (61% and 51% for RF and SVM respectively). The results of sensitivity analysis (all RapidEye bands, NDVI and NDVI-RE) further confirmed that NDVI-RE and red-edge band of RapidEye image outperformed the other input variable (NDVI, Blue, Green Red and Near Infrared bands of RapidEye image) in classifying the defoliation levels (**Fig 8.2**). Although the result seems plausible, the study further tested the reliability of one of the classifiers used (RF) to ascertain the level of validity of its accuracies in classifying the insect defoliated levels in mopane woodland.

Table 8. 3: Class accuracies and percentage deviations for all RapidEye bands and all RapidEye bands excluding red-edge based on SVM, RF classification algorithm.

	RF			SVM		
	AB	AB-RE	Dev. (%)	AB	AB-RE	Dev. (%)
PD	56	52	4	68	60	8
R	96	40	56	88	44	44
UD	84	80	4	84	80	4
Overall Accuracy (%)	79	58	21	80	61	19

 * AB = All bands, AB-RE= All bands excluding red-edge
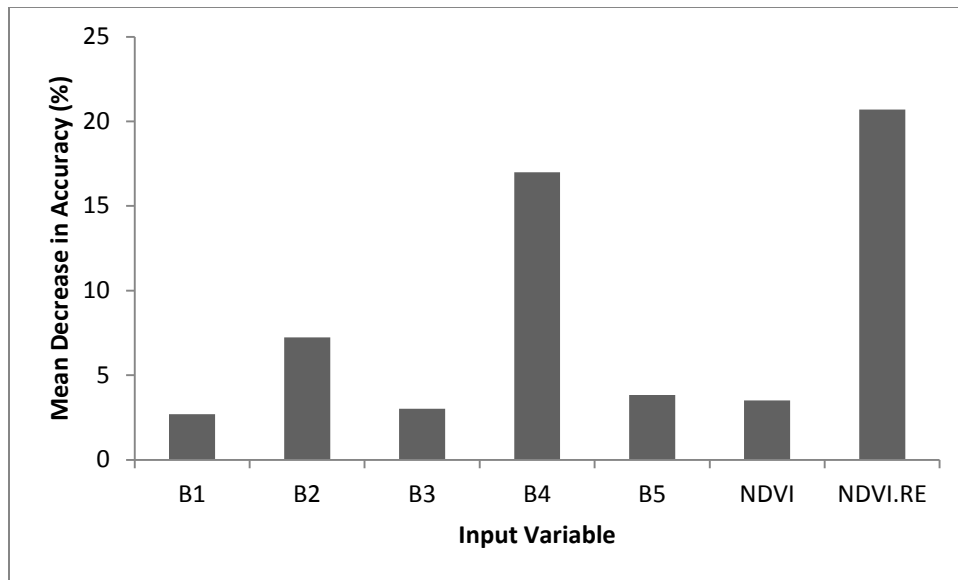
Figure 8. 2: Ranking of 5 bands of RapidEye image, NDVI and NDVI-RE using mean decrease accuracy for classifying levels of defoliation.

8.5 Evaluating the Reliability and Stability of Internal Accuracy Assessment of Random Forest when Compared with Independent Dataset

Classification accuracies are of utmost important for any classification in remote sensing (Foody, 2004; Manandhar *et al.*, 2009; Pontius and Millones, 2011). The accuracy of any classification depends on the available training samples used as well as the independent samples set aside for testing the performance of the classification (Congalton and Green, 1999). However, currently there are no agreements in the remote sensing world on the number of samples necessary for training and testing data and hence validating classification because sample size relates directly to the power of analysis which is a function of sample site variability (Foody *et al.*, 2006). Practically, in most of remote sensing applications for insect defoliation monitoring, the number of training samples is limited for comprehensive monitoring because of difficulty in collecting data before and after defoliation (Hall *et al.*, 2007). This is particularly true for defoliation in semi-arid environment such as mopane woodland, where collecting such sufficient training and test samples is difficult due to poor accessibility. Given these problems, the challenge was to test the reliability and stability of existing classifier for effective processing and classification of insect defoliation levels in mopane woodland using new multispectral RapidEye imagery.

In this thesis (Chapter 6), we tested the reliability and stability of the internal validation technique of RF algorithm. There have been debates on using RF internal validation technique for validation because it is assumed to be reliable and stable. While previous studies such as

Lawrence *et al.* (2006), Adam *et al.* (2011), Breiman and Cutler (2012) and Waske *et al.* (2007) have supported this claim, others such as Bylander (2002), Ruiz-Gazen and Villa (2007), Statnikov *et al.* (2008), Strobl and Zeileis (2008) and Menze *et al.* (2009) have contested the validity of this claim. Firstly, those who oppose its reliability have only done so using dataset other than remote sensing. Secondly, those who support the claim that internal validation of RF is reliably have randomly used various splitting options for training and testing (independent dataset), however, the rationale behind this has not been statistically justified. Therefore, the study evaluated the reliability and stability of the RF internal validation process with independent dataset when splited over different options for insect defoliation mapping using multispectral RapidEye Image. Furthermore, the study statistically validates the best split option for defoliation classification using RF and multispectral imagery (RapidEye).

Results showed that 40% -60%, 50% -50%, 60% -40% and 70% -30% split options were statistically not different (95% significance level) from each as compared with the other split options (**Table 8.4**). However, 70% -30% split option outperforms other split options because it had the lowest mean standard errors (0.2351) and that a much narrower distribution (CL = 0.0094) was observed from the standard errors produced. Similarly, the study also discovered that the internal estimate of accuracy of RF was able to provide relatively lower error rates (0.2319) for what is small datasets for classification as compared to other validation techniques used in this study **(Fig 8.3)**. Moreover, it was observed that the errors produced by the internal validation methods of RF was relatively stable as can be seen from the result of confidence interval (0.0006) obtained in this study. It was therefore concluded that for using RF in insect defoliation classification, 70-30% split option is preferable for validating classification independently although the robustness of the defoliation classification may be better achieved using internal bootstrapping accuracy estimate in RF.

Table 8. 4: P values of the Turkey post Hoc test performed to compare the errors produced from different splitting options

	20-80	30-70	40-60	50-50	60-40	70-30	80-20	RF Validating
20-80	1	**< 0.0001**	**0.002**	**0.005**	**< 0.0001**	**< 0.0001**	**0.034**	**< 0.0001**
30-70	**< 0.0001**	1	**0.000**	**< 0.0001**	**0.011**	**0.034**	**< 0.0001**	**< 0.0001**
40-60	**0.002**	**0.000**	1	1.000	0.975	0.877	**0.010**	0.806
50-50	**0.005**	**< 0.0001**	1.000	1	0.902	0.722	**0.010**	0.877
60-40	**< 0.0001**	**0.011**	0.975	0.902	1	1.000	**0.006**	0.902
70-30	**< 0.0001**	**0.034**	0.877	0.722	1.000	1	**0.036**	0.975
80-20	**0.034**	**< 0.0001**	**0.010**	**0.010**	**0.006**	**0.036**	1	**0.002**
RF Validating	**< 0.0001**	**< 0.0001**	0.806	0.877	0.902	0.975	**0.002**	1

* Highlighted values are significantly different.

Figure 8. 3: Histogram showing the frequency of standard error produced after classification for different holdout options

8.6 Quantifying Leaf Area Index (LAI) for Different Defoliation Levels using Radiative Transfer Model

Relationship between biophysical characteristics such as leaf area index (LAI) of canopies and reflectance has previously been studied with great success (Asner, 1998; Atzberger, 2004; Richter *et al.*, 2012; Ustin *et al.*, 2009). The estimation of LAI from remote sensing are usually based on empirical method, spectral vegetation indices or radiative transfer (RT) inversion model (Rullan-Silva *et al.*, 2013). The empirical methods have been observed to be sensor specific and dependent on site and sampling condition, hence they are expected to change in space and time (Colombo *et al.*, 2003; Meroni *et al.*, 2004). Moreover, LAI estimation using the empirical methods still face a lot of challenges with regards to the need for collecting in situ calibration datasets leading to high cost and labour intensive measurements (Richter *et al.*, 2011). To this end, RT models provide an alternative approach as it assumes that the spectral variation of canopy reflectance as a function of canopy's leaf and soil background could be accurately described using physical laws through the interaction and transfer of radiation inside the canopy (Meroni *et al.*, 2004).

Although, advances have been made in the use of hyperspectral (Adelabu *et al.*, 2014) and multispectral (Adelabu *et al.*, 2013) dataset for discriminating insect defoliation levels, the knowledge of biophysical variables such as LAI may help to understand the influence of the insects on the canopy at each of the defoliation levels. In this thesis (Chapter 7), LAI was estimated for different defoliation levels based on simulated data from RapidEye image by inverting the canopy radiative transfer model PROSAIL H (Jacquemoud *et al.*, 2009). We compared the estimated LAI with other results in the literature.

The estimated mean LAI are within the LAI values obtained by other simulation studies (Colombo *et al.*, 2003; Weiss *et al.*, 2000) (**Table 8.5**). Furthermore, the estimated LAI were also in the range of LAI estimated by previous studies that have attempted to classify defoliation levels using remote sensing data (Hall *et al.*, 2003b). The study further observed that the NDVI-RE has the strongest relationship with the estimated LAI while the red band (band 2) has a weak relationship (**Fig 8.4**). The result is in line with other studies that have conducted relationship analysis between LAI values and reflectance from different regions of electromagnetic spectrum (Vina *et al.*, 2011).

Table 8. 5: Descriptive statistics of the estimated LAI from the Inverted PROSAIL H RT model

Defoliation Levels	Mean	Maximum	Minimum	Std. Dev.
PD	4.622328	5.020619	4.233532	0.140263
UD	5.123606	5.969437	4.494172	0.164477
R	4.147604	4.647977	3.433718	0.170675

The study confirmed that it is possible to estimate LAI through the inversion of a radiative transfer model using relatively cheap and available new generational multispectral imagery such as RapidEye with comparable values to those of empirical approaches (from literature).
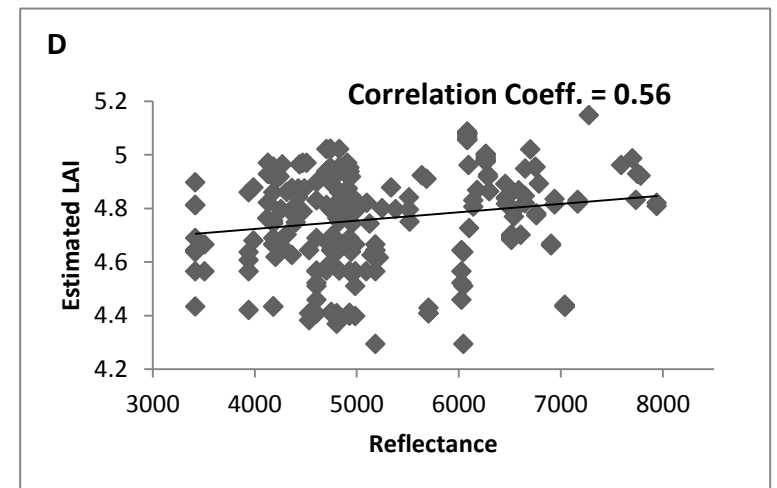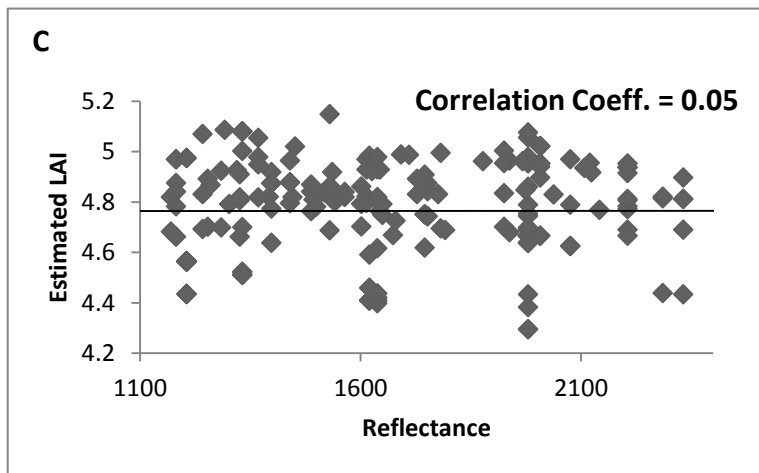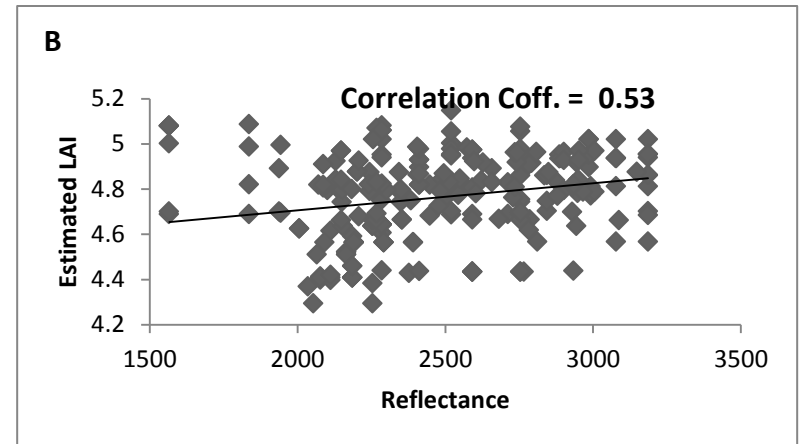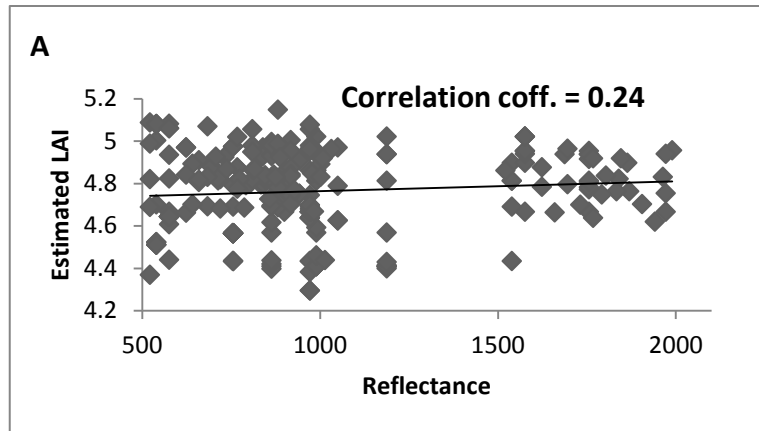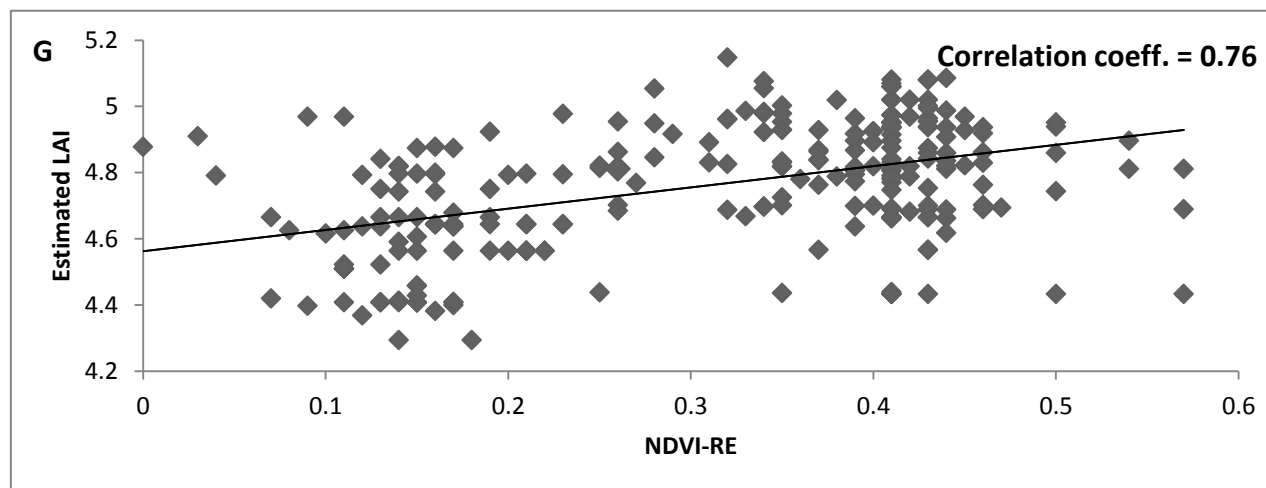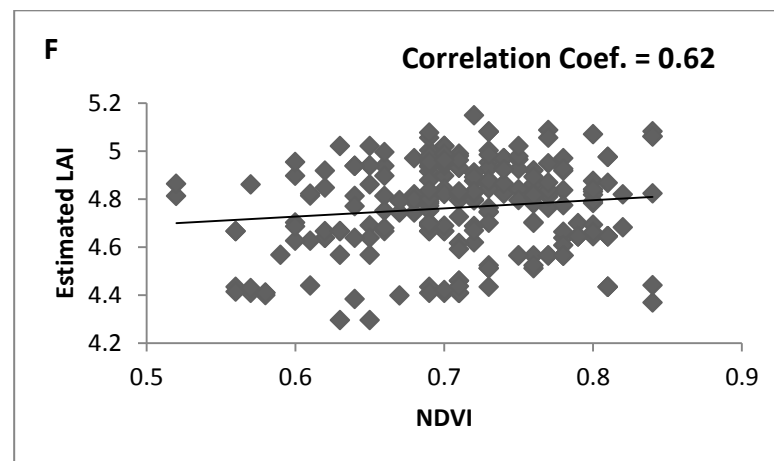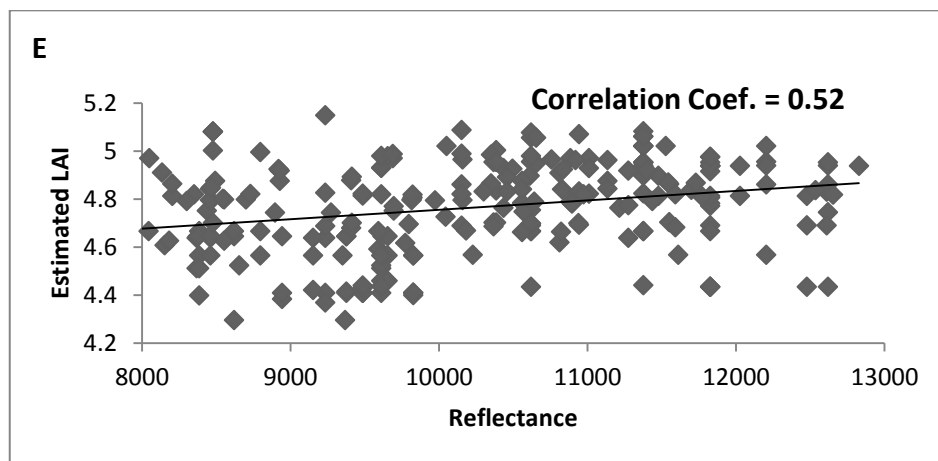
Figure 8. 4: Pearson correlation coefficient of LAI and different bands of RapidEye and selected vegetation indices, A=Band 1, B=Band 2, C=Band 3, D=Band 4, E=Band 5, F=NDVI, G=NDVI-RE.

8.7 Conclusion

The main aim of this study was to investigate the capabilities of remote sensing techniques in classifying different levels of insect defoliation in mopane woodland using different remote sensing sensors. The findings reported in this thesis are that the information contained in some remote sensing sensors and the use of advance classification algorithm can accomplish these tasks. The main conclusions are based on the following findings from the different objectives addressed in this study:

1. Canopy reflectance from relatively cheap and readily available high resolution multispectral data with strategic bands such as RapidEye has a strong potential to be used for discriminating *Colophospermum mopane* and its co-existing species using machine learning algorithm such as SVM and RF. The result implies that the challenges facing researchers in classifying tree species in semi-arid environment could be minimized through the use of remote sensing data with strategic bands. Furthermore, the results permitted us to use ground based hyperspectral data to identify defoliation levels in mopane trees.

2. Using the ground based hyperspectral data, the RF algorithm could be used for discriminating the spectral differences among the different levels of change in canopy after insect defoliation with relatively high accuracy. The analysis of the ground based hyperspectral data also showed the importance of the red-edge and near-infrared regions in discriminating the insect defoliation levels. This result prompts us to downscale to readily available and cheap multispectral imagery with strategic bands.

3. The inclusion of the red-edge channel in the RapidEye generally benefits the image classification of insect defoliated levels in mopane woodland.

4. Similarly, the study concludes that vegetation indices derived from red-edge band such has NDVI-RE have stronger capability to discriminate insect defoliation levels as compared with indices derived from traditional remote sensing bands.

5. The internal accuracy assessment of machine learning RF algorithm is relatively reliable and stable for insect defoliation classification and it is worth considering as a desirable technique for future insect defoliation applications especially in complex environments such as semi-arid environment where usually no convenient or sufficient field data are available.

6. Biophysical variables such as LAI of insect defoliated levels could be estimated through the inversion of radiative transfer model such as PROSAILH using relatively

cheap and available new generational multispectral imagery such as RapidEye with comparable values to those of conventional or empirical approaches.

8.8 The Future

Remote sensing is an integral and essential tool for the collection of data needed to support decisions and action programs to improve forest health (Zhou *et al.*, 2010). While not all attempts to use remote sensing in forest health protection have proven successful, many have shown to meet data requirements, and have proven to be cost-effective alternatives to ground data acquisition. The findings from this thesis contribute to existing research in general and further support scientific knowledge of environmental management practices and food security in Africa and other sites around the world more sensitive to global changes. Moreover, the findings of this thesis could lay the foundation for possible management practices that will enable efficient and sustainable use of the resources emanating from mopane woodland. Therefore, the following recommendations are suggested for future research:

1. The present study used ground based hyperspectral data to discriminate the levels of defoliation in mopane woodland; it will however be good for future research to test the ability of airborne or spaceborne hyperspectral imagery in detecting and classifying the defoliation levels.

2. New generation multispectral data contains strategic bands and do not require complex processing techniques and are available and relatively inexpensive. In this regard, the capability of multispectral sensors other than RapidEye (e.g. Sumbandilasat, Sentinel -2, and Worldview-2) in classifying the different defoliated levels should be tested.

3. In order for remote sensing methods to become operational for classifying these different defoliation levels, future research is needed to investigate the optimal spatial resolution and pixel size that could better classify the different levels of insect defoliation.

4. Further research should investigate and measure the biophysical and biochemical variables (including LAI, Nitrogen, and Tannin etc.) of mopane canopies at different defoliation levels with the aim of relating it canopy reflectance.

5. Although the reliability of the internal accuracy assessments of RF in classifying different defoliation levels was tested in this study, more research is still needed on the strength of RF and SVM used in this study as compared to other classifiers such as artificial neural network which has been proved to be successful in remote sensing. This is particularly essential for insect defoliation classification in areas such as semi-arid environments where it may be difficult to collect samples because of accessibility.

References

Adam, E., and O. Mutanga, 2009. Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing* 64, 612-620.

Adam, E., and O. Mutanga, 2010. Improving the spectral discriminating of papyrus (Cyperus papyrus L.) and its co-existent species at canopy level with hyperspectral indices and random forest algorithm. *In* "The 8th Conference of the African Association of Remote Sensing for the Environment (ARSE 2010)", Addis-Ababa, Ethiopia.

Adam, E. M., O. Mutanga, D. Rugege, and R. Ismail, 2009. Field spectrometry of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands of St Lucia, South Africa. *In* "Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009", Vol. 4, pp. IV-260-IV-263.

Adam, E. M., O. Mutanga, D. Rugege, and R. Ismail, 2011. Discriminating the papyrus vegetation (*Cyperus papyrus L.*) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing* 33, 552-569.

Adam, E. M., O. Mutanga, D. Rugege, and R. Ismail, 2012. Discriminating the papyrus vegetation (Cyperus papyrus L.) and its coexistent species using random forest and hyperspectral data resampled to HYMAP. *International Journal of Remote Sensing* 33, 552-569.

Adelabu, S., O. Mutanga, E. Adam, and M. A. Cho, 2013. Exploiting Machine Learning Algorithms for Tree Species Classification in a Semi-Arid woodland Using RapidEye image *Journal of Applied Remote Sensing* 7, 073480.

Adelabu, S., O. Mutanga, E. Adams, and R. J. Sebego, 2014. Spectral Discrimination of Insect Defoliation Levels in Mopane Woodland using Hyperspectral Data. *IEEE Journal of selected topics in Earth Observations and Remote Sensing* 7, 177-186.

Adelabu, S., O. Mutanga, and M. A. Cho, 2012. A Review of Remote Sensing of Insect Defoliation and its Implications for the Detection and Mapping of Imbrasia belina Defoliation of Mopane Woodland. *The African Journal of Plant Science and Biotechnology* 6, 1-13.

Allen, T. R., and J. A. Kupfer, 2000. Application of spherical statistics to change vector analysis of Landsat Data: southern appalachian spruce-fir forests. *Remote Sensing of Environment* 74, 482-493.

ASD, 2005. "Analytical Spectral Devices, Inc., Fieldspec Handheld 2 TM: User's Guide, Version 4.05," Analytical Spectral Devices, Boulder,USA.

Asner, G. P., 1998. Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment* 64, 234-253.

Asner, G. P., J. A. Hicke, and D. B. Lobell, 2003a. Per-pixel analysis of forest structure: Vegetation indices, spectral mixture analysis and canopy reflectance modeling. *In* "Methods and Applications for Remote Sensing of Forests: Concepts and Case Studies." (M. A. Wulder and S. E. Franklin, eds.). Kluwer Academic Publishers, New York.

Asner, G. P., J. M. O. Scurlock, and J. A. Hicke, 2003b. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecology and Biogeography* 12, 191-205.

Asner, G. P., C. A. Wessman, and D. S. Schimel, 1998. Heterogeneity of Savanna Canopy Structure and Function from Imaging Spectrometry and Inverse Modeling. *Ecological Applications* 8, 1022-1036.

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment* 93, 53-67.

Bacour, C., F. Baret, D. Béal, M. Weiss, and K. Pavageau, 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment* 105, 313-325.

Barry, K., C. Stone, and C. Mohammed, 2008. Crown-scale evaluation of spectral indices for defoliated and discoloured eucalypts. *International Journal of Remote Sensing* 29, 47–69.

Benediktsson, J. A., and J. R. Sveinsson, 2004. Random forest classification of multisource remote sensing and geographic data. *In* "2004 IEEE International Geoscience and Remote Sensing Symposium", Vol. 2, pp. 1049-1052.

Bennett, D. D., and B. M. Tkacz, 2008. Forest health monitoring in the United States: a program overview. *Australian Forestry* 71, 223-228.

Bhalotra, Y. P. R., 1987. Climate of Botswana, Part 2: Elements of Climate. Botswana Meteorological Services, Gaborone.

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

Breiman, L., and A. Cutler, 2012. State of the Art of Data Mining Using Random Forest. *In* "Salford Data Mining Conference", San Diego, USA.

Burges, C. J. C., 1998 A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121-167.

Burns, G., and A. Joyce, 1981. Evaluation of land cover change detection techniques using Landsat MSS data. *In* "7th Pecora Symposium", pp. 252-260, Sioux Falls, South Dakota.

Bylander, T., 2002. Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine Learning* 48, 287-297.

Campbell, P. K. E., B. N. Rock, M. E. Martin, C. D. Neefus, J. R. Irons, E. M. Middleton, and J. Albrechtova, 2004. Detection of initial damage in Norway spruce canopies using hyperspectral airborne data. *International Journal of Remote Sensing* 25, 5557−5583.

Carleer, A., and E. Wolff, 2004. Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering and Remote Sensing* 70, 135-140.

Carter, G., 1994. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing* 15, 697-703.

Carter, G. A., and A. K. Knapp, 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany* 88, 677−684.

Carter, G. A., and R. L. Miller, 1994. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment* 50, 295-302.

Chan, J. C. W., and D. Paelinckx, 2008. Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment* 112, 2999-3011.

Cheng, P., and C. Chaapel, 2008. Increased image collection opportunities, Digital Globe's worldview-1 satellite. *In* "Geoformatics online magazine". http://www.pcigeomatics.com/pdf/WorldView-1.pdf.

Cho, M., and A. Skidmore, 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment* 101, 181-193.

Cho, M., A. Skidmore, and C. Atzberger, 2008. Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimations using properties optique spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined

leaves (SAILH) simulated data. *International Journal of Remote Sensing* 29, 2241-2255.

Cho, M. A., P. Debba, R. Mathieu, L. Naidoo, J. A. van Aardt, and G. P. Asner, 2010. Improving Discrimination of Sa-vanna Tree Species Through a Multiple-Endmember Spectral Angle Mapper Approach: Canopy-Level Analysis. *IEEE Transactions on Geoscience and Remote Sensing* 48, 4133-4142.

Ciesla, W. M., 2003. European Woodwasp: A potential threat to North America's conifer forests. *Journal of Forestry*, 18-23.

Ciesla, W. M., and R. E. Acciavatti, 1982. Panoramic aerial photography for mapping gypsy moth defoliation. *In* "Report 83-1", pp. 1-17. Fort Collins, CO.

Clerke, W. H., and C. W. Dull, 1990. Evaluating the utility of SPOT digital imagery for delineating and categorizing gypsy moth defoliation. *In* "In Protecting Natural Resources with Remote Sensing, The Third Forest Service Remote Sensing Application Conference", pp. 13-21, Tucson,Arizona.

Clevers, J. G., 1997. A Simplified approach yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment* 61, 221-228.

Cohen, W. B., and M. Fiorella, 1998. Comparison of methods for detecting conifer forest change with thematic mapper imagery. *In* "Remote sensing change detection: environmental monitoring methods and applications" (C.D Lunetta and R. S. Elvidge, eds.), pp. 89-102. Ann Arbor Press, Chelsea,Michigan, USA.

Collantes, H. G., E. Gianoli, and H. M. Niemeyer, 1999. Defoliation affects chemical defenses in all plant parts of rye seedlings. *J. Chem. Ecol.* 25, 491-499.

Collins, J. B., and C. Woodcock, 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing Environment* 56, 66–77.

Colombo, R., D. Bellingeri, D. Fasolini, and C. M. Marino, 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment* 86, 120-131.

Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Macé, and A. Pragnére, 2002. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 1-15.

Congalton, R. G., and K. Green, 1999. "Assessing the Accuracy of Remotely Sensed Data: Principles and Practices," Lewis Publishers, Boca Raton.

Cook, B., P. Bolstad, J. Martin, F. Heinsch, K. Davis, and W. Wang, 2008. Using lightuse and production efficiency models to predict photosynthesis and net carbon exchange during forest canopy disturbance. *Ecosystems* 11, 26−44.

Coops, N., M. Stanford, K. Old, M. Dudzinski, D. Culvenor, and C. Stone, 2003. Assessment of Dothistroma needle blight of Pinus radiata using airborne hyperspectral imagery. *Phytopathology* 93, 1524–1532.

Coops, N. C., C. Stone, D. S. Culvenor, and L. Chisholm, 2004. Assessment of crown condition in eucalypt vegetation by remotely sensed optical indices. *Journal of Environmental Quality* 33, 956–964.

Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, 2004. Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing* 25, 1565−1596.

Darvishzadeh, R., C. Atzberger, A. K. Skidmore, and A. A. Abkar, 2009. Leaf area index derivation from hyperspectral vegetation indices and the red edge position. *International Journal of Remote Sensing* 30, 6199-6218.

Darvishzadeh, R., A. Skidmore, C. Atzberger, and S. Wieren, 2008. Estimation of vegetation LAI from hyperspectral reflectance data: effects of soil type and plant architecture. *Int J Appl Earth Obs Geoinform* 10, 358-373.

Dash, J., and P. J. Curran, 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Advances in Space Research* 39, 100-104.

de Beurs, K. M., and P. A. Townsend, 2008. Estimating the effect of gypsy moth defoliation using MODIS. *Remote Sensing of Environment* 112, 3983–3990.

DeFries, R. S., and J. C. W. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment* 74, 503-515.

Diago, M. P., M. Vilanova, and J. Tardaguila, 2010. Effects of timing of manuals and mechanical early defoliation on the Aroma of Vitis Vinifera L.Tempranillo wine. *American Journals of Enology and Viticulture* 61, 382-391.

Dieterle, F. J., 2003. Multianalyte quantifications by means of integration of artificial neural networks,genetic algorithms and chemometrics for time-resolved analytical data, Universität Tübingen, Tübingen.

Ditlhogo, M., J. Allotey, S. Mpuchane, G. Teferra, B. A. Gashe, and B. A. Siame, 1996. Interactions between the mopane caterpillar, Imbrasia belina, and its host,

Colophospermum mopane in Botswana. *In* "Management of mopane in Southern Africa" (C. Flower, Wardell-Johnson, G. and Jamieson, A, ed.), pp. 46-49.

Do, F., A. Rocheteau, A. L. Diagne, V. A. Goudiaby, A. Granier, and J. P. Lhomme, 2008. Stable annual pattern of water use by Acacia tortilis in Sahelian Africa. *Tree Physiology* 28, 95-104.

Dolph, R. E. J., 1980. "Budworm activity in Oregon and Washington, 1947-1979 pacific Northwest region."

Dull, C., D. Rubel, B. Spears, and R. Uhler, 1990. Integration of remote sensing and GIS databases to monitor forest conditions in the southern region. *In* " Protecting Natural Resources with Remote Sensing, The Third Forest Service Remote Sensing Application Conference", Tucson Arizona.

Durbha, S. S., R. L. King, and N. H. Younan, 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment* 107, 348-361.

Dye, D. G., and C. J. Tucker, 2003. Seasonality and trends of snow-cover, vegetation index, and temperature in northern Eurasia. *Geophysical Research Letters* 30 58 (1) − 58 (4).

ENVI, 2006. "Environment for Visualising Images," ITT industries, Inc, USA.

Fan, L., S. Liu, H., , C. Bernhofer, H., , H. Liu, and F. Berger, H., 2007. Regional land surface energy fluxes by satellite remote sensing in the upper Xilin River watershed (Inner Mongolia, China). *Theor Appl Climatol* 88, 231-245.

Fassnacht, K. S., S. T. Gower, M. D. MacKenzie, E. V. Nordheim, and T. M. Lillesand, 1997. Estimating the leaf area index of North Central Wisconsin forests using the Landsat thematic mapper. *Remote Sensing of Environment* 61, 229-245.

Foody, G. M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing* 70, 627-633.

Foody, G. M., and A. Mathur, 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103, 179−189.

Foody, G. M., A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, 2006. Training set size requirements for the classification of a specific class. *Remote Sensing of Environment* 104, 1-14.

Franklin, S. E., 2001. "Remote Sensing for Sustainable Forest Management," Lewis Publishers, Boca Raton, Florida.

Fraser, R. H., and R. Latifovic, 2005. Mapping insect-induced tree defoliation and mortality using coarse spatial resolution imagery. *International Journal of Remote Sensing* 26, 193−200.

Gamon, J., J. Pen˜uelas, and C. B. Field, 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment* 41, 35-44.

Gamon, J. A., and J. S. Surfus, 1999. Assessing leaf pigment content and activity with a reflectometer. *New Phytologist* 143, 105–117.

Gitelson, A., and M. Merzlyak, 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: experiments with autumn chesnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology* 22, 247–252.

Gitelson, A. A., G. P. Keydan, and M. N. Merzlyak, 2006. Three band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters* 33.

Gitelson, A. A., M. N. Merzlyak, and H. K. Lichtenthaler, 1996. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology* 148, 501-508.

Gitelson, A. A., Y. Zur, O. B. Chivkunova, and M. N. Merzlyack, 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology* 75, 272–281.

Gonsamo, A., 2011. Normalized sensitivity measures for leaf area index estimation using three-band spectral vegetation indices. *Int. J. Remote Sens* 32, 2069-2080.

Goodwin, N., N. C. Coops, and C. Stone, 2005. Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances. *International Journal of Applied Earth Observation and Geoinformation* 7, 11–28.

Greenberg, J. A., S. Z. Dobrowski, C. M. Ramirez, J. L. Tuil, and S. L. Ustin, 2006. Acbottom–up approach to vegetation mapping of the Lake Tahoe Basin usingchyperspatial image analysis. *Photogrammetric Engineering and Remote Sensing* 72, 581−589.

Greenberg, J. P., A. Guenther, P. Harley, . , L. Otter, E. M. C. Veenendaal, N. Hewitt, A. E. James, and S. M. Owen, 2003. Eddy flux and leaf-level measurements of biogenic VOC emissions from mopane woodland of Botswana. *J. Geophys. Res.,* 108, 8466.

Greyling, M., F. H. van der Bank, J. P. Grobler, and D. C. J. Wessels, 2001. Allozyme variation in two populations of the Mopane worm, Imbrasia belina (Saturniidae), and

the effect of developmental stage and staggered generations. *South African Journal of Animal Science* 31, 15-24.

Gupta, R. K., D. Vijayan, and T. S. Prasad, 2003. Comparative analysis of red-edge hyperspectral indices. *Advances in Space Research* 32, 2217-2222.

Hall, R. J., D. P. Davidson, and D. R. Peddle, 2003a. Ground and remote estimation of leaf area index in Rocky Mountain forest stands, Kananaskis, Alberta. *Canadian Journal of Remote Sensing* 29, 411–427.

Hall, R. J., E. H. Hogg, J. P. Brandt, B. S. Case, R. A. Fernandes, C. Butson, and S. G. Leblanc, 2003b. Relating aspen defoliation to changes in leaf area derived from field and satellite remote sensing data. *Can. J. Rem. Sens* 29, 299-313.

Hall, R. J., R. S. Skakun, and E. J. Arsenault, 2007. Remotely sensed data in the mapping of insect defoliation. *In* "Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches" (W. A. Michael and S. E. Franklin, eds.), pp. 85-111. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA.

Heller, R. C., J. J. Ullimon, and H. N. Anderson, 1981. The use of photo interpretation and remote sensing techniques to establish hazard rating criteria for spruce budworm susceptible stands. (Forest and Service, eds.), Vol. CANUSA- West Final Report, pp. 48. United States Department of Agriculture, USA.

Horning, N., 2010. Random Forests : An algorithm for image classification and generation of continuous fields data sets. *In* "International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences".

Houborg, R., H. Soegaard, and E. Boegh, 2007. Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sensing of Environment* 106, 39-58.

Hrabar, H., D. Hattas, and J. du Toit, 2009a. Differential effects of defoliation by mopane caterpillars and pruning by African elephants on the regrowth of Colophospermum mopane foliage. *Journal of Tropical Ecology* 25, 301-309.

Hrabar, H., D. Hattas, and J. T. du Toit, 2009b. Intra-specific host preferences of mopane moths (Imbrasia belina) in mopane (Colophospermum mopane) woodland. *African Zoology* 44, 131-140.

Hsu, C. W., C. C. Chang, and C. J. Lin, 2009. A practical guide to support vector classification. *In* "Technical Note", Department of Computer Science and Information Engineering, National Taiwan Univ., Taiwan.

Hudak, A. T., E. K. Strand, L. A. Vierling, J. C. Byrne, J. U. H. Eitel, S. Martinuzzi, and M. J. Falkowski, 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sensing of Environment* 123, 25-40.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83, 195−213.

Ismail, R., and O. Mutanga, 2011. Discriminating the early stages of Sirex noctilio infestation using classification tree ensembles and shortwave infrared bands. *International Journal of Remote Sensing* 32, 4249-4266.

Ismail, R., O. Mutanga, and U. Bob, 2007. Forest health and vitality: The detection and monitoring of Pinus patula trees infected by Sirex noctilio using digital multispectral imagery (DMSI). *Southern Hemisphere Forestry Journal* 69, 39-47.

Ismail, R., O. Mutanga, L. Kumar, and U. Bob, 2008. Determining the optimal resolution of remotely sensed data for the detection of *Sirex noctilio* infestations in *Pinus patula* plantations in KwaZulu-Natal,South Africa. *The South African Geographical Journal* 90, 196-204.

Jacquemoud, S., and F. Baret, 1990. PROSPECT: a model of leaf optical properties spectra. *Remote Sensing of Environment* 34, 75-91.

Jacquemoud, S., F. Baret, B. Andrieu, F. M. Danson, and K. Jaggard, 1995. Extraction of vegetation biophysical parameters by inversion of the PROSPECT +SAIL models on sugar-beet canopy reflectance data - Application to TM And AVIRIS sensors. *Remote Sensing of Environment* 52, 163-172.

Jacquemoud, S., W. Verhoef, F. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, and e. al., 2009. PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment* 113, 56-66.

Jensen, J., 2005. " Introductory digital image processing: a remote sensing perspective," 3rd/Ed. Pearson Education, Inc.

Joria, P. E., and S. C. Ahearn, 1991. A comparison of the SPOT and Landsat Thematic Mapper satellite systems for detecting gypsy moth defoliation in Michigan. *Photogrammetric Engineering and Remote Sensing* 57, 1605−1612.

Kantola, T., M. Vastaranta, X. Yu, P. Lyytikäinen, M. Holopainen, M. Talvitie, S. Kaasalainen, S. Solberg, and J. Hyypaa, 2010. Classification of defoliated tres using tree-level airborne laser scanning data combined with aerial images. *Remote Sensing of Environment* 2, 2665-2679.

Khanna, S., A. Palacios-Orueta, Whiting, M.l.,, S. L. Ustin, D. Riaño, and J. Litago, 2007. Development of angle indexes for soil moisture estimation, dry matter detection and landcover discrimination. *Remote Sensing of Environment* 109, 154 - 165.

Kharuk, V. I., K. Rnason, and M. Dvinskaya, 2007. Evidence of evergreen conifer invasion into larch dominated forests during recent decades in central Siberia. *Eurasian Journal of Forest Research* 10, 163–171.

Kim, H. O., J. O. Yeom, and Y. S. Kim, 2011. Agricultural land cover classification using rapideye satellite imagery in South Korea - first result. *In* "Remote Sensing for Agriculture, Ecosystems, and Hydrology XIII" (Christopher M. U. Neale and A. Maltese, eds.), Vol. 8174. SPIE Proceedings.

Kimes, D., Y. Knjazikhin, J. L. Privette, A. Abuelgasim, and F. Gao, 2000. Inversion methods for physically-based models. *Remote Sensing Reviews* 18, 381-440.

Kodani, E., Y. Awaya, K. Tanaka, and N. Matsumura, 2002. Seasonal patterns of canopy structure,biochemistry and spectral reflectance in a braod-leaved deciduous Faus crenata canopy. *Forest Ecology and Management* 167, 233-249.

Koike, T., 1987. Seasonal changes in the photosynthetic rate at light saturation and respiration rate of deciduos braod-leaved trees. *In* "Transition 98th Meeting", Vol. 98, pp. 389-390. Japan Forest Society, Japan.

Kovacs, K., K. J. Ranson, and V. I. Kharuk, 2005. Detecting siberian silk moth damage in central Siberia using multi-temporal MODIS data. *In* "International Workshop on the Analysis of Multi-temporal Remote Sensing images", Biloci, Missippi, USA.

Krahwinkler, P., and J. Rossman, 2011. Using Decision Tree Based Multiclass Support Vector Machines for Forest Mapping. *In* "IEEE International Geoscience and Remote Sensing Symposium" (Lena Halounová, ed.), pp. 307-318, Vancouver.

Kulkarni, V. Y., and P. K. Sinha, 2013. Random Forest Classifiers : A Survey and Future Research Directions. *International Journal of Advanced Computing* 36, 1144-1153.

Kumar, L., K. S. Schmidt, S. Dury, and A. Skidmore, 2001. Review of hyperspectral remote sensing and vegetation science. *In* "Imaging spectrometry: basic principles and prospective applications" (F. van der Meer and S. M. de Jong, eds.). Kluwer Academic Press, Dordrecht, The Netherlands.

Kuusk, A., 1991. The hot spot effect in the plant canopy reflectance. *In* "Photon-Vegetation Interactions", pp. 139-159. Springer.

Lange, H., and S. Solberg, 2008. Leaf area index estimation using LIDAR and forest reflectance modelling of airborne hyperspectral data. *In* "Proc IGARSS Int Conf on "The next generation"", pp. 475-478, Boston (USA).

Lawrence, R., and M. Labus, 2003. Early detection of Douglas-fir beetle infestation with subcanopy resolution hyperspectral imagery. *West. J. Applied Forestry* 18, 202-206.

Lawrence, R. L., S. D. Wood, and R. L. Sheley, 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* 100, 356-362.

Lee, W., V. Alchanatis, C. Yang, M. Hirafuji, D. Moshou, and C. Li, 2010. Sensing technologies for precsision specialty crop production. *Computer and Electronic in Agriculture* 74, 2-33.

Lėonard, J., 1949. Notulae systematicae IV: Caesalpiniaceae- Amherstieae africanae americanaeque. *Bulletin du Jardin Botanique de L'ėtat (Bruxelles)* 19, 388-391.

Liaw, A., and M. Wiener, 2002. Classification and regression by randomForest. *R News* 2, 18-22.

Liu, X., M. Song, D. Tao, L. Zicheng, L. Zhang, C. Chen, and J. Bu, 2013. Semi-supervised Node Splitting for Random Forest Construction. *In* "IEEE Conference on Computer Vision and Pattern Recognition".

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

Lucas, R., A. Rowlands, O. Niemann, and R. Merton, 2004. Hyperspectral sensors and applications. *In* "Advanced image processing techniques for remotely sensed hyperspectral data", pp. 11–49. Springer, Berlin.

Luther, J. E., S. E. Franklin, J. Judak, and J. Meades, 1997. Forecasting the susceptibility and vulnerability of balsam fir stands to insect defoliation with Landsat Thematic Mapper data. *Remote Sensing of Environment* 59, 77−91.

Makhado, A., J. Potgieter, D. Wessels, A. Saidi, and K. Masehela, 2012. Use of Mopane Woodland Resources and Associated Woodland Management Challenges in Rural Areas of South Africa. *Ethnobotany Research & Applications* 10, 369-379.

Malone, S., 2001. Assessment of soybean leaf area for redefining management strategy for Leaf-Feeding insects, Virginia Polytechnic Institute & State University, Blacksburg.

Manandhar, R., I. O. A. Odeh, and T. Ancev, 2009. Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing* 1, 330-344.

Mansour, K., 2013. Comparing the New Generation World View-2 to Hyperspectral Image Data for Species Discrimination. *International Journal of Development Research* 3, 008-013.

Mantlana, K. B., 2008. Seasonal and inter-annual variations of leaf-level photosynthesis and soil respiration in the representative ecosystems of the Okavango Delta, Botswana, Wageningen University, Netherlands.

Mapaure, I., 1994. The Distribution of Mopane. *Kirkia* 15, 1-5.

Melgani, F., and L. Bruzzone, 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens* 42, 1778-1790.

Menze, B. H., B. M. Kelm, R. Masuch, U. Himmelreich, W. Petrich, and F. A. Hamprecht, 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10, 213.

Meroni, M., R. Colombo, and C. Panigada, 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment* 92, 195-206.

Milton, E. J., M. E. Schapeman, K. Anderson, M. Kneubuhler, and N. Fox, 2009. Progress in field spectroscopy. *Remote Sensing of Environment* 113, 92-109.

Mitchell, M. W., 2011. Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. *Open Journal of Statistics* 1, 205-211.

Mohammed, G. H., W. D. Binder, and S. L. Gillies, 1995. Chlorophyll fluorescence: a review of its practical forestry applications and instrumentation. *Scandinavian Journal of Forest Research* 10, 383-410.

Mojeremane, W., and T. Kgathi, 2005. Seed treatments for enhancing germination of Colophospermum mopane seeds: A Multipurpose Tree in Botswana. *Journal of Biological Sciences* 5, 309.

Morgan, T., K. Riley, R. Tannebring, and L. Veldhuis, 2010. Evaluating the Impacts of Small-Scale Greenspace: A Case Study of Harlem Place in Downtown Los Angeles. , University of California, Califonia.

Muchoney, D. M., and B. N. Haack, 1994. Change detection for monitoring forest defoliation. *Photogrammetric Engineering and Remote Sensing* 60, 1243−1251.

Musvoto, C., I. Mapaure, T. Gondo, A. Ndeinoma, and T. Mujawo, 2007. Reality and Preferences in Community Mopane (Colophospermum Mopane) Woodland Management in Zimbabwe and Namibia. *World Academy of Science Engineering and Technology* 28, 83-87.

Mutanga, O., and E. Adam, 2011. High density biomass estimation: Testing the utility of Vegetation Indices and the Random Forest Regression algorithm *In* "34th International Symposium on Remote Sensing of Environment", Sydney, Australia.

Mutanga, O., and A. K. Skidmore, 2004a. Integrating imaging spectroscopy and neural networks to map tropical grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment* 90, 104-115.

Mutanga, O., and A. K. Skidmore, 2004b. Narrow band vegetation indices solve the saturation problem in biomass estimation. *International Journal of Remote Sensing* 25, 3999-4014.

Mutanga, O., and A. K. Skidmore, 2007. Red-edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, 34-42.

Mutanga, O., A. K. Skidmore, and H. H. T. Prins, 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa using continuum removed absorption features. *Remote Sensing of Environment* 89, 393-408.

Mutanga, O., J. A. N. Van Aardt, I. Riyad, A. Fethi, and K. Lalit, 2009. Imaging spectroscopy (Hyperspectral Remote sensing) in Southern Africa: an overview. *South African Journal of Science* 105, 193-198.

Naughton, D., A. Brunn, J. Czapla-Myers, S. Douglass, M. Thiele, H. Weichelt, and M. Oxfort, 2011. Absolute radiometric calibration of the RapidEye multispectral imager using the reflectance-based vicarious calibration method. *Journal of Applied Remote Sensing* 5, 053544.

Nelson, R. F., 1983. Detecting forest canopy change due to insect activity using Landsat MSS. *Photogrammetric Engineering and Remote Sensing* 49, 1303-1314.

Niemann, K. O., G. Quinn, D. G. Goodenough, F. Visintini, and R. Loos, 2012. Addressing the Effects of Canopy Structure on the Remote Sensing of Foliar Chemistry of a 3-Dimensional, Radiometrically Porous Surface. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5, 584 - 593.

Nitze, I., U. Schulthess, and A. H., 2012. Comparison of Machine Learning Algorithms Random Forest, Artificial Neural Network And Support Vector Machine to Maximum

Likelihood for Supervised Crop Type Classification. *In* "Proceedings of the 4th GEOBIA", pp. 035-040, Rio de Janeiro - Brazil.

Özçift, A., 2011. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Computers in Biology and Medicine* 41, 265-271.

Palgrave, K. C., 1983. "Trees of southern Africa," Struik Publishers, Cape Town.

Paulick, V., 2003. Cycles of Life. Namib Desert Environmental Education Trust, Namibia.

Pen˜uelas, J., F. Baret, and I. Filella, 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetical* 31, 221–230.

Peña, M. A., and S. H. Altmann, 2009. Use of satellite-derived hyperspectral indices to identify stress symptoms in an Austrocedrus chilensis forest infested by the aphid Cinara cupressi. *International Journal of Pest Management* 55, 197-206.

Petropoulos, G. P., C. Kalaitzidis, and K. Prasad Vadrevu, 2012. Support Vector Machines and Object-Based Classification for Obtaining Land-Use/Cover Cartography from Hyperion Hyperspectral Imagery. *Computers and Geosciences* 41, 99 - 107.

Pietrzykowski, E., C. Stone, E. Pinkard, and C. Mohammed, 2006. Effects of mycosphaerella leaf disease on the spectral reflectance properties of juvenile eucalyptus globulus foliage. *Forest Pathology* 36, 334-348.

Pontius, J., R. A. Hallet, and M. E. Martin, 2005. Assessing hemlock decline using visible and near-infrared spectroscopy: indices comparison and algorithm development. *Applied Spectroscopy* 59, 836–84.

Pontius Jr., R., and M. Millones, 2011. Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407-4429.

Pontius, R. G., and M. Millones, 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407-4429.

Pouteau, R., A. Collin, and B. Stoll, 2011. A comparison of machine learning algorithms for classification of tropical ecosystems observed by multiple sensors at multiple scales. *In* "34th International Symposium for Remote Sensing of Environment", Sydney.

Pu, R., P. Gong, G. Biging, and M. R. Larrieu, 2003. Extraction of Red Edge Optical Parameters from Hyperion Data for Estimation of Forest Leaf Area Index. *IEEE Transactions on geoscience and remote sensing* 41, 916-921.

Quackenbush, L. J., and Y. Ke, 2007. Investigating New Advances in Forest Species Classification. *In* "Proceedings of 2007 ASPRS Annual Conference," Tampa, Florida.

Radeloff, V. C., D. J. Mladenoff, and M. S. Boyce, 1999. Detecting jack pine budworm defoliation using mixture analysis: separating effects from determinants. *Remote Sensing of Environment* 69, 156 - 169.

RapidEye, 2011. Satellite imagery product specifications. *[online] Available at:* *http://www.RapidEye.de/upload/-RE_Product_Specifications_ENG.pdf* *[Accessed 01 Nov. 2011]*.

Redo, D. J., and A. C. Millington, 2011. A hybrid approach to mapping land-use modification and land-cover transition from MODIS time-series: a case study from the Bolivian seasonal tropics. *Remote Sensing of Environment* 115, 353-372.

Richter, K., C. Atzberger, T. B. Hank, and W. Mauser, 2012. Derivation of biophysical variables from Earth observation data: Validation and statistical measures. *J. Appl. Remote Sens* 6, 063557.

Richter, K., C. Atzberger, F. Vuolo, and G. D'Urso, 2011. Evaluation of sentinel-2 spectral sampling for radiative transfer model based LAI estimation of wheat, sugar beet, and maize. *IEEE Journal of Selected Topics in Applied Earth Observations in Remote Sensing* 4, 458-464.

Ringrose, S., W. Matheson, B. Mogotsi, and F. Tempest, 1989. Nature of the darkening effect in drought affected savannah woodland environments relative to soil reflectance. *Remote Sensing of Environment* 25, 519-524.

Riordan, C. J., 1981. Change detection for resource inventories using digital remote sensing data. *In* "Workshop on In-Place Resource Inventories: Principles and Practices", pp. 278-283. University of Maine, Orono, Maine.

Rodriguez-Galiano, V., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. Rigol- Sanchez, 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67, 93-104.

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering, 1973. Monitoring vegetation systems in the Great Plains with ERTS. *In* "In 3rd ERTS Symposium", pp. 309–317. NASA SP-351 I, Washington D.C.

Ruiz-Gazen, A., and N. Villa, 2007. Storms prediction: Logistic regression vs. random forest for unbalanced data *Case studies in business, industry and government statistics* 1, 91-101.

Rullan-Silva, C. D., A. E. Olthoff, J. A. Delgado de la Mata, and J. A. Pajares-Alonso, 2013. Remote Monitoring of Forest Insect Defoliation -A Review. *Forest Ecosystem* 22, 377-391.

Sah, S., 2013. A Multi-Temporal Fusion-Based Approach for Land Cover Mapping in Support Of Nuclear Incident Response, Pune University, Chester F. Carlson Center for Imaging Science of the College of Science Rochester Institute of Technology.

Santos, M., J. Greenberg, and S. Ustina, 2010. Using hyperspectral remote sensing to detect and quantify southeastern pine senescence effects in red-cockaded woodpecker (Picoides borealis) habitat. *Remote Sensing of Environment* 114, 1242-1250.

Sanz-Cortiella, R., J. Llorens, A. Escola`, J. Arno´-Santorra, M. Ribes-Dasi, J. Masip-Vilalta, F. Camp, F. Gracia-Aguila, F. Solanelles-batlle, S. Plana-DeMarti, T. Palleja-Cabre, J. Palacin-Roja, E. Gregorio-Lopez, I. Del-Moral-Martinez, and J. Rossel-Polo, 2011. Innovative LiDAR 3D dynamic measurement system to estimate fruit-tree leaf area. *Sensors* 11, 5769-5791.

Schlerf, M., and C. Atzberger, 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing. *Remote Sensing of Environment* 100, 281-294.

Schlerf, M., and C. Atzberger, 2012. Vegetation Structure Retrieval in Beech and Spruce Forests Using Spectrodirectional Satellite Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5, 8 - 17.

Schuster, C., M. Forster, and B. Kleinschmit, 2012. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing* 33, 5583-5599.

Sebego, R. J., and W. Arnberg, 2002. Interpretation of mopane woodlands using air photos with implication on satellite image classification. *International Journal of Applied Earth Observation and Geo-information* 4, 119–135.

Sebego, R. J., W. Arnberg, B. Lunden, and S. Ringrose, 2008. Mapping of Colophospermum mopane using Landsat TM in eastern Botswana. *South African Geographical Journal* 90, 41-53.

Seidl, R., P. M. Fernandes, T. F. Fonseca, F. Gillet, A. M. Jönsson, K. Merganicová, S. Netherer, A. Arpaci, J. D. Bontemps, and H. Bugmann, 2011. Modelling natural disturbances in forest ecosystems: A review. *Ecological Modelling* 222, 903-924.

Sims, D. A., and J. A. Gamon, 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing Environment* 81, 337–354.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10, 989-1003.

Smith, K. L., M. D. Steven, and J. J. Colls, 2004. Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks. *Remote Sensing of environment* 92, 207-217.

Somers, B., J. Verbesselt, E. M. Ampe, N. Sims, W. W. Verstraeten, and P. Coppin, 2010. Spectral mixture analysis to monitor defoliation in mixed aged Eucalyptus globules Labill plantations in southern Australia using Landsat 5TM and EO-1 Hyperion data. *International Journal of Applied Earth Observations and Geoinformation* 12, 270-277.

Souza, A. A., L. S. Galvao, and J. R. Santos, 2010. Relationships between Hyperion-derived vegetation indices, biophysical parameters, and elevation data in a Brazilian savannah environment. *Remote Sensing Letters* 1, 55-64.

Stack, J., A. Dorward, T. Gondo, P. Frost, F. Taylor, and N. Kurebgaseka, 2003. Mopane Worm Utilisation and Rural livelihoods in Southern Africa. *In* "Presentation for International Conference on Rural Livelihoods, Forests and Biodiversity", Bonn, Germany.

Statnikov, A., L. Wang, and C. Aliferis, 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics* 9, 1-10.

Stimson, H. C., D. D. Breshears, S. L. Ustin, and S. C. Kefauver, 2005. Spectral sensing of foliar water conditions in two co-occurring conifer species: Pinus edulis and Juniperus monosperma. *Remote Sensing of Environment* 96, 108−118.

Stone, C., and N. C. Coops, 2004. Assessment and monitoring of damage from insects in Australian eucalypt forests and commercial plantations. *Australian Journal Entomology* 43, 283– 292.

Strobl, C., and A. Zeileis, 2008. "Danger: High Power! Exploring the Statistical Properties of a Test for Random Forest Variable Importance." Department of Statistics, University of Munich, Munich, Germany.

Stumpf, A., and N. Kerle, 2011. Object - oriented mapping of landslides using random forests. *Remote sensing of environment* 115, 2564-2577.

Styles, C. V., and J. D. Skinner, 2000. The influence of large mammalian herbivores on growth form and utilization of mopane trees, Colophospermum mopane, in Botswana's Northern Tuli Game Reserve. *African Journal of Ecology* 38, 95-101.

Tapsall, B., P. Milenov, and K. Tasdemir, 2010. Analysis of RapidEye imagery for annual landcover mapping as an aid to European Union (EU) common agricultural policy. *In* "ISPRS Technical Commission VII Symposium - 100 Years" (ISPRS, ed.), Vol. Vol. XXXVIII pp. 568-573, Vienna, Austria. IAPRS.

Thenkabail, P. S., R. B. Smith, and E. D. Pauw, 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment* 71, 158－182.

Thomas, V., P. Treitz, J. H. McCaughey, T. Noland, and L. Rich, 2008. Canopy chlorophyll concentration estimation using hyperspectral and lidar data for a boreal mixedwood forest in northern Ontario, Canada. *International Journal of Remote Sensing* 29, 1029–52.

Tigges, J., T. Lakes, and P. Hostert, 2013. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of environment* 136, 66-75.

Timberlake, J., 1996. Colophospermum mopane - a tree for all seasons. *In* "Sustainable Management of Indigenous Forests in the Dry Tropics" (P. T. Mushove, Shumba, E.M. and Matose, F., ed.). Zimbabwe Forestry Commission/SAREC, Harare.

Townsend, P. A., K. N. Eshleman, and C. Welcker, 2004. Remote estimation of gypsy moth defoliation to assess variations in stream nitrogen concentrations. *Ecological Applications* 14, 504−516.

Treitz, P. M., and P. J. Howarth, 1999. Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems. *Progress in Physical Geography* 23, 359–390.

Tucker, C. J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of environment* 8, 127-150.

Ustin, S. L., S. Jacquemoud, A. Palacios-Orueta, L. Li, and M. L. Whiting, 2009. Remote sensing based assessment of biophysical indicators for land degradation and desertification,. *In* "Recent advances in remote sensing and geoinformation processing for land degradation assessment" (R. A., J., Hill ed.), pp. 15-44. CRC Press.

Vaiphasa, C., A. K. Skidmore, W. F. de Boer, and T. Vaiphasa, 2007. A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, 225-235.

Van der Sanden, J. J., A. Deschamps, S. J. Thomas, R. Landry, and R. J. Hall, 2006. Using MERIS to assess insect defoliation in Canadian Aspen Forests. *In* "International Geoscience and Remote Sensing Symposium 2006 and 27th Canadian Symposium on Remote Sensing,Remote Sensing-A natural Global Partnership", Denver, Colorado.

Vapnik, V., 1998. "Statistical learning theory. Support Vector Machines for Pattern Recognition," John Wiley & Sons, New York.

Veenendaal, E. M., K. B. Mantlana, N. W. Pammenter, P. Weber, P. Huntsman-Mapila, and J. Lloyd, 2008. Growth form and seasonal variation in leaf gas exchange of Colophospermum mopane savanna trees in northwest Botswana. *Tree Physiology* 28, 417-24.

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sensing of Environment* 16, 125-141.

Vina, A., A. A. Gitelson, A. L. Nguy-Robertson, and Y. Peng, 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment* 115, 3468-3478.

Vogelmann, J. E., B. N. Rock, and D. M. Moss, 1993. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing* 14, 1563–1575.

Waske, B., V. Heinzel, M. Braun, and G. Menz, 2007. Random forests for classifying multi-temporal sar data. *In* "Envisat Symposium", Montreux, Switzerland.

Weiskittel, A. R., and D. A. Maguire, 2007. Response of Douglas-fir leaf area index and litterfall dynamics to Swiss needle cast in north coastal Oregon, USA. *Ann For Sci* 64, 121-132.

Weiss, M., and F. Baret, 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment* 70, 293-306.

Weiss, M., F. Baret, R. B. Myneni, A. Pragnere, and Y. Knyazikhin, 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* 20, 3-22.

Wessels, I., C. Waal, and W. D. Boer, 2007. Induced chemical defences in Colophospermum mopane trees. *African Journal Of Range and Forage Science* 27, 141–147.

Williams, D. L., 1975. Computer analysis and mapping of gypsy moth defoliation levels in Pennsylvania using Landsat-1 digital data. *In* "NASA Earth Presentations Vol. 1-A", pp. 167-181.

Williams, D. L., and M. L. Stauffer, 1978. Monitoring gypsy moth defoliation by applying change detection techniques to Landsat imagery. *In* "Symposium on Remote Sensing for Vegetation Damage Assessment", pp. 221-229.

Williams, D. L., M. L. Stauffer, and K. C. Leung, 1979. A forester's look at the application of image manipulation techniques to multitemporal Landsat data. *In* "Fifth Annual Symposium on Machine Processing of Remotely Sensed Data: Purdue University, Lafayette", pp. 368-375.

Wolter, P. T., P. A. Townsend, B. R. Sturtevant, and C. C. Kingdon, 2008. Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario. *Remote Sensing of Environment* 112, 3971−3982.

Wulder, M., and S. Franklin, 2003. "Remote Sensing of Forest Environments: Concepts and Case Studies," Kluwer Academic, Boston.

Wythers, K. R., P. B. Reich, and D. P. Turner, 2003. Predicting leaf area index from scaling principles: corroboration and consequences. *Tree Physiology* 23, 1171-1179.

Yao, Y., Q. Liu, and X. Li, 2008. LAI retrieval and uncertainty evaluations for typical row-planted crops at different growth stages. *Remote Sensing of Environment* 112, 94−106.

Zarco-Tejada, P. J., J. R. Miller, G. H. Mohammed, and T. L. Noland, 2000. Chlorophyll fluorescence effects on vegetation apparent reflectance:Leaf-level measurements and model simulation. *Remote Sensing of Environment* 74, 582−595.

Zhou, L., R. Kaufmann, Y. Tian, R. Mynenin, and C. Tucker, 2010. Monitoring forest plant biodiversity changes and developing conservation strategies: a study from Camili Biosphere reserve area in NE Turkey. *Biologia* 65, 843-852.

Zhou, L., R. K. Kaufmann, Y. Tian, R. B. Mynenin, and C. J. Tucker, 2003. Relation between interannual variations in satellite measures of northern forest greenness and climate between 1982 and 1999. *Journal of Geophysical Research* 108, 4004.

Classification image of the best indices (NDVI-RE) for Chapter 5