The remote sensing of forest canopy gaps in a selectively logged submontane tropical forest reserve in Kenya

Colbert Mutiso Jackson

A thesis submitted to the Faculty of Science, at the University of Witwatersrand, in fulfilment of the academic requirements for the degree of Doctor of Philosophy in Geography and Environmental Sciences

November 2022
Johannesburg
South Africa

Abstract

Forests constitute 31% (about 4 billion ha) of the land area of the earth, and tropical forests cover is about 2 billion ha. Tropical forests play a significant role in supporting the earth's life and natural ecosystems. But many conservation and protection efforts have not been effective, as they are being cleared in many countries for timber and expansion of agricultural land. Few undisturbed tropical forests remain today, and unsustainable selective logging (SL) is probably the single biggest factor contributing to the global degradation of tropical forests. The amount of forest degradation that is not detected using currently available remote sensing (RS) techniques is unknown. Many methods used to map SL in tropical forests using low/medium spatial resolution datasets have a high rate of false detections. As a result, reliable and operational methods for monitoring SL in tropical forests ought to be utilized. Recently, very high resolution (VHR) RS datasets have caught the interest of researchers studying SL in tropical forests.

Therefore, this study was aimed to apply spectral-texture analysis approach to detect canopy gaps caused by illegal logging of *Ocotea usambarensis* (East African camphor) in Mt. Kenya Forest Reserve (MKFR) in Chuka, Tharaka Nithi County, Kenya using the VHR WorldView-3 (WV-3) satellite data. Several features were derived from the WV-3 data—however, a large number of features results to longer computing time, and might result in reduced classification accuracy. Therefore, feature ranking measures—the mean decrease accuracy, the mean decrease Gini, and the pairwise feature correlation were used. The Jeffries-Matusita distance, the transformed divergence index, the G-statistic, and the Euclidean distance were used to calculate the separability of the forest landscape classes. First, the study reviewed and discussed RS techniques used to map SL in the tropical forests. Second, the threatened trees species (TS) in the selectively logged MKFR were mapped. Third, gaps in the forest canopy were detected and quantified using two approaches—initially, only spectral features were used to detect gaps in the forest canopy. A total of 55 spectral features were extracted from the WV-3 dataset—23 means (of 15 vegetation indices–VIs and 8 visible-near-infrared–VNIR bands), and 23 standard deviations–SDs (of 15 VIs and 8 VNIR bands). Also extracted were 8 ratios (of 8 VNIR bands), and 1 brightness feature (average of the means of bands 1 to 8). The study also explored the potential of rich textural features combined with color to model canopy gaps using GLCM-, LBP-, and MLBP-based rotation-invariant feature descriptors derived from WV-3 imagery. Due to their excellent performance and clear logic, two advanced machine learning (ML) classification models—the random forest (RF) and support vector machine (SVM) models were used to identify and classify canopy gaps in the spectral and texture domains of the WV-3 data. During the training process the learning parameters of RF (*mtry* and *ntree*) and SVM (γ and C) algorithms were optimised to obtain the best possible settings. Finally, the study reviewed Kenya's forest policy and law on participatory forest management (PFM).

The best tree species classification results reported F1-scores of 68.56 ± 2.6% and 64.64 ± 3.4% for RF and SVM, respectively. The RF and SVM models used to map canopy gaps using the spectral features reported average overall accuracies (OAs) of 92.3 ± 2.6% and 94.0 ± 2.1%, respectively. Average kappa coefficients (κ) were 0.88 ± 0.03 for RF and 0.90 ± 0.02 for SVM. The user's accuracy (UA) and producer's accuracy (PA) were in the range of 84–100%. The OA for the classification of canopy gaps using textural/spectral features reported values between 80 (RF, block F's MLBP/ASM) and 95.1% (SVM, block E's MLBP/CON). The average OA scores were 84.68 ±3.1, 84.54 ±2.5, 84.86 ±3.0, 86.46 ±3.9, 87 ±4.0, and 85.44 ±3.7 for image blocks A, B, C, D, E, and F, respectively, for the RF classifier, and 85.44 ±3.6, 87.2 ±1.8, 86.3 ±4.3, 89.84 ±2.8, 87.28 ±4.5 and 86.12 ±3.6 for the SVM classifier. The

UA and PA were in the range of 67-75% and 77-100% for the univariate LBP and MLBP models, respectively. Texture fused with colour resulted to higher classification accuracies.

Overall, the approach used in this study demonstrated improved ability of VHR satellite data and ML classification models to accurately map fine canopy gaps resulting from SL. Knowledge about where canopy gaps are located and how they are distributed is critical in accurate estimation of carbon densities of forests, and also for managing the proliferation of invasive species, among other applications. LiDAR datasets acquire the three-dimensional (3-D) structure of forest vegetation—repeat surveys can thus detect the removal of individual trees. The integration of optical images and LiDAR data may boost canopy gap classification.

Preface

This doctoral thesis describes work undertaken as part of a program of study at the School of Geography, Archaeology, and Environmental Studies, University of the Witwatersrand, Johannesburg from January 2018 to January 2022, under the supervision of Professor Elhadi Mohammed I. Adam (School of Geography, Archaeology, and Environmental Studies, University of the Witwatersrand, Johannesburg).

This document is the result of my investigations, therefore I have exercised reasonable care to ensure that this work is original, and where otherwise stated, all used sources of literature are listed at the end of each chapter. This doctoral thesis has not previously been submitted/accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Colbert Mutiso Jackson ______________________ Date: 11/11/2022

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

Prof. Elhadi Mohammed I. Adam ______________________ Date: 11.11.2022

Declaration 1-Plagiarism

I, Colbert Mutiso Jackson, 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 is detailed in the thesis and the References section.

Signed ____________________

Declaration 2-Publication and manuscripts

1. **Jackson,** C.M., and Adam, E. (2020). Remote sensing of selective logging in tropical forests: current state and future directions. *iForest - Biogeosciences and Forestry*, 13, 286–300.
2. **Jackson,** C.M., and Adam, E. (2021). Machine learning classification of endangered tree species in a tropical submontane forest using WorldView-2 multispectral satellite imagery and imbalanced dataset. *Remote Sensing*, 13, 4970.
3. **Jackson,** C.M., and Adam, E. (2022). A machine learning approach to mapping canopy gaps in an indigenous tropical submontane forest using WorldView-3 multispectral satellite imagery. *Environmental Conservation*, 49, 255–262.
4. **Jackson,** C.M., and Adam, E. (in review). Feature extraction and classification of WorldView-3 imagery using GLCM - and MLBP-based rotation-invariant feature descriptors for detection of canopy gaps. *Tropical Conservation Science*.
5. **Jackson,** C.M., and Adam, E. (in review). An assessment of forest policy and law on participatory forest management for sustainable forest management in Kenya. *Restoration Ecology*.

Dedication

To my dear wife, Charity, and our adorable children, Lulu and Lamont for their constant support and prayers for my success

Acknowledgments

This thesis represents the culmination of a process made possible by the contributions of numerous institutions and individuals, for which I am grateful. I would like to express my heartfelt appreciation to the people and organizations listed below for their contributions to this project. The research was carried out at the University of the Witwatersrand's School of Geography, Archaeology, and Environmental Studies, Faculty of Science, in Johannesburg, South Africa. I am extremely grateful to this institution for covering my tuition and providing me with research facilities as well as the time to work on this study.

I am extremely grateful to my supervisor, Prof. Elhadi Mohammed I. Adam, for accepting me as a Ph.D. student and guiding me through the entire Ph.D. process. Throughout my time at the University of the Witwatersrand, he advised this thesis and mentored me. His broad knowledge and deep insights have been a valuable asset to my success. I enjoyed working on such an interesting and complex topic as the use of very high resolution (VHR) WorldView-3 imagery to detect selective logging in Kenya's Mount Kenya Forest Reserve. Thank you for your moral support, encouragement, and inspiration to complete the research. It was always a relief to share taxing issues with you because you always had a solution and a very encouraging attitude. Your suggestions and criticisms influenced the final shape of the thesis. Thank you, Prof. Elhadi Mohammed I. Adam, for allowing me to teach GEOG 3017: Introduction to GIS and Remote Sensing as a lab assistant, tutor, and lecturer. I also want to thank the leadership, lecturers, and support staff in the School of Geography, Archaeology, and Environmental Studies for their help and enabling learning environment.

I would like to thank the Kenya Forest Service leadership at the Head Office in Nairobi, particularly the office of the Chief Conservator of Forests, for allowing me to collect data in Mount Kenya Forest Reserve. Mr. James Mburu, Tharaka Nithi County Ecosystem Conservator, and Mr. John Maina, Chief Forester, Chuka Forest Station, deserve special thanks for making it possible for me to visit Mt. Kenya Forest Reserve. I would also like to thank Mwebia and Keter for showing me around the research area and providing armed security while I was collecting field data. They were extremely dedicated to their work. They were extremely helpful and made a significant contribution to this research. I am eternally grateful to both. Thank you for agreeing to be my field assistant, Mr. David Kithinji. You worked tirelessly.

Many thanks to my family for their unwavering support while I was studying. Thank you from the bottom of my heart for your love, care, understanding, and encouragement. Your help has been unwavering. You have always shown your affection. I am especially grateful to

Charity, my dear wife, for her moral support. Thank you, Charity, for being a devoted wife, a true friend, and supportive during all of the difficulties we faced while I was away at school. My Ph.D. would not have been completed without her unwavering support. My success is her reward.

Despite coming from different countries, backgrounds, cultures, and age groups, I am grateful to my fellow Ph.D. candidates for the challenges and "home feeling" they provided. This provided me with a reasonably comfortable and conducive study environment. While I may not be able to thank everyone, I would like to express my gratitude to everyone who contributed to the success of my studies in some way.

Finally, I thank God for the gifts of life, good health, and wisdom, which allowed me to study uninterrupted.

Table of Contents

List of figures

List of tables

Abbreviations and acronyms

AGB:	Above-Ground Biomass
AIGF:	Area-Integrated Gap Fraction
ALS:	Airborne Laser Scanning
ARI:	Anthocyanin Reflectance Index
ASM:	Angular Second Moment
AutoMCU:	Automated Monte Carlo Unmixing
CARI:	Chlorophyll Absorption Ratio Index
CCA:	Contextual Classification Algorithm
CFA:	Community Forest Association
CIPFA:	The Chartered Institute of Public Finance and Accountancy
CLAS:	Carnegie Landsat Analysis System
CON:	Contrast
COR:	Correlation
CRI-1:	Carotenoid Reflectance Index 1
CRI-2:	Carotenoid Reflectance Index 2
CV:	Cross-Validation
DAP:	Digital Aerial Photograph
ENT:	Entropy
EO:	Earth Observation
ETM+:	Enhanced Thematic Mapper Plus
GFW:	Global Forest Watch
GLCM:	Grey Level Co-occurrence Matrix
GoK:	Government of Kenya
GPS:	Global Positioning System
GSFD:	Gap Size-Frequency Distribution
HOM:	Homogeneity
HR:	High Resolution
HYDICE:	Hyperspectral Digital Collection Experiment
IEA:	Institute of Economic Affairs
ITC:	Individual Tree Crown
J-M:	Jeffries-Matusita
κ:	Kappa
KFS:	Kenya Forest Service
KWS:	Kenya Wildlife Service
LAI:	Leaf Area Index
LBP:	Local Binary Pattern
LDA:	Linear Discriminant Analysis
LiDAR:	Light Detection and Ranging
MCARI:	Modified Chlorophyll Absorption Ratio Index
MDA:	Mean Decrease in Accuracy
MDG:	Mean Decrease in Gini
MDS:	Multi-Dimensional Scaling
MENR:	Ministry of Environment and Natural Resources
MKFR:	Mt. Kenya Forest Reserve
ML:	Machine Learning
MLBP:	Multivariate Local Binary Pattern
MSAVI:	Modified Soil-Adjusted Vegetation Index

MSAVIaf: Modified Soil Adjusted Vegetation Index aerosol free
NBR: Normalized Burn Ratio
NDFI: Normalized Difference Fraction Index
NDVI: Normalized Difference Vegetation Index
NEMA: National Environment Management Authority
NIR: Near Infra-Red
NPCI: Normalized Pigment Chlorophyll Index
NPV: Non-Photosynthetic Vegetation
OA: Overall Accuracy
OBIA: Object-Based Image Analysis
OLI: Operational Land Imager
OOB: Out-Of-Bag
PA: Producer's Accuracy
PFM: Participatory Forest Management
PRI: Photochemical Reflectance Index
PSND: Pigment-Sensitive Normalized Difference
PSRI: Plant Senescence Reflectance Index
RADAR: Radio Detection and Ranging
RBF: Radial Basis Function
RENDVI: Red Edge Normalized Difference Vegetation Index
REPI: Red-Edge Position Index
RF: Random Forest
RIL: Reduced Impact Logging
RGB: Red, Green, Blue
RMSE: Root Mean Square Error
RS: Remote Sensing
SAM: Spectral Angle Mapper
SAR: Synthetic Aperture Radar
SDI: Shadow Detection Index
SFM: Sustainable Forest Management
SIFI: Secretariat for International Forestry Issues
SIPI: Structurally Insensitive Pigment Index
SL: Selective Logging
SRred/green: Simple ratio red/green
SMA: Spectral Mixture Analysis
SVM: Support Vector Machine
TD: Transformed Divergence
TM: Thematic Mapper
TS: Tree Species
UA: User's Accuracy
UNESCO: United Nations Educational, Scientific and Cultural Organization.
VARI: Visible Atmospherically Resistant Index
VHR: Very High Resolution
VI: Vegetation Index
VNIR: Visible and Near Infra-Red
WV-3: WorldView-3
WV-2: WorldView-2

CHAPTER ONE

General introduction

1.1 Status of tropical forests

Tropical forests are home to more than half of the Earth's species diversity and provide several essential biological services (Solberg *et al*., 2008). They control global weather patterns and, more importantly, play an important role in the global carbon cycle (Da Ponte *et al*., 2015), storing large amounts of carbon while producing large amounts of the Earth's oxygen (Solberg *et al*., 2008). Many indigenous cultures and peoples rely on tropical forests and the land they occupy (Solberg *et al*., 2008). Despite substantial attempts to create novel methods to identify deterioration, little is known about the pace and severity of tropical forest degradation. However, some researchers estimate that unsustainable selective logging (SL) is destroying about 20% of humid tropical forests (Asner *et al*., 2009). This uncertainty, in turn, has made it difficult to understand and estimate the effects of forest degradation on tropical biodiversity (Borrego and Skutsch, 2014). Thus, reliable and operational systems, such as remote sensing (RS), are required to monitor selective logging (SL) in tropical forests.

1.2 Defining Selective logging and canopy gap

1.2.1 Selective logging

According to Laurance *et al*. (2014), the term "selective logging" involves the felling of trees in large-scale for industrial use, as well in small-scale for local use. In this thesis report, selective logging is defined as a type of timber extraction in which a tree(s) from a selected high-value species, are harvested without necessarily displaying any logging infrastructure (Asner *et al*., 2005; Andersen *et al*., 2014).

1.2.2 Canopy gap

The concept of "gaps" and their proper definition and measurement have been the subject of some discussion. Different authors have proposed various methods for defining and measuring canopy gaps. The extended gap concept is used by Runkle (1981) in defining a canopy gap as covering the stems of surrounding trees. But for RS purposes a tree trunk's location may not be easily determined from the top, because the location of the stem's base may not necessarily correspond with the centre of the crown (Gaulton and Malthus, 2010). According to Gaulton and Malthus (2010), ‘a gap boundary is defined by a line at ground level (the drip-line) located vertically beneath the inner most point reached by the foliage of a tree crown at any level, at that point on the gap perimeter.’ A ground level line that runs vertically beneath the innermost point of a tree crown is how Gaulton and Malthus (2010) define a gap boundary. Brokaw (1982) defines a canopy gap as a 'hole' in a forested area reaching an average height of about 2 m, and the smallest size of a gap—20 to 40 m^2. Gaps are defined by Fox *et al*. (2000) in relation to the difference in height from the bordering canopy. A gap, according to Koukoulas and Blackburn (2004), is a 'hole' in the canopy of forest due to the removal of a tree(s). Gaulton and Malthus (2010) state that a gap has a minimum area ≥5 m^2 and height ≤ 10 m. Getzin *et al*. (2014), mapped 1 m^2 canopy gaps using unmanned aerial vehicle (UAV) data—pixel size of 7 cm. Gaps in a forest measure at least width of 2 m, a maximum vegetation height of 3 m, and a minimum area of 50 m^2 (Bonnet *et al*., 2015). Therefore, the minimum size of a gap and gap closure height, normally varies from forest to forest, and affected by factors such as the size of tree crowns, understory height, and advance regeneration, among others (Gaulton and Malthus, 2010).

Canopy gaps are usually small (<1000 m^2) (Betts *et al*., 2005). In the past, the gap's size was determined by first measuring its length and width from the ground, and then computing its area using a circle or ellipse's formula (e.g. Runkle, 1992). Ground-based techniques can be challenging, especially in rough terrain. Furthermore, the results of ground surveys can be subjective—depending on the surveyor (e.g. Nakashizuka *et al*., 1995). Additionally, in field surveys it can be challenging to distinguish fine canopy gaps and the ones where the understory vegetation is dense. However, suggestions have been proposed to increase the ground surveys' objectivity—such as measuring the height of canopy at specific locations on a predetermined grid (Brokaw and Grear, 1991) or conducting regular line-intersect sampling (Battles *et al*., 1996). Nonetheless, these techniques require a sizable amount of time and manpower (Nakashizuka *et al*., 1995). The most appropriate solution was to observe forests from above rather than below through RS. Satellite and airborne data with very high-resolution (VHR) are appropriate for precisely delineating forest canopy gaps, as well as individual tree crowns.

Only gaps with a minimum area of 100 m^2 were considered in this study since it is such gaps that are significant for carbon dynamics (Negrón-Juárez *et al*., 2011), and to rule out small gaps that were not likely caused by felled *Ocotea* usambarensis (East African camphor) trees.

1.3 Detection of canopy gaps in tropical forests

A challenging task is the detection and delineation of a gap (Vepakomma *et al*., 2008). According to Nyamgeroh (2015), there have been few studies on canopy gap delineation, and even fewer are the methods to delineate canopy gaps. Forest canopy gaps have been studied using Landsat data with either the mono-temporal (e.g. Asner *et al*., 2005 and Negrón-Juárez *et al*., 2011)) or the multi-temporal approaches (e.g. Wang *et al*., 2019). However, Landsat's pixel size (30 m) is incapable of detecting small canopy gaps (Clark *et al*., 2004). Additionally, the detectability of the effects of SL on Landsat images is 1 to 3 year after the disturbance event (Costa *et al*., 2019). High resolution (HR) multispectral sensors, e.g. SPOT and IKONOS were used to solve the spatial resolution problem (Clark *et al*., 2004).

Different mapping methods depend on their ability to identify gaps in the forest canopy. Small canopy gaps can be detected using very high resolution (VHR) optical and light detection and ranging (LiDAR) data (Asner *et al*., 2013). Calibrated VHR synthetic aperture radar (SAR) data have been promising in detecting canopy gaps created through the removal of individual trees (Mitchell *et al*., 2017). Traditional aerial photography was successfully used for mapping canopy gaps before the introduction of HR optical data (Malahlela *et al*., 2014). But these methods fall short in capturing crucial biochemical traits like chlorophyll content, which are crucial for canopy gap mapping (Malahlela *et al*., 2014). Furthermore, its data acquisition is cost-intensive, but technological advances have revitalized its use through UAVs. Spaias *et al*. (2016) and Ota *et al*. (2019) used UAV data acquired before and after logging to map small-scale canopy gaps. An increase in shadow was used as an indicator of the location of logging events by Spaias *et al*. (2016). According to Ota *et al*. (2019), change of above-ground biomass (AGB) was related to disappearing tree crowns due to SL. Spaias *et al*. (2016) used hyperspectral data—its application is limited because it is computationally demanding particularly if cloud computing resources were not available.

A canopy structure can be described in three-dimensional (3-D) by use of a canopy height model (CHM). Several studies, e.g. Asner *et al*. (2013), Boyd *et al*. (2013), Kent *et al*. (2015), Wedeux and Coomes (2015), and Dalagnol *et al*. (2019) have applied CHMs. The CHMs were derived from LiDAR data from a single acquisition to delineate canopy gaps with exceptional accuracy and consistency in tropical areas. Andersen *et al*. (2014) used repeat LiDAR data and applied a simple differencing of LiDAR CHMs to detect disappearing tree crowns. Airborne LiDAR data have been used to acquire precise measurement of AGB stock and change—several studies used a single acquisition, e.g. D'Oliveira *et al*. (2012), Ellis *et al*. (2016), Melendy *et al*. (2018), and Pearson *et al*. (2018). Others used before and after logging datasets, e.g., Englhart *et al*. (2014), Silva *et al*. (2017), Pinagé *et al*. (2019), and Rex *et al*. (2020). Nonetheless, airborne LiDAR's coverage of relatively small spatial extents and high data acquisition costs limits its application, especially in tropical areas.

Descals *et al*. (2017) proposed an automated processing workflow in bi-temporal VHR SmallSat imagery to enable the detection of SL. Dalagnol *et al*. (2019) compared the potential of VHR satellite data to that of airborne LiDAR to detect tree loss.

According to Yang *et al*. (2015), broad-band multispectral images can yield promising results for identifying canopy gaps. The visual interpretation of VHR multi-date satellite imagery is a promising way to detect and quantify canopy gaps created by logging activities with fairly low uncertainty (Clark et al., 2004). Nonetheless, spatially precise validation data is not readily available, and automated approaches based on VHR satellite data to accurately map canopy gaps over expansive areas are lacking (Dalagnol *et al*., 2019).

In order to quantify illegal logging in a tropical forest, this study was aimed to evaluate whether gaps in forest canopies can be detected and mapped using the very high resolution (VHR) WorldView-3 multispectral imagery. High resolution (HR) WorldView-2 image and Google Earth were used to provide historical data. Accurate quantification of canopy gaps from disappearing tree crowns has a crucial contribution in calculating carbon densities of forests, as well as modelling the effects of forest degradation on tropical biodiversity. Selective logging (SL), a major cause of forest degradation in tropical forests, results to long-term carbon emissions into the atmosphere. The accuracy of carbon estimates can be improved, provided canopy gaps are accurately identified and correctly estimate trees in per unit area of forest. This can be used by forest managers to implement on-the-ground conservation and restoration projects in Mt. Kenya Forest Reserve (MKFR).

1.4 Research objectives

The primary focus of this study was to use a remote sensing (RS) technique—the spectral-texture analysis approach—to explore the potential of the very high resolution WorldView-3 (WV-3) multispectral dataset to identify and quantify canopy tree loss associated with selective logging (SL) in a closed-canopy submontane tropical forest. The gaps in the study area were vegetated—the canopy gaps had low vegetation inside them—an initial stage of vegetation recovery from SL disturbance. Two advanced machine learning (ML) classification models—the random forest (RF) and support vector machine (SVM) models were used to identify and classify canopy gaps in the spectral and texture domains of the WV-3 multispectral imagery. The specific objectives were:

1. To review the status, trends, potentials, and challenges and recommended future directions in remote sensing (RS) techniques used to map selective logging (SL) in the tropical forests from 1992 to 2019;
2. To use species-level tree classification to map threatened trees species in a selectively logged sub-montane heterogeneous tropical forest—MKFR;
3. To explore the potential of spectral features from the visible-near-infrared bands and vegetation indices to model canopy gaps caused by logging of *Ocotea* trees;
4. To apply a texture-spectral analysis approach—using GLCM-, LBP-, and MLBP-based rotation-invariant feature descriptors—to exploit the potential of the very high resolution WorldView-3 multispectral imagery to model canopy gaps; and
5. To assess how inadequacies in Kenya's forest policy and law on participatory forest management led to illegal activities in forests in Kenya.

1.5 Scope of the study

This study used multispectral RS data to discriminate among tree species under the threat of SL in Mt. Kenya Forest Reserve (MKFR) in Chuka, Tharaka Nithi County, Kenya. The applications of spectral features extracted from the WV-3 VNIR bands and vegetation indices (VIs) were used to detect canopy gaps resulting from SL of *Ocotea usambarensis* (East African camphor). Fused texture and spectral features extracted using GLCM-, LBP-, and MLBP-based models were also explored to detect gaps in forest canopy. Two machine learning (ML) classification algorithms—random forest (RF) and support vector machine (SVM) were used. The study collected a field dataset of georeferenced canopy gaps in the study site, and historical data—Google Earth and WorldView-2 datasets. The Participatory Forest Management (PFM) policy and law were then evaluated in order to determine the efficacy of changes made to Kenya's policies, legislation, and institutional arrangements to support transitions toward Sustainable Forest Management (SFM).

Mt. Kenya Forest Reserve (MKFR) is an Afromontane tropical forest ecosystem that supports a diverse range of biodiversity and provides valuable ecosystem services, as well as being a major forested water catchment area in Kenya (MENR, 2016). For decades, MKFR has been subjected to illegal logging for its commercially valuable reserves of indigenous timber, as well as other illegal activities that have resulted in reduced ecosystem services (KWS, 2010).

1.6 The study area

The forest reserve was established in 1932 and placed under the management of the Forest Department (now known as the Kenya Forest Service) with the primary goal of preserving and developing the forest. This included creating plantations to replace harvested indigenous stands, regulating resource access, and preserving a forest industry (KWS, 2010). Because of its remarkable ecosystems and natural beauty, the forest reserve was designated a UNESCO World Heritage Site in 1997 (NEMA, 2010). Mt. Kenya Forest Reserve (MKFR) is located in Central Kenya and includes parts of the counties of Meru, Tharaka-Nithi, Embu, Kirinyaga, and Nyeri. Chuka Forest, located in Tharaka-Nithi County, is part of the Mt. Kenya ecosystem and encompasses approximately 21,740 ha (Figure 1.1). Six locations—Kiang'ondu, Magumoni, Mitheru, Mugwe, Mwonge, and Thuita—make up the forest-adjacent area. Chuka Forest is located between longitudes $37^0$19'0"E and $37^0$36'0"E, and latitudes $0^0$11'0"S and $0^0$19'30"S. The MKFR spans a range of elevation, slope, and aspect positions. The bedrock of the study area is formed by phonolites from volcanic events about 2 million years ago, but the inorganic part of the soils was formed from volcanic ashes and pyroclastic rocks (Lange *et al*., 1997). In the 2700-3300 m zone, dark andosols with low bulk density and high organic matter content predominate. The topsoil has a high silt content, whereas the B-horizon has a high clay content (Lange *et al*., 1997).

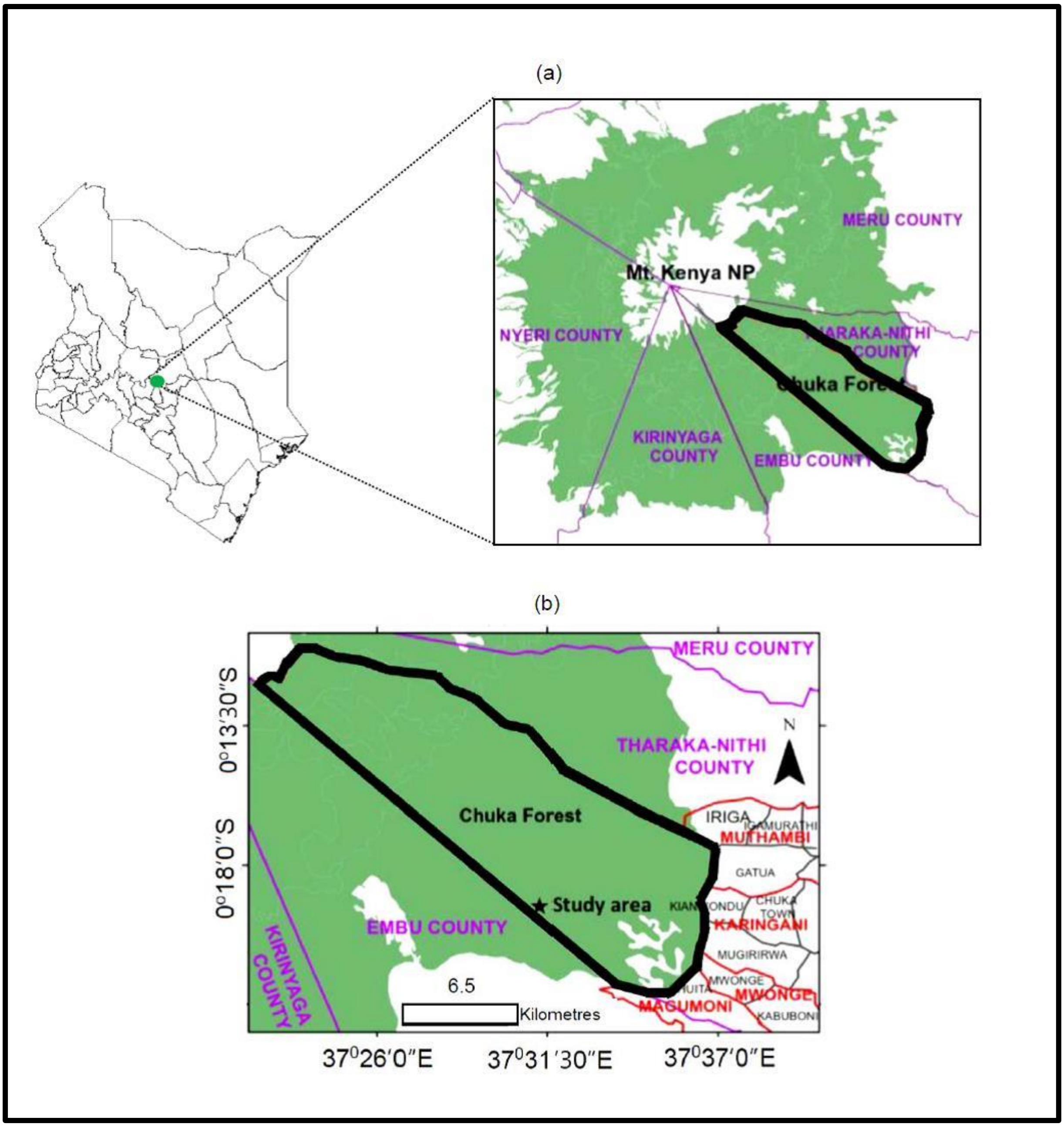


Figure 1.1. *Map of the study area in Mt. Kenya Forest Reserve (MKFR) showing the location of (a) Chuka Forest in MKFR, and (b) Chuka Forest and the adjacent locations.*

Annual rainfall follows a bimodal distribution—March-May are the long rains whereas October-December are the short rains (Figure 1.2). Rainfall is light on the lower slopes but heavy at higher altitudes (KFS, 2010). The highest rainfall (about 2500 mm) occurs between 2750 and 3750 m above sea level, while above 4500 m, most precipitation falls as snow or hail, and frosts are common above 2500 m above sea level (KFS, 2010). January and February are the driest months, and the trade wind system has the greatest impact on the windward side (Wooller *et al*., 2002). Large diurnal oscillations characterize the climate, but the annual amplitude in the monthly means is small (Lange *et al*., 1997). The eastern and southern sides are wetter (2500 mm), while the northern side is drier (900 mm) (Nyongesa and Vacik, 2019).

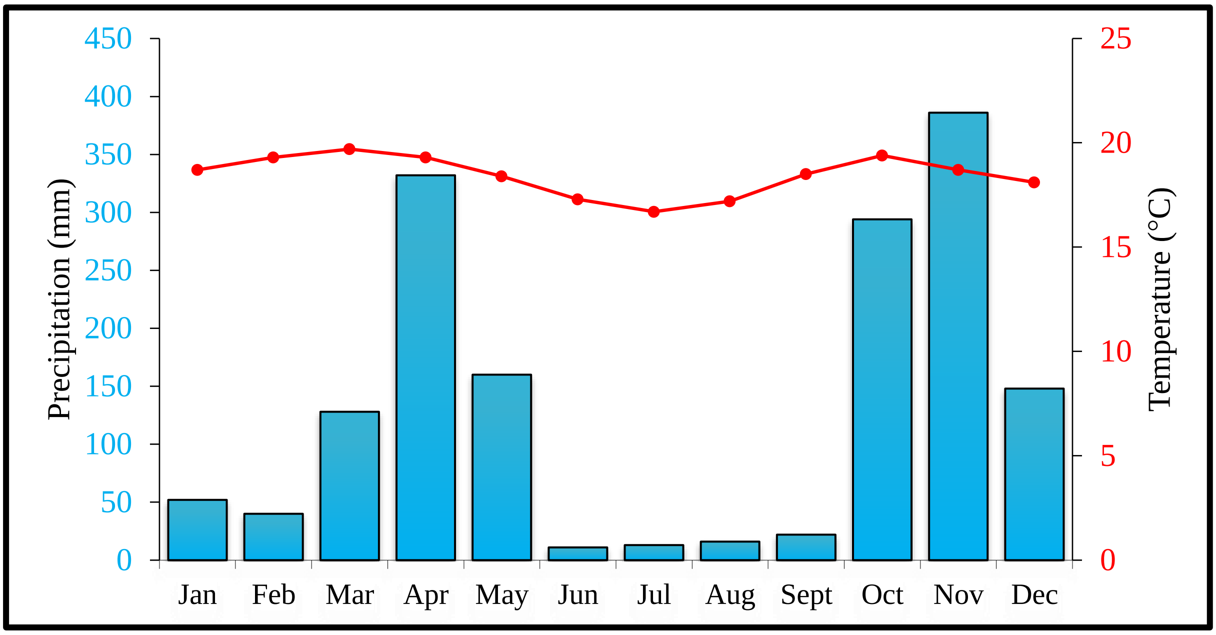


Figure 1.2. *The climate of the study area; mean monthly rainfall (bars) and mean monthly temperature (red dotted line).*

Mt. Kenya shows a marked vegetational gradient dictated by altitude and rainfall amount. The lower tree line of the forest belt is due to agricultural and pastoral activities (Lange *et al*., 1997). In the southern parts of the mountain, cultivated land extends up to nearly 1800 m, in the eastern and western approximately 2400 m, and in the northern it is about 2900 m. About 3400 m in the south and west of the mountain lies the upper tree line whereas in the north the tree line is at 3000 m (Lange *et al*., 1997). The vegetation on the lower slopes of Mt. Kenya is montane forest. Characteristic tree species in the indigenous forest include *Podocarpus latifolia*, *Nuxia congesta*, *Newtonia buchananii*, *Calodendrum capense*, *Croton megalocarpus*, *Juniperus procera*, *Ocotea usambarensis*, and *Olea europaea spp africana*. Then next to the montane forest is the bamboo forest (which extends from 2,550 m), scrub, and moorland (from 3,000 to 3,500 m). Characteristic species in this zone include *Hagenia abyssinica* and *Hypericum spp*. The moorland gives way to Afro-Alpine vegetation. The bare rock, ice, and snow are at the highest altitude (Nature Kenya, 2019). *Ocotea usambarensis*, normally targeted for its hardwood which has excellent decay and insect resistance, also forms the evergreen sub-montane forests on the extremely humid southern, south eastern, and eastern slopes at 1500 to 2500 m. They never form pure stands and prefer humid Niti- and Acrisols

(Bussmann and Beck, 1995). Nyayo Tea Zone, established by Legal Notice No. 265 of 1986 provides a buffer zone to check against human encroachment into MKFR.

1.7 Thesis overview

The thesis aims to achieve the main objectives of this study through five research papers that have been submitted to international peer-reviewed journals, three of which have already been published, one is being reviewed, and the other one is being prepared for journal submission. Each was written as a stand-alone article that can be read independently of the other parts of the thesis but draws separate conclusions that are linked to the overall research objectives and questions. As a result, there are overlaps and replications in the sections "Introduction" and "Method" in the various chapters. When one considers the rigorous peer-review procedure and the fact that the various chapters are papers that can be read independently without losing the overall context, this problem is of relatively minor importance. The thesis is divided into seven chapters.

The second chapter conducts a pantropical meta-analysis to determine the current state of knowledge on remote sensing techniques used to detect and monitor SL in tropical forests. The chapter examines and discusses the current state of RS techniques used to detect and monitor SL disturbances in tropical forests. The analyses focused on the years 1992-2019. The number of studies on RS for SL has steadily increased over time, necessitating their review to inform current research and forest management techniques. A variety of peer-reviewed articles are discussed to assess the applicability and accuracy of various methods in various situations. Major issues with existing approaches are identified, and future needs were discussed.

Using RF and SVM classifiers and WorldView-2 (WV-2) multispectral imagery, Chapter 3 examines the effects of imbalanced data on identifying and mapping tree species under threat in a selectively logged sub-montane heterogeneous tropical forest. Three datasets derived using three techniques—oversampling, undersampling, and combined sampling techniques were compared with the original dataset. The chapter lists the key spectral bands that were determined to be crucial in mapping the endangered tree species in the research area.

Chapter 4 detects canopy gaps caused by illegal logging of *Ocotea* trees using very high-resolution multispectral satellite data and machine learning algorithms, as well as historical data from WV-2 and Google Earth. Thus, the role of vegetation indices derived from WorldView-3 data in mapping canopy gaps was investigated.

The Grey-Level Co-Occurrence Matrix (GLCM), which was fused with the Multivariate Local Binary Pattern (MLBP) to detect canopy gaps, is the focus of Chapter 5. It is nearly 50 years old and was used for the analysis of grey level texture based on statistical approaches. Three WorldView-3 (WV-3) bands' worth of spectral and textural data were combined to enhance classification outcomes.

Chapter 6 assesses Kenya's forest policy and law on Participatory Forest Management (PFM) to know the extent to which adjustments in laws, regulations, and institutional structures intended to speed up the transition to sustainable forest management (SFM) have failed. Several steps have been proposed to simplify and rationalize Kenya's forest policies and laws.

The study is summarized in Chapter 7. The findings and conclusions from the preceding chapters are summarized. Some suggestions for future research on the use of remote sensing to map canopy gaps in tropical forests are made.

At the end of the thesis, there is a single reference list.

CHAPTER TWO

Literature review

This chapter is based on:

Jackson, C.M., and Adam, E. (2020). Remote sensing of selective logging in tropical forests: current state and future directions. *iForest - Biogeosciences and Forestry,* 13, 286–300.

Abstract

This chapter reviewed and discussed the status of remote sensing (RS) techniques applied in detecting and monitoring selective logging (SL) disturbance in tropical forests. The number of studies on remote sensing (RS) for selective logging (SL) has grown steadily over the years, therefore, there was need for an updated review to guide forest management practices and current research. The analyses concentrated on the period 1992-2019. Accurate and precise detection of selectively logged sites in a forest is crucial for analysing the spatial distribution of forest disturbances and degradation. Remote sensing (RS) can be used to monitor SL activities over tropical forests, which otherwise require labour-intensive and time-consuming field surveys. Peer-reviewed articles were searched in *ScienceDirect*®, *Web of Science*®, and *Scopus*® web-based databases containing the largest abstract and citation databases of peer-reviewed literature. The key terms used were ("Selective Logging" OR "Canopy Gap" OR "Canopy Opening" OR "Deforestation" OR "Forest Degradation" OR "Forest Disturbance") AND "Tropical Forests" AND "Remote Sensing" in the title, abstracts and the keywords in the search. About 5546 articles were found in *ScienceDirect*, 2395 in *Web of Science*, and 888 articles in *Scopus* (as of June 20, 2019). After considering the relevant articles, abstracts of 328 articles were read through—information was extracted from 110 relevant articles that discussed the RS of SL in tropical forests.

The literature review showed that there was increased publication activity over the years. This could be related to the availability of information, awareness about RS technology, and growing attention around the monitoring of SL and related activities in tropical forests. A large number of studies (54%) used low/medium spatial resolution datasets to map SL in tropical forests. Only 28% and 18% used very high resolution (VHR) and high-resolution (HR) sensors, respectively. Low and medium spatial resolution datasets are too coarse to detect less intensive SL. Many of the methods for mapping SL in tropical forests using low/medium spatial resolution datasets have a high rate of false detections. Recently, VHR satellite and unmanned aerial vehicle- (UAV-) based datasets have caught the interest of researchers mapping disappearing tree crowns due to SL tropical forests.

Keywords: Tropical Forest Disturbance. Selective Logging. Forest Degradation. Forest Canopy Gaps. Disturbance Mapping. Remote Sensing. Forest Monitoring.

2.1 Introduction

Tropical forests are housing more than half of the diversity of life on earth and offering several vital biological applications such as pest control, pollination, and seed dispersal (Solberg *et al*., 2008). Tropical forests regulate global weather patterns, and more importantly, play a key role in the global carbon cycle (Da Ponte *et al*., 2015), keeping large amounts of carbon and producing great supplies of the Earth's oxygen (Solberg *et al*., 2008). Tropical forests and the land they occupy support numerous indigenous cultures and peoples (Solberg *et al*., 2008). Despite the remarkable values and functions of tropical forests, many conservation and protection efforts have not been effective, as a result, the forests are being cleared in many countries for timber and agricultural land expansion (Gibson *et al*., 2011). Forest disturbance from selective logging (SL) and associated degradation by forest fires may have enduring effects on forest dynamics and composition (Asner *et al*., 2006), thus interfering with forest health and the availability of essential ecosystem functions and services (Gibson *et al*., 2011). A study in the Brazilian Amazon showed an increase in total forested areas affected by SL and forest fires from approximately 11,800 to 35,600 km^2 in 1992 and 1999, respectively (Matricardi *et al*., 2013). Asner *et al*. (2005) indicated that each year, SL can expand over as much forested area as does deforestation, with logged areas ranging in size from 12,075 to 19,823 km^2 between 1999 and 2002. The study showed that between 1999 and 2004, about 76% of all timber harvest practices led to high levels of canopy damage sufficient to leave forests prone to drought and fire (Asner *et al*., 2005). Currently, few truly undisturbed tropical forests exist, and arguably the single most important cause of tropical forest degradation worldwide is unsustainable SL (Miettinen *et al*., 2014). More than 400 million hectares of natural tropical forests have been degraded since 1980 (Edwards *et al*., 2014). Selective logging (SL) is a form of extraction of timber where a group of trees from selected species, the high-value tree species, are removed from the forest (Andersen *et al*., 2014). Global demand for precious and rare tropical timber, such as ebony and rosewood, is expected to continue to grow, and the international market has *c*. 50–90% of tropical wood harvested illegally (Nellemann, 2012). Thus, there is a need for sustainable forest management (SFM). Sustainable forest management (SFM) is the process of managing a forest to reduce forest degradation and deforestation by ensuring the sustainability of forest resources, protection, and conservation of genetic diversity and to ensure the sustainable exploitation of the biological resources, and enhancing the full valuation of forest goods and services (Poudyal *et al*., 2018). However, SFM is suffering from a variety of obstacles related to (Chia *et al*., 2020):

- governance issues (e.g. poorly defined tenure and resource rights, inadequate transparency and accountability, corruption, and limited involvement of relevant actors in the formulation of management plans);
- economic issues (e.g. high opportunity cost of maintaining forests, high transaction cost for better forest management, low financial returns from improved forest management, and unattractive incentives);
- regulatory and legislative issues (e.g. poor regulatory framework, lack of political will and incentive to implement regulations, and unrealistic legislation); and
- knowledge and capacity issues (e.g. poor understanding on the benefits to improve forest management, inadequate financial and material resources, and limited human resources to enforce and monitor regulations).

However, researchers and practitioners should not give up the SFM idea, but rather the effort to enhance it must be redoubled and refined.

Selective logging (SL) practices determine the outcome of SFM (Poudyal *et al*., 2018), thus an important component of SFM is monitoring of the forest status. Reliable and operational systems for monitoring SL in tropical forests have to be utilized (Hirschmugl *et al*., 2017). Such a system should be able to provide an estimate of baseline forest conditions in a spatially explicit fashion, departures from which can be used to assess current and previous trends of forest degradation and deforestation (Verbesselt *et al*., 2010). Remote sensing (RS) can be used to assess and monitor SL over tropical forests (Anwar and Stein, 2012), in a spatially and temporally continuous manner (Banskota *et al*., 2014). Accurate and precise detection of selectively logged sites in a forest is crucial for analysis of the spatial distribution of forest disturbances and degradation (Anwar and Stein, 2012). The RS methods that have been developed to detect SL in tropical forests detect intensive timber harvest (> 20 m^3 ha^{-1}). Therefore, medium spatial resolution datasets like Landsat are normally considered too coarse to detect less intensive SL (Hethcoat *et al*., 2018). Pinage *et al*. (2016) showed that the intensity of canopy impacts may vary according to the SL activity. High-intensity logging causes high forest damage that is long-lasting, and detectable on satellite imagery and *vice versa*. Soil fraction images obtained from spectral mixture modelling of multispectral or hyper-spectral data serve as a suitable approach for the detection of SL (Souza *et al*., 2005; Matricardi *et al*., 2010). Forests with obvious selective logging have well-defined logging infrastructure and extensive canopy degradation. Forests where subtle SL is taking place normally show less canopy perforation or visible infrastructure. Therefore, RS techniques may not easily differentiate them from undisturbed forests (Asner *et al*., 2004). Methods required for monitoring SL at high temporal resolution are not available. Some of the existing methods for mapping SL mostly come with numerous false detections, and existing techniques for minimizing them either impair the temporal accuracy or increase the omission error for the forest disturbance (Hamunyela *et al*., 2016). The extent of forest degradation as a result of SL using currently available techniques is unknown (Hethcoat *et al*., 2018). New change detection approaches can improve the detection of SL to achieve accurate mapping and quantification of forest loss (Hamunyela *et al*., 2016).

Over the last decades, there has been a rapid growth in the number of studies that investigated the use of RS for SL (Hethcoat *et al*., 2018). Providing an overview of the RS data and techniques that have been used in SL to identify the challenges and opportunities is essential. Such an overview would be useful practically in forest management and scientifically by highlighting the priorities and remaining research gaps for further investigation. Several review studies have been analyzed to evaluate the application of RS in mapping deforestation and forest degradation in tropical forests. Solberg *et al*. (2008) analysed the state of the art of RS techniques, detailing the relevant sensors and algorithms, usable datasets, and information on the leading institutions for research and development on techniques that might lead to operational monitoring of tropical forests. Miettinen *et al*. (2014) critically discussed available approaches for large-area forest degradation monitoring with satellite RS data at high to medium spatial resolution, in Southeast Asia. Da Ponte *et al*. (2015) provided an overview of the RS-based studies of tropical forest dynamics in Latin America, categorizing the existing studies based on selected sensors and data analysis methodologies. The review has considered both large-scale as well as small-scale forest changes solely induced by anthropological activities. Mitchard (2016) provides a review of earth observation (EO) methods for detecting and measuring forest change in the tropics. The study describes current and emerging EO technologies, and how these can be used to map forests and forest changes. Hirschmugl *et al*. (2017) reviewed the current state of the art in RS-based monitoring of forest disturbances and forest degradation. This literature reviews focus on Europe’s temperate forests and Africa’s tropical evergreen forests, using optical EO data. Mitchell *et al*. (2017) reviewed the RS approaches for monitoring forest degradation in support of countries' measurement, reporting, and verification (MRV) systems for reducing emissions from deforestation and forest degradation, conservation of existing forest carbon stocks, sustainable forest management and enhancement of forest carbon stocks (REDD+). The paper reviewed forest degradation that leads to canopy gaps that are detectable using RS, taking into consideration both forests within and outside the tropics. Therefore, there was need for an updated overview of the methods used in detecting and monitoring SL in tropical forests only. This chapter had the main focus laid on types of sensors and methodologies of data analysis, and the major challenges and further research needed to explore the use of RS in monitoring small-scale forest disturbance due to SL in tropical forests. The analyses concentrated on the period from 1992 to 2019.

2.2 Methods

2.2.1 Database search

The peer-reviewed articles published between 1992 and 2019 were searched in the web-based databases *ScienceDirect*®, *Web of Science*®, and *Scopus*®. They contain the largest abstract and citation databases of peer-reviewed literature, and the University of Wisconsin–Madison rates *ScienceDirect*®, *Web of Science*®, and *Scopus*® among the top ten web-based databases (UW Madison Libraries, 2019). The two most well-established databases that cover the widest scope of English-language literature, while focusing on peer-reviewed publications are *Scopus*® and *Web of Knowledge*® (Kleinschroth *et al*., 2016). Neither database is inclusive but complements the other.

The key terms used were (“Selective Logging” OR “Canopy Gap” OR “Canopy Opening” OR “Deforestation” OR “Forest Degradation” OR “Forest Disturbance”) AND “Tropical Forests” AND “Remote Sensing” in the title, abstracts and the keywords in the search. About 5546 articles were found in *ScienceDirect*, 2395 in *Web of Science*, and 888 articles in *Scopus* (as of June 20, 2019). Bibliographies of the articles were iteratively scanned until no new relevant articles were identified. All abstracts were scrutinized to filter out irrelevant articles. The remaining articles were read and retained only if they were relevant. Articles with irrelevant titles were excluded. After bringing the relevant articles together, the abstracts of 328 articles were read through. Finally, information was extracted from 110 relevant articles that explicitly discussed the RS of SL in tropical forests.

2.2.2 Content analysis

For the selected 110 articles, the same set of attributes for analysis was extracted (Table 2.1). In particular, noted were publication details, reference data availability, and the geographical location of the study area (i.e. if the study was conducted within the tropics). To characterize the sensor applied in each study, the name of the sensor, properties, and also the platform were recorded. The image processing techniques used to detect and quantify the areas under SL have been discussed. The auxiliary data used, the measure of accuracy, and the level of accuracy achieved were also investigated.

Table 2.1: *Information extracted from the literature on the remote sensing techniques applied in detecting and monitoring selective logging in tropical forests.*

Attribute	Description
Publication details	The year of publication, journal type, and national affiliation of the authors and co-authors.
Reference data	How reference data used for model calibration and validation were collected.
Geographical location	Location of the study area. If no location was given in the paper, Google Maps was used to determine the approximate location, or Landsat WRS2 reference was used to identify the centre location of the Landsat footprint used in the analysis.
Sensor	The sensor type used in the study.
Platform	The platform of the sensor, i.e., space-borne, air-borne, or unmanned aerial vehicle.
Spatial properties	The sensor resolution used: coarse resolution (>100 m), medium resolution (10-100 m), high resolution (5-10 m), and very high resolution (<5 m); the spatial scale of the studies.
Temporal properties	The temporal resolution of the sensor, and temporal scale of the remote sensing data (single date, bi-temporal, multi-temporal, or time series analysis), and sensor archive.
Spectral properties	The spectral properties used in the analysis: Vegetation Index (if a single vegetation index or band was used, then the index name was noted), Multi (if multi-spectral bands or multiple spectral indices were used), Hyper (if hyperspectral bands or multiple narrow-band indices were used), Lidar, and SAR.
Techniques	The techniques employing satellite data used to detect and quantify selective logging.
Accuracy measure	The measure of accuracy reported in the study.
Accuracy	Level of accuracy, according to accuracy measure.

2.3 Results of literature review

2.3.1 Publication details

Watrin and Rocha (1992) provided some of the first RS estimates of the area affected by SL (Souza *et al*., 2005). After the publication of the second article by Stone and Lefebvre (1998), the application of RS in SL analyses of tropical forests rose steadily (Figure 2.1). However, no relevant articles were published in 1993-1997, 1999, 2001, and 2011. The first major publication activity took place between 2002 and 2010, with an average of four papers per year, while the second considerable publication activity occurred between 2012 and 2019, with an average of nine articles per year. This shows that publication activity has more than doubled between 2012 and 2019. This could be related to the availability of information, awareness about remote sensing technology, and growing attention around the monitoring of SL and related activities in tropical forests.

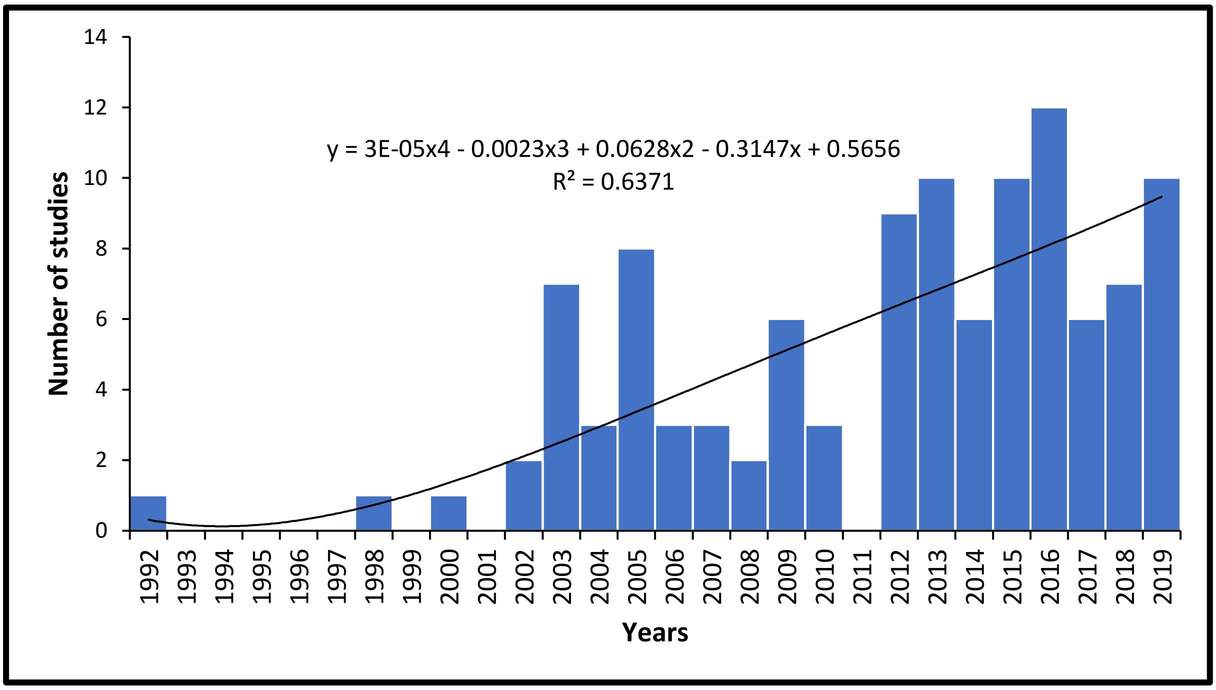


Figure 2.1. *Temporal distribution of published articles where remote sensing analysed selective logging in tropical forests.*

The results from the content analysis showed that using RS to examine SL in tropical forests is being accepted by a rising number of scientific disciplines; the articles are published in about fifty different scientific journals. About half of publications are appearing in journals specifically focusing on RS applications (Souza and Barreto, 2000; Souza *et al*., 2003, 2005; Asner *et al*., 2005; Koltunov *et al*., 2009; Matricardi *et al*., 2010, 2013; Heiskanen *et al*., 2015; Dalagnol *et al*., 2019), while the other half is distributed among other journal categories (Asner *et al*., 2004; Furusawa *et al*., 2004; Burivalova *et al*., 2015; Ellis *et al*., 2016; Sofan *et al*., 2016). The distribution between the journal categories has articles published in information and communication technology, ecological, cartography, interdisciplinary, and natural hazards-oriented journals.

The majority of the articles' first authors have been affiliated with institutions located in Europe (De Wasseige and Defourny, 2004; Rahm *et al*., 2013; Wang *et al*., 2019), North

America (Stone and Lefebvre, 1998; Wang *et al*., 2005; Ellis *et al*., 2016), South America (Souza *et al*., 2003; Graça *et al*., 2015; Conde *et al*., 2019), and Asia (Furuisawa *et al*., 2004; Sofan *et al*., 2016; Qu *et al*., 2018). Institutions in Africa do not have any representation (Figure 2.2). Only 0.7% of the articles have African co-authorship (Burivalova *et al*., 2015; Descals *et al*., 2017; Kankeu *et al*., 2016).

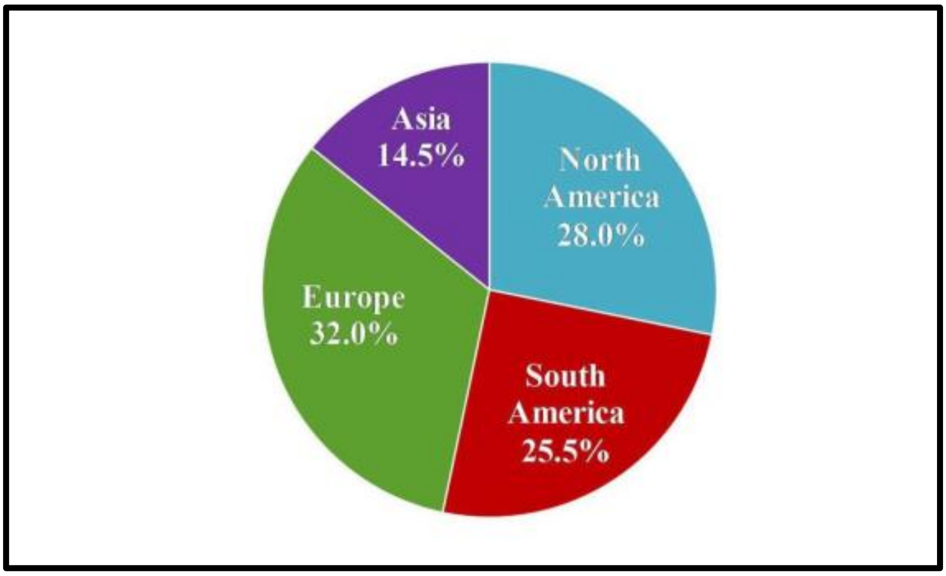


Figure 2.2. *Affiliation of lead authors in published articles where remote sensing analysed selective logging in tropical forests.*

Overall, 510 researchers from universities and research institutions around the world appear in the 110 articles (Table 2.2). The United States of America has 146 researchers who are from American universities and research institutions. The most featured institutions are Michigan State University (Matricardi *et al*., 2005, 2007, 2010), Carnegie Institution of Washington, Stanford University (Asner *et al*., 2004; Broadbent *et al*., 2008), Woods Hole Research Center (Laporte and Lin, 2003; De Grandi et al., 2015) and United States Department of Agriculture (USDA) Forest Service (Andersen *et al*., 2014). Brazil has contributed 111 researchers. Universities or research institutions hosting researchers at the time a study was being conducted and/or published, determined which countries the researchers were listed under. Eraldo A.T. Matricardi, for example, is listed under the USA because he published an article (Matricardi *et al*., 2007) on selective logging while at the Department of Geography, Michigan State University. In Matricardi *et al*. (2013), the same author is listed under Brazil because at the time of the publication of the article he was at the Department of Forestry, University of Brasilia.

Table 2.2: *Distribution of researchers in remote sensing of selective logging in tropical forests.*

Country	Lead author	Co-author	Country	Lead author	Co-author	Country	Lead author	Co-author
USA	26	120	France	2	16	PNG	0	2
Brazil	26	85	Austria	2	7	Norway	0	1
Italy	7	35	Netherlands	2	7	Laos	0	1
UK	6	30	F. Guiana	2	5	Cameroon	0	1
Germany	6	18	Mexico	1	5	Laos	0	1
Finland	6	9	Switzerland	1	2	Brunei	0	1
Japan	5	14	China	1	0	Bolivia	0	1
Australia	4	9	Malaysia	1	1	Argentina	0	1
Indonesia	4	9	Puerto Rico	0	3	Madagascar	0	1
Canada	3	6	Singapore	0	3	S. Africa	0	1
Belgium	3	1	Peru	0	2	Nepal	0	1

2.3.2 Geographical information

This study established the location of the study areas within the tropical region, where latitude and longitude coordinates were indicated. In the cases where there were no coordinates provided in the paper, Google Maps™ was used to find the approximate location, or Landsat World Reference System 2 (WRS2) was used to spot the centre location of the Landsat scene used in the analysis. As indicated in Figure 2.3, the scientific activities' geographic distribution is uneven in the tropical forests. Significantly more research has been done in South America, and about eighty-nine percent occurred in the Brazilian Amazon, specifically in Para, Rondonia and Mato Grosso states, where most of the disturbance has taken place. In Asia, the majority of the research was conducted in Kalimantan (Borneo). In Africa such studies were conducted in Cameroon. Cameroon has the highest percentage of previously logged forests than its Congo Basin neighbours since it has a higher population density (De Grandi *et al.*, 2015).

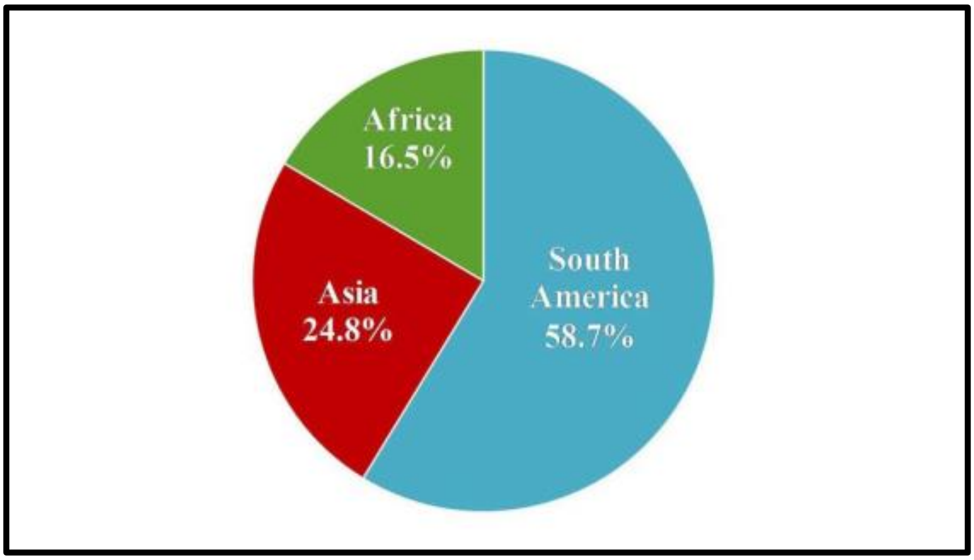


Figure 2.3. *Spatial distribution of studies where remote sensing analyzed selective logging in tropical forests.*

2.3.3 Sensors used to assess selective logging in tropical forests

Overall, twenty-six different sensors were used in the studies reviewed. Optical, RADAR (radio detection and ranging), and LiDAR (light detection and ranging) are identified as the three types of EO data, each with different characteristics. A significant number of articles utilized optical sensors (Figure 2.4a), such as the Landsat sensor (Costa *et al*., 2019), while others used the RapidEye (Franke *et al*., 2012; Deutscher *et al*., 2013), IKONOS (Read *et al*., 2003; Furusawa *et al*., 2004), *Satellite Pour l'Observation de la Terre* (SPOT)-4 (Guitet *et al*., 2012; Sofan *et al*., 2016), and SPOT-5 (Pithon *et al*., 2013). Very high resolution (VHR) and high resolution (HR) data were usually utilized as a single dataset or in conjunction with Landsat imagery as reference data. GeoEye-1 (Dalagnol *et al*., 2019), QuickBird (Hirschmugl *et al*., 2014), Pleiades (Langner *et al*., 2018), and the Chinese-Brazil Earth Resources Satellite 2B (CBERS-2B) High-Resolution Camera (HRC)—panchromatic, 2.5 m resolution)—(Anwar and Stein, 2012) were used as reference data. Wang *et al*. (2005) used IKONOS 1-m pansharpened imagery to validate canopy fractional cover maps resulting from Landsat Enhanced Thematic Mapper Plus (ETM+) data. In the published articles, the most recent Sentinel-2 (Lima *et al*. 2019), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)—(Broadbent *et al*., 2006), the MODerate Resolution Imaging Spectroradiometer (MODIS)—(Neba *et al*., 2014), and SmallSat (Descals *et al*., 2017) were also used to detect SL in tropical forests. Lima *et al*. (2019) used Sentinel-2 and Landsat-8

Operational Land Imager (OLI) images for monitoring SL in the Brazilian Amazon. Landsat-8 detected 36.9% more area of SL than Sentinel-2 data. Logging infrastructure was better mapped from Sentinel-2 (43.2%) than Landsat-8 (35.5%) data. Neba *et al*. (2014) reported that SL reduces aboveground biomass (AGB) stock, and through linear regression modelling discovered that logging roads and normalized difference vegetation index (NDVI) values derived from MODIS 250 m can indirectly determine above-ground biomass (AGB) logged. Unmanned digital aerial photographs (DAPs – Ota *et al*., 2019), and LiDAR data (Englhart *et al*., 2013; Kent *et al*., 2015; Melendy *et al*., 2018). Qu *et al*. (2018) estimated leaf area index (LAI) from LiDAR height percentile metrics and compared it with MODIS product in a selectively logged tropical forest area in Eastern Amazonia. Wedeux and Coomes (2015) employed airborne laser scanning (ALS) data to measure the canopy of old-growth and selectively logged peat swamp forest across a peat dome in Central Kalimantan, Indonesia.

The tropical areas are usually characterized by high cloud cover, thus RADAR sensors that can infiltrate clouds have been favoured in many cases (Figure 2.4b). Rauste *et al*. (2013) developed a technique to map SL in the northern Republic of Congo using ALOS PALSAR imagery acquired before and after the logging activities and attained an overall accuracy of 70.4%. The same technique was applied for TerrSAR-X data and achieved an overall accuracy of 53.6%. Lei *et al*. (2018) developed a new approach using TanDEM-X data to detect and quantify SL events in Tapajos National Forest, south of Santarem, Pará in the Brazilian Amazon region. A comparison of TanDEM-X results with ALOS-2 data qualitatively matches, confirming both the location and the epoch of the disturbance event. In the assessment of SL in tropical forests, a majority of the studies used medium spatial resolution sensors (52.2%), followed by VHR sensors (28%), HR sensors (18%), and coarse resolution (1.8%).

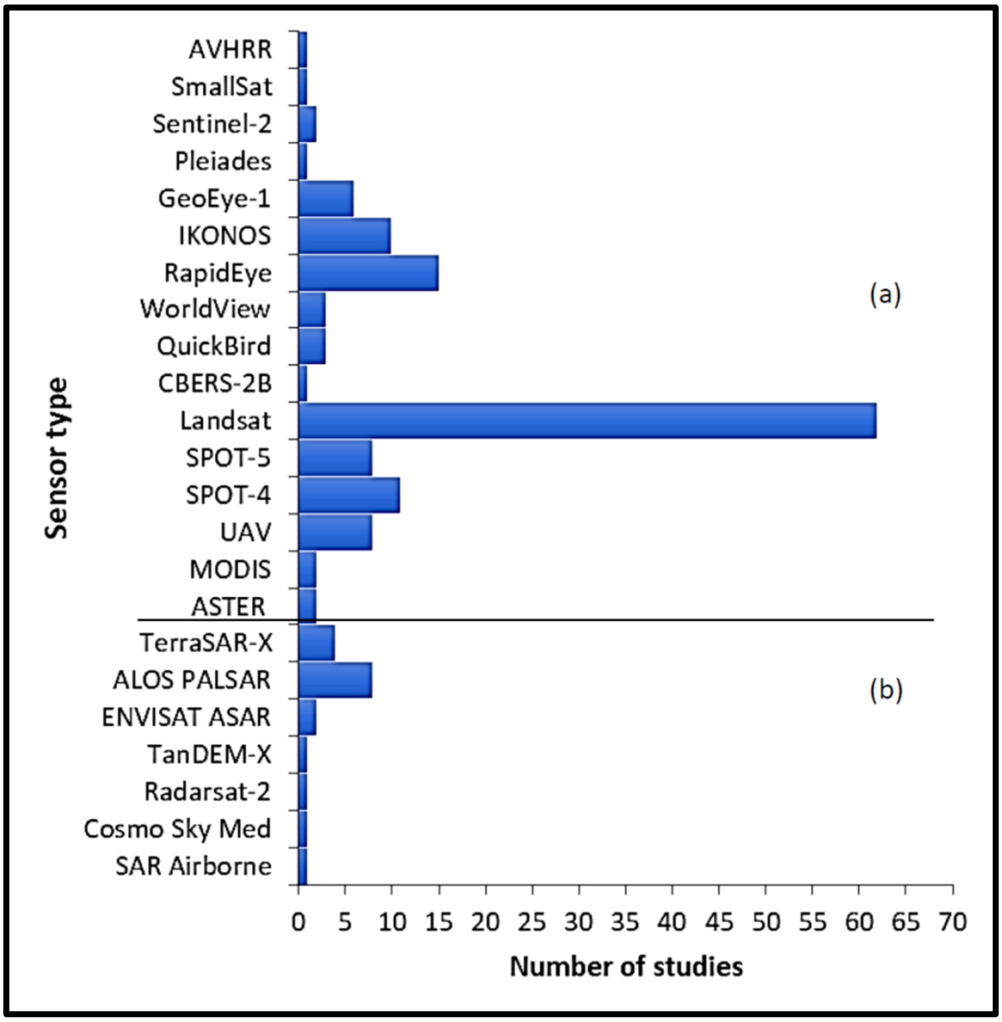


Figure 2.4. *Sensors used to assess selective logging (SL) in tropical forests. (a) Frequency of optical sensors; (b) frequency of RADAR sensors.*

Regarding the sensors and area covered by each study, Monteiro *et al*. (2003) used Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) to detect area affected by logging in three study areas covering 900 km^2, located in Sinop, Cláudia and Marcelândia in the state of Mato Grosso, Brazil. Field data were used to test the accuracy of the spectral mixture models to estimate the total area affected by logging (recent, old, and repeated). Marcelândia reported the greatest accuracy (80%), followed by Cláudia (73%) and Sinop (69%). The lowest area affected by logging was reported in Sinop (10,731 ha), followed by Marcelândia (19,391 ha) and Cláudia (25,276 ha). Souza *et al*. (2003) mapped forest canopy damage associated with logging and burning in a study site approximately 1600 km^2 in Paragominas, northeast of Pará, and achieved an overall accuracy of 86% of the forest degradation map (non-forest, degraded forest, and logged forest). As well, a high correlation (R^2 = 0.97) was observed between the total live AGB of degraded forest classes and the non-photosynthetic vegetation (NPV) fraction image. The NPV fraction improved the ability to map old selectively logged forests. Matricardi *et al*. (2007) applied visual interpretation and a textural algorithm on Landsat TM and ETM+ to identify and map selectively logged forests in a study site, approximately 1,300,550 km^2 in Brazil's Legal Amazon. The results indicated that by 1992 about 5980 km^2 of the forest had been logged and during the 1992-1996 and 1996-1999 periods the logged area expanded by 10,064 and 26,085 km^2, respectively. Visual interpretation and semi-automated techniques showed almost the same overall accuracy (OA) of 92.8% and 90.2%, respectively. Visual interpretation and semi-automated techniques combined produced an OA of 92.9%. In a study site covering about 4700 km^2 in a remote area of tropical peat swamp forest in Central Kalimantan, Indonesia, Franke *et al*. (2012) detected logging activities and the impact of fire by a pixel-based spectral mixture analysis (SMA). Forest, non-forest, and logging trails could be differentiated with an OA of 91.5%. Baldauf (2013) using TerraSAR-X detected the locations of extracted trees in a study site covering about 300 km^2 in Caracaraí municipality, federal-state Roraima, Northern Brazil, with an OA of 98.25%. Sofan *et al*. (2016) applied a spectral index derived from Landsat ETM+ and SPOT-4, the Normalized Difference Fraction Index (NDFI), then compared it with the Normalized Burn Ratio (NBR) and NDVI, to have enhanced detection of forest canopy damage caused by SL activities and associated forest fires in West Kalimantan. The study site covered about 100,000 km^2. The authors found that NDFI has higher accuracy (95%) to classify the degradation of forests due to logging and burning activities.

Other studies used the coefficient of determination (R^2) as an overall measure of accuracy. Asner *et al*. (2004) in a study site covering about 450 km^2 in Fazenda Cauaxi and Paragominas Municipality of Pará State, Brazil used Landsat ETM+ to assess the landscape and regional dynamics of canopy damage following SL (forest canopies, exposed non-photosynthetic vegetation, and soils). Spectral mixture analyses (SMA) of the Landsat ETM+ data showed that the spectral mixture model applied in this study, together with a combination of field- and image-derived spectral end-member bundles ($0.74 \leq R^2 \geq 0.96$), provided several information about forest canopy damage due to SL. Souza *et al*. (2009) in a study covering 62 transects (each 0.5 ha) within five regions of the Brazilian Amazon, used Landsat ETM+ and SPOT-4 to detect and map the extent of forest degradation based on canopy damage and small clearings. A high correlation ($R^2 = 0.7134$) showed the relationship between AGB and NDFI values for the degraded forest of Paragominas (Pará) and Sinop (Mato Grosso). Matricardi *et al*. (2013) used Landsat TM and ETM+ to survey forest degradation due to selective logging activities and forest fires in Acre, Amapá, Rondônia, and Roraima states, plus parts of Mato

Grosso, Maranhao, and Tocantins. The entire study site covered approximately 1,574,350 km^2. An empirical linear relationship between fractional coverage values derived from each VI and those calculated from hemispherical photographs was applied in the study. The modified soil-adjusted vegetation index (MSAVI) presented the highest value of the coefficient of determination (R^2) for Rondônia and Acre study sites (0.74 and 0.81, respectively). MSAVI and MSAVIaf showed the best overall performance under clear sky conditions for the Mato Grosso, Rondônia, and Acre study sites. However, MSAVIaf showed the best performance under smoky conditions for the Mato Grosso study site (Matricardi *et al*., 2010). Based on these results, the authors also selected MSAVIaf as the optimum vegetation index to be used in this basin-wide analysis.

2.3.4 Spatial scale

Throughout the articles reviewed, only local and regional spatial scales have been utilized to examine SL in tropical forests. Local-scale defines a study covering a relatively small area in a region or country. Regional studies differ in scale, considering not only broader areas such as part of or the entire Amazon in Brazil, but also a study covering an entire country. The local scale studies were more frequent and constituted eighty-five percent of the articles published (Wang *et al*., 2005; D'Oliveira *et al*., 2012; Antropov *et al*., 2015; Grecchi *et al*., 2017; Ota *et al*., 2019), while only fifteen percent of the studies monitored SL at a regional scale (Asner *et al*., 2005; Broadbent *et al*., 2008; Matricardi *et al*., 2013; Langner *et al*., 2018).

2.3.5 Temporal scale

As far as the temporal scale of the studies is concerned, it can be concluded that detecting SL in tropical forests is achieved using a single date (composite image), bi-temporal, multi-temporal, and time-series data analysis. The current review showed that about ten percent of the studies utilized single-date analysis. Read *et al*. (2003) merged 1-m and 4-m resolution single date IKONOS data to evaluate and monitor logging impacts, north of the Rio Amazon near Itacoatiara, Amazonas. Wang *et al*. (2005) acquired one scene of Landsat ETM+ on June 18, 2000, to measure forest degradation caused by SL in the Amazonian state of Mato Grosso, Brazil. A linear mixture model was applied in the optimal vegetation index domain, the MSAVI to derive tropical forest fractional cover, and to study the capability to detect SL in an area that consisted of clear-cut areas, and undisturbed forests, and degraded forests (Wang et al., 2005). The canopy fractional cover maps derived from ETM+ imagery were validated with 1-m pan-sharpened IKONOS imagery. Tangki and Chappell (2008) used a Landsat-5 TM acquired in

March 1997 to quantify mean tree biomass across a selectively logged forest in the Ulu Segama Forest Reserve, Borneo. The results indicated a major difference in the stand biomass from 172 t ha^{-1} in a forest area that had high-lead logging to 506 t ha^{-1} in a similarly sized area of undisturbed forest. Anwar and Stein (2012) used a Landsat-5 TM image acquired on 21 July 2008 to detect changes in forest canopy due to SL in the south-western Brazilian Amazon.

Thirty-one percent of the studies used bi-temporal data analysis. Asner *et al*. (2002) used two Landsat-7 ETM+ scenes acquired on July 13, 1999, and July 31, 2000, to quantify the extent and intensity of SL and the canopy closure, years after logging operations in Fazenda Cauaxi, Pará State. Andersen *et al*. (2014) applied a simple differencing of 2010 and 2011 LiDAR canopy height models to identify areas where canopy over 30 m tall had been removed in natural tropical forests in the Western Brazilian Amazon. Pinage *et al*. (2019) used airborne LiDAR in 2012 and 2014 to determine the areas with canopy gaps caused by reduced-impact selective logging (RIL) in a forest in the Eastern Brazilian Amazon. Bi-temporal analysis detects SL in forests by comparing two RS datasets obtained at different points in time (Da Ponte *et al*., 2015).

About thirty-four percent of studies applied multi-temporal analysis. Qi *et al*. (2002) made use of Landsat ETM+ images which were acquired in 1992, 1996, and 2000 to map and quantify SL in the Amazon State of Mato Grosso, Brazil. The results were validated using an IKONOS image acquired on June 13, 2000. Souza *et al*. (2003) used one multispectral SPOT-4 scene acquired in August of 1999, and four Landsat TM images acquired in 1988, 1991, 1995, and 1996, to assist in the identification of forest degradation age in mapping forest canopy damage associated with logging and burning in the Paragominas study. The resultant forest degradation map, with an OA of 86%, showed 35% of the forest area as intact forest, 56% as logged forest, and 9% of the forest area was classified as degraded forest. To assess old selectively logged forests and old burned forests at "Fazenda Agrosete" in the Paragominas logging centre, north-east of Pará state, Souza and Roberts (2005) applied four Landsat TM images acquired in 1984, 1988, 1991, and 1996 and one SPOT-4 image acquired in 1999. Matricardi *et al*. (2013) estimate showed that about 5467, 7618, and 17,437 km^2 were new areas of SL and/or forest fires in 1992, 1996, and 1999, respectively, and 2.4% and 1.3% of the total detected selectively logged and burned forests, respectively, took place within protected areas. Sofan *et al*. (2016) applied spectral indices analysis on Landsat-7 ETM+ data acquired on July 23, 2006, May 7, 2007, August 5, 2008, July 31, 2009, and SPOT-4 data dated July 31, 2009, May 16, 2012, and October 15, 2012, to enhance detection of forest canopy gaps due to SL activities and associated forest fires in West Kalimantan, Indonesia. Unlike bi-temporal analysis, multi-temporal assessments constitute longer periods which allow quantification of selectively logged areas, as well as offering information concerning the status of the forest (Da Ponte *et al*., 2015). The remaining twenty-five percent comprise the time series analysis. De Wasseige and Defourny (2004) developed an operational system to detect and monitor SL activities in a study that took place in the Ngoto forest, the south-western part of the Central African Republic, using time series of six SPOT multi-spectral (from December 95 to July 96) and two Landsat TM (November 1990 and January 1995) images. The study concluded that despite rapid canopy closure, more than 50% of the logging roads and skid trail surface network was still visible five years after the last timber extraction. Asner *et al*. (2005) applied an automated RS technique to map SL in the Brazilian Amazon, using Landsat ETM+ imagery from 1999 to 2002. Matricardi *et al*. (2005) estimated the area affected by logging in the Sinop region, State of Mato Grosso, Brazil by applying an annual time-series analysis based on 11-year series of Landsat imagery from 1992 to 2002. The study showed that, due to rapid regrowth and deforestation, evidence of logging activities normally vanished in a period of 1 to 3 years. Koltunov *et al*. (2009) present a large-scale study of the relationships between SL and forest phenology using time-series analysis of MODIS satellite data of selectively logged forests in Mato Grosso, Brazil. The time-series analysis shows that SL leading to about 5-10% canopy damage contributed a remarkable 54% of all year 2000 logging observations. Generally, a low level of logging intensity causes low forest damage and, therefore, contributes to short-lived forest phenology changes, while high-intensity logging causes high forest damage that is long-lasting and detectable on satellite imagery (Pinage *et al*., 2016). Souza *et al*. (2013) using annual Landsat imagery from 2000 to 2010, quantified degradation and deforestation rates in the Brazilian Amazon over 10 years. The results showed a major decline in annual deforestation rates by 46% at the end of 2005 and a 20% increment in annual forest degradation. Based on radar sensors, Antropov et al. (2015) evaluated the use of bi-temporal mosaics of ALOS PALSAR data collected in 2007 and 2010 to detect and monitor SL activities in the northern part of the Republic of the Congo. In a second study, a time series of strip-map C-band SAR data was assessed. The technique involved the analysis of textural features of SAR backscatter temporal log-ratio image. This is the first demonstration of C-band SAR-based mapping of selectively logged areas. Verhegghen *et al*. (2015) utilized a time series analysis of SPOT-4 and RapidEye images covering the North of the Republic of Congo, acquired in the period from February to June 2013 to assess forest degradation from SL. The analysis highlights very rapid changes in the forest cover, as most logging gaps varnished a span of a few months. Different from bi-temporal analysis, time series can provide a quasicontinuous history of SL and regeneration processes (Da Ponte *et al*., 2015). The detection of SL signals using RS is difficult and they are rapidly lost within a very short time due to vegetation regrowth, and this is why the time factor is very crucial in detecting SL in tropical forests (Hirschmugl *et al*., 2014).

2.3.6 Methods employed to map and characterize selective logging in tropical forests

A variety of mapping techniques have been tried and utilized in studies of SL in tropical forests (Souza *et al*., 2013). These techniques differ in their mapping goals, approach, geographic extent, and map accuracies. Visual interpretation of satellite imagery has been widely used in mapping SL in tropical forests (Read et al., 2003; Furusawa *et al*., 2004). Furusawa *et al*. (2004) evaluated the effects of a SL operation in New Georgia Island, Solomon Islands employing supervised classification and onscreen visual interpretation, in association with detailed field observation. The selectively logged area accounted for 7.4% of the original forest and thus causing substantial damage. Read *et al*. (2003) concluded that the use of a combination of high-resolution satellite images, taken immediately after logging, and detailed field observations may serve as the most useful way of analysing selection logging. However, visual interpretation is considered to entail more workload and the skills of the interpreter determine the precision of the final product (Da Ponte *et al*., 2015). A technique often applied in detecting SL is the Spectral Mixture Analysis (SMA). To map SL in the state of Pará, Souza *et al*. (2003) applied spectral mixture models to calculate fraction images from SPOT-4 imagery and concluded that more field studies are needed for SL assessment in other regions with different types of forests. Soil fraction images acquired through spectral mixture models assist to identify small forest perforations with spatial patterns showing at the sub-pixel scale in RS imagery (Souza *et al*., 2003). Selective logging (SL) increases fire susceptibility, owing to a large amount of available fuel in the form of slash piles and collateral damage caused by the logging operations (Souza and Barreto, 2000). An index based on SMA known as the Normalized Difference Fraction Index (NDFI) was proposed by Souza *et al*. (2005) to enhance the detection of forest canopy damage due to SL activities and associated forest fires. Analysis of the NDFI data is enabled by a contextual classification algorithm (CCA) that facilitates accurate detection of logging and fire-derived canopy gaps (Souza *et al*., 2005). Others have used SMA together with geographic information system (GIS) techniques (Souza and Barreto, 2000; Monteiro *et al*., 2003). Asner *et al*. (2006) created a measure, Area-Integrated Gap Fraction (AIGF) of canopy damage by converting remotely sensed subpixel fractional changes in canopy cover into spatial estimates of forest canopy gap fraction. Asner et al. (2005) advanced the computational analysis of Landsat ETM+ satellite data using the Carnegie Landsat Analysis System (CLAS) to map degraded forests by SL in the Brazilian Amazon with 86% OA. This approach provides automated image analysis using atmospheric modelling (entails detection of forest canopy openings, surface debris, and bare soil exposed by forest disturbances) and pattern-recognition techniques. A wide range of spectral variables (Table 2.1) was used in enhancing the detection of SL because they are sensitive to vegetation greenness (Souza *et al*., 2005). Broadbent *et al*. (2006) used a probabilistic spectral mixture model, i.e. automated Monte-Carlo unmixing (AutoMCU) analysis to generate per-pixel fractional cover estimates of photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and soil. Results were compared with the NDVI—NDVI, and the PV and NPV fractions in felling gaps >400 m^2 were distinguishable from unlogged forest values for up to six months after logging, and those <400 m^2 were distinguishable for up to three months after harvest. Matricardi *et al*. (2013) applied visual interpretation and semi-automated techniques to detect and quantify basin-wide forest canopy damage due to SL in the Brazilian Amazon. The semi-automated technique involved an application of a textural algorithm to detect patios. A buffer was applied around the patios to estimate the amount of forest affected by logging. The authors also used a semi-automated RS technique based on NPV fraction images derived from SMA, to detect burned forests in the Amazon state of Mato Grosso. The results based on Landsat imagery showed a significant rise (~300%) in forest degradation by SL and burning in the Brazilian Amazon between 1992 and 1999. Souza et al. (2013) used SMA, NDFI, and knowledge-based decision tree classification (DTC) to quantify deforestation and degradation (logged and burned forests) in the Brazilian Amazon. High spatial resolution SPOT 2.5 m pixel imagery was used to generate geolocated points to assess the accuracy of the classified images. The study mapped and assessed the accuracy to quantify forest (97%), deforestation (85%), and forest degradation (82%) with an OA of 92%. The authors concluded that the total forest area degraded by SL and fires each year was equal to or exceeded the area deforested, leading to more fragmented landscapes. Sofan *et al*. (2016) detected forest canopy damage due to SL and forest fires in Kapuas Hulu and Sintang districts of West Kalimantan, Indonesia by applying the SMA approach and spectral indices analysis. Landsat-based monitoring is inadequate for monitoring SL due to its low spatial resolution (Da Ponte *et al*., 2015), and in solving this problem other studies have used object-based change detection techniques (Burivalova *et al*., 2015; Grecchi *et al*., 2017), trajectory-based analysis (Shimizu *et al*., 2017), and the use of an airborne laser scanner (ALS) which determines the three-dimensional (3-D) structure of forest vegetation, and are thus used to estimate AGB or forest structure changes (Englhart *et al*., 2013; Kent *et al*., 2015; Pinage *et al*., 2019).

2.3.7 Accuracy assessment

The selected articles varied enormously in the level of details about the map validation process, as well as on the metrics used to report accuracy. The study area size ranged from about one km^2 to a regional scale, covering about 4,000,000 km^2, but ~34% of cases (n = 37) dealt with areas covering less than or equal to 1000 km^2. Only 80% of the papers reviewed reported the measure of accuracy. All four standard accuracy assessment metrics (overall accuracy- OA, kappa- κ, user's accuracy- UA, and producer's accuracy- PA) were reported in 28.2% of cases. Another ~5% of the cases included only OA. The accuracies attained when detecting and monitoring SL in tropical forests were explored, and the lowest OA was recorded by Hernandez-Gómez *et al*. (2019), distributed as 13.9% in Petcacab and Caobas, 14.5% in Felipe Carrillo Puerto, and 19.1% in Noh Bec (Yucatan, Mexico). The study site was approximately 1600 km^2 and used Landsat ETM+/OLI. Over 68% of the observed disturbed areas were misclassified as unlogged pixels in all *ejidos*, but the CLAS algorithms worked well in detecting forest degradation from logging in other countries of the tropics, such as Brazil (Asner *et al*., 2005, 2006; Broadbent *et al*., 2008). Among the highest OA were 96.7% and 95.7% for Sentinel-2 and Landsat-8, respectively in a study by Lima *et al*. (2019) while mapping forest disturbance caused by SL in the south of the Brazilian State of Amazonas. The study site was approximately 5625 km^2. An assessment of the impacted area by SL led to a discovery of 1143 ha and 1197 ha of disturbed forest on Sentinel-2 and Landsat-8 data, respectively. Despite its coarser spatial resolution, Landsat-8 has a good potential to identify logging features. These results agree with a recent study carried out by Hethcoat *et al*. (2018) in the State of Rondônia, Brazil. However, the area mapped as logged with Landsat-8 data is larger in comparison with Sentinel-2-based results. Sentinel-2 leads to a more precise pixel-based mapping of SL due to its higher spatial resolution, making it possible to map smaller disturbances and to map larger disturbances more precisely.

In a study site covering about 30,000 km^2, Matricardi *et al*. (2010) in the analysis of SL and forest fire impacts on natural forests, attained an OA of 95.8% for burned forests detection based on non-photosynthetic vegetation fraction image. Selective logging (SL) was found to be responsible for disturbing the largest proportion (31%) of natural forests, followed by deforestation (29%). Altogether, SL and forest fires affected approximately 40% of the study site area. Rahm *et al*. (2013) recorded an OA of 91.5%, 89.7%, and 88.9% on RapidEye data at 5, 10, and 20 m, respectively, while detecting forest degradation in a study site (15,000 km^2) in the province of Bandundu in the Democratic Republic of Congo. From a forest extent of 34,233 ha on the RapidEye image of 2011, approximately 388 ha were deforested and 267 ha were degraded in August 2012. Costa *et al*. (2019) applied a semi-automatic technique based on texture algorithm and visual interpretation to map selectively logged forests, in portions of the states of Mato Grosso, Para, and Rondonia. Based on the results obtained, it was estimated that the overall accuracies are greater than 91%. The study discovered that forests affected by SL increased in Mato Grosso and Rondônia, while a decrease in forests affected by logging activities was observed in Pará.

Qi *et al*. (2002) used Landsat images to detect SL in the Amazon State of Mato Grosso, Brazil. Since extensive field measurements are limited due to labour intensity in the tropical areas, the authors used the IKONOS image to validate the fractional forest cover estimates. Fractional covers from the ETM+ image were plotted against those from IKONOS. The ETM+ and IKONOS fractional cover values fitted well, with R^2 = 0.96. D'Oliveira *et al*. (2012) mapped forest biomass in areas of low-intensity logging using airborne scanning LiDAR in Antimary State Forest, Acre State, Western Brazilian Amazon. A systematic random sample of fifty 0.25-ha ground plots was measured and used to construct LiDAR-based regression models for AGB (R^2 = 0.72). The total and mean AGB estimates obtained using the synthetic estimator (total 231,694 Mg; mean 231.7 Mg ha^{-1}) nearly matched those obtained using the model-assisted estimator (total 231,589 Mg ± 5.477 SE; mean 231.6 Mg ha^{-1} ± 5.5 SE; ± 2.4%). They were more precise than plot-only simple random sample estimator (total 230,872 Mg ± 10,477 SE; mean 230.9 Mg ha^{-1} ± 10.5 SE; ± 4.5%). Englhart *et al*. (2013) analyzed multi-temporal LiDAR and field inventory measurements to study forest dynamics and AGB changes in undisturbed, selective logged, and burned peat swamp forests in Borneo. Above-ground biomass (AGB) regression models were developed based on field inventory measurements and LiDAR-derived height histograms for 2007 (R^2 = 0.77) and 2011 (R^2 = 0.81) were quantified, as well as changes in canopy height and AGB. Intact forests had on average 20 t ha^{-1} AGB with a canopy height increase of 2.3 m over the four years. Selectively logged forests showed an average AGB loss of 55 t ha^{-1} within 30 m and 42 t ha^{-1} within 50 m of detected logging trails, although the mean canopy height increased by 0.5 m and 1.0 m, respectively. Burned forests lost 92% of the initial biomass. To quantify landscape-scale changes in canopy structure due to logging, Wedeux and Coomes (2015) using airborne laser scanning (ALS) data, applied gap size-frequency distributions (GSFDs) in a selectively logged peat, Mawas Conservation Area, Central Kalimantan, Indonesia. This technique was validated, as it yielded a fit going through the origin and with an $R^2 = 0:88$ between predicted and measured peat values in 33 plots where peat data were available.

The validation methods used in SL studies in tropical forests are field data, RS data, a combination of field data and RS data, and data from model simulations (Figure 2.5).

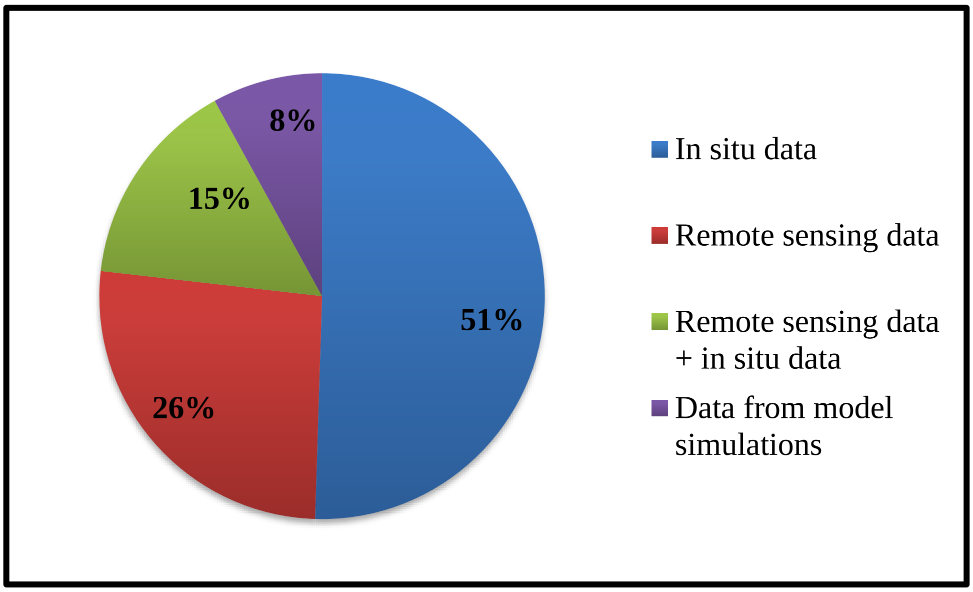


Figure 2.5. *Validation methods used in selective logging (SL) studies in tropical forests.*

2.4. Discussion

The following section discusses the main findings of this review as well as the emerging need for further studies on RS techniques applied in detecting and monitoring SL in tropical forests. The resultant needs of studies are addressed to offer appropriate solutions to guide tropical forest management practices and current research.

2.4.1 Application of remote sensing

The analysis shows that the number of journals publishing RS research has broadened with time in their scientific disciplines, meaning that the RS users have diversified. The publications have shifted from strictly RS-oriented journals into journals specialized in forest research, geography, biological conservation, geophysical research, and interdisciplinary science. This suggests that RS is most likely to be accepted as a dependable information source by a larger group of scientific researchers (Karlson and Ostwald, 2015). More integration of RS into scientific disciplines dealing with other types of research questions about forest disturbances is required to further promote the scientific involvement of this rapidly developing technology.

The need of having improved quality and accessibility of RS data applicable at different spatial scales will possibly facilitate this trend. Improved access to relevant satellite data and the best available methods are key to operational forest degradation monitoring (Da Ponte *et al*., 2015).

Despite steady progress in RS technologies, the spatial extent, temporal resolution, and availability of data acquired are inadequate to meet the growing demand for information. There is a need for more accessible, fast, and precise information on the world's tropical forest dynamics. Production of forest disturbance maps is only useful if the stakeholders can access and query the data (Mitchard, 2016). Improved availability and accessibility of high resolution (HR) and very high resolution (VHR) imagery and the development of algorithms able to handle complex data structures will potentially detect and monitor SL in tropical forests (Kuemmerle *et al*., 2013). Deforestation and degradation processes take place daily and on different scales, yet several countries do not have the economical means or governmental support to establish monitoring programs meant to monitor forest disturbance (Da Ponte *et al*., 2015). Examples of already implemented programs are the Brazilian governmental project "Basin Restoration Program" (PRODES), carried out through INPE, monitoring the forests over the Legal Amazon region (Solberg *et al*., 2008). Complementary to PRODES, since May 2005, is a near real-time monitoring system called *Detecção de Desmatamento em Tempo Real* (DETER), able to detect forest disturbances larger than 25 ha (Solberg *et al*., 2008) in a 15-day interval using MODIS and CBERS satellite data. In 2008, a new programme, *Mapeamento da Degradação Florestal na Amazônia Brasileira* (DEGRAD), was introduced to assess degradation especially from selectively logged activities by using Landsat images and CBERS (Da Ponte *et al*., 2015). Started in July 2015, MapBiomas is an initiative involving universities, non-governmental organizations (NGOs), and technology companies (including Google) working together to understand Brazilian territory transformations from 1985 to the present, based on the annual mapping of land cover land use (Wang *et al*., 2019). MapBiomass uses Google Earth Engine® to process and distribute satellite images quickly. The platform is open-access and provides data, codes, and methodologies to users, allowing researchers to use MapBiomass´ maps and other products (Wang *et al*., 2019). Such programs are needed elsewhere within the tropical forests. Ideally, all data, including forest inventory datasets would be made available as open data. This often maximizes the use of the product, leading to innovative users who could not have been envisaged in advance (Mitchard, 2016). Thus, open data policies will go a long way in supporting routine monitoring efforts.

2.4.2 Geographic distribution of the scientific activity

The scientific activities related to forest disturbance due to SL in the tropical forests are not evenly distributed. This trend is considerably set by research preferences and investments put in place by institutions within a country, further heightened by the way large international research projects are allocated, and the bilateral associations among national governments and universities (Da Ponte *et al*., 2015; Karlson and Ostwald, 2015). It can as well be inferred that the shortage of studies in some tropical regions is due to a lack of expertise in the area. Data gaps mainly exist in developing countries due to inconsistent data acquisition and distribution frameworks, yet it is where deforestation and forest degradation are extensive (Kuemmerle *et al*., 2013). The areas within the Congo Basin have been experiencing armed conflict, hence researchers are concerned about their security. The low number of authorship with African affiliations is a proof that the research carried out in the tropics has majorly been initiated, conducted, and funded by foreign institutions, usually located in Europe (Deutscher *et al*., 2013; Hirschmugl *et al*., 2014; Neba *et al*., 2014) and North America (Laporte and Lin, 2003). Research funding institutions in the United States of America have a very strong presence in the Brazilian Amazon (Souza and Barreto, 2000; Broadbent *et al*., 2006; Koltunov *et al*., 2009; Qu *et al*., 2018; Dalagnol *et al*., 2019). Brazilian research funding institutions have also funded several studies in the Brazilian Amazon (Souza and Barreto, 2000; Souza and Roberts, 2005; Conde *et al*., 2019; Costa *et al*., 2019; Pinage *et al*., 2019). The Brazilian Amazon, being the world's largest contiguous rain forest (Anwar and Stein, 2012), and endowed with high levels of biodiversity and its importance in terms of ecological resources (Da Ponte *et al*., 2015), has attracted more research. This is due to (Da Ponte *et al*., 2015):

- the fact that the Brazilian Amazon has remarkable rates of forest disturbances, especially in Rondônia and Mato Grosso states;
- the availability of expertise in RS and related technologies;
- the influence of government support and awareness;
- international interest from the scientific community, and
- availability of financial support to conduct research.

At present, there is a need to increase the study to other regions within tropical areas, such as tropical Africa, which are disadvantaged in RS technology and knowledge. This can be achieved through multilateral research collaborations, readily available *in situ* data and research infrastructure, and political stability. In areas experiencing social unrest, high-resolution imagery may complement *in situ* data, thus reducing the costs and risks associated with field campaigns. Improved support for capacity building of local research institutions and universities will initiate an increased usage of RS. Remote sensing (RS) capacity majorly depends on RS data availability. Currently, a variety of high-quality RS data, such as the data archive of the Landsat sensor or the high-spatial and temporal resolution of the Sentinel satellites, is freely available and in the future, it is expected there will be free data for even higher spatial and temporal resolutions.

2.4.3 Multi-temporal monitoring of selective logging

The majority of the articles reviewed monitored SL using bi-temporal and multi-temporal analysis. Relatively, fewer studies utilized the continuous spectral-temporal trajectories for the detection and analyses of SL in tropical forests. Mapping and monitoring SL requires data with adequate temporal and spatial resolution. As a result of insignificant seasonal changes in dense tropical forests, and the persistent cloud cover for most of the year, the potential benefits of time series analysis play a key role in the multi-temporal studies (Miettinen *et al*., 2014). The initial signs of logging activity visible from satellite data vanish in the humid tropics quickly as a result of the fast regrowth of secondary vegetation (Langner *et al*., 2018). Harvest intensity is more directly related to the number of felling gaps, although that relationship may be affected by overlying felling gaps (Broadbent *et al*., 2006). Canopy openness defines the ability of remote sensors to detect many of the ground disturbances indicative of logging activities (Broadbent *et al*., 2006). Selective logging (SL) at moderately high intensities poses large-scale impacts on the canopy and an abundance of spectrally distinct features like log landing decks or large road networks (Hethcoat *et al*., 2018). Selective logging (SL) at lower logging intensities, e.g. a single tree fall or an under-canopy skid trail, has small-scale impacts on the canopy and can be impossible to detect with moderate spatial resolution imagery where only a small proportion of an imaged pixel is affected by the target feature (Read *et al*., 2003). Also, some logging impacts such as the loss of green cover associated with a felled tree are ephemeral, and infrequent repeat observation opportunities over cloudy tropical areas make it challenging to detect SL (Melendy *et al*., 2018). Over the last two decades, large-region evaluation of SL has been demonstrated with moderate resolution optical imagery (Souza *et al*., 2005), but with uncertainty resulting from sensor limitations. This promotes the significance of regular data acquisition to monitor keenly the temporal changes of reflectance signatures (Miettinen *et al*., 2014). The fast regrowth vegetation that covers the scars in the canopy is also detectable in satellite data, prolonging the available time for degradation monitoring, but due to the atmospheric conditions within the tropical forests, it has been impossible to obtain continuous time series with high to medium resolution data until very recently (Miettinen *et al*., 2014).

Fused satellite observations can be utilized in creating a dense time series to track forest disturbances such as SL (Mitchell *et al*., 2017). This offers a great opportunity to explore the full potential of time series analysis for forest disturbance mapping. Therefore, monitoring logging activities might be enhanced by increasing intra-annual data availability for both medium- and high-spatial-resolution sensors (Da Ponte *et al*., 2015). Dense time series provides a high possibility for monitoring SL but requires considerable investment in data processing for it needs extensive computing resources. Methods are required for analysing dense time series of satellite image time series that provide high spatial details. The opportunity has several challenges and needs new methods that can efficiently handle dense satellite image time series that enable temporal analysis while accounting for a spatial context. This would enable the monitoring of SL in tropical forests with unprecedented detail.

2.4.4 Remote sensing techniques applied to selective logging in tropical forests

The need to research tropical forest degradation emerged in the 1990s as the spatial extent of SL was found to be not accounted for in deforestation studies (Hirschmugl *et al*., 2014). Various RS methods have been used to detect and estimate the intensity of SL in tropical forests, but these methods are less developed (Hethcoat *et al*., 2018). The majority of RS-based research, regarding the monitoring of SL in tropical forests used moderate resolution RS imagery, such as Landsat imagery, putting into consideration that it is cost-free and acquired at regular intervals (Souza et al., 2013; Mitchard, 2016). Most authors (Asner *et al*., 2002; Shimabukuro *et al*., 2014) acknowledged that they can detect selectively logged areas at moderately high intensities (>20 m^3 ha^{-1}; 3-7 trees ha^{-1}), but their methods' ability to quantify the extent of lower logging intensities (<20 m^3 ha^{-1}) and duration of logging damage using Landsat imagery is not known (Hethcoat *et al*., 2018). Moderate resolution RS imagery presents spectral confusion of canopy gaps due to natural disturbance (e.g. windfall) with canopy gaps caused by selective harvesting (Asner *et al*., 2002). The selective harvest gaps are often sub-pixel in scale (Souza *et al*., 2005). The broad spectral range covered by Landsat wavelengths can limit the detection of spectrally subtle changes (Asner *et al*., 2002). Fusion techniques of panchromatic and multispectral images are commonly applied to enhance the imagery (Read *et al*., 2003), as well as upscaling of the spatial resolution using vegetation indices (Arroyo-Mora *et al*., 2009).

Early studies assessed SL by visual interpretation on Landsat images (Laporte and Lin, 2003; Read *et al*., 2003; De Wasseige and Defourny, 2004; Furusawa *et al*., 2004). Visual interpretation has been used to quantify clearings in forests and also to identify areas degraded by SL activities (Souza et al., 2003; Matricardi *et al*., 2005). Da Ponte et al. (2015) state that the method is time-consuming and ineffective in quantitative, wall-to-wall assessments of forest disturbance dynamics. The skills of the interpreter determine the precision of the final product. The Spectral Mixture Analysis (SMA) method, although widely used in the Brazilian Amazon, a major challenge with the SMA method by Franke *et al*. (2012) in Central Kalimantan is that choosing the endmember would have to consider the differences in spectral reflectance as a result of topography and atmospheric conditions, and this limits the use of SMA analyses over large areas. The post-classification method has been used to generate independent classification results from two images acquired in different periods and then compared pixel-by-pixel (Hirschmugl *et al*., 2014), or object-by-object (Burivalova *et al*., 2015). This method can only be used in a bi-temporal data analysis, and therefore it needs multiple bi-temporal comparisons (Da Ponte *et al*., 2015). Object-based analyses focus on physically meaningful features rather than pixels. This is achieved through image segmentation into individual tree crown (ITC) polygon objects, which reduces the canopy's high local spectral variability, and minimizes the mismatch of tree crown geo-misallocation (Dalagnol *et al*., 2019). Win *et al*. (2012) used the image differencing method using two SPOT 5 pan-sharpened images acquired in October 2007 and January 2009 to analyse canopy changes from forest harvesting. Other complex methods were developed, such as the LandTrendr algorithm, usually applied to a single spectral index or variable, such as Tasseled Cap Wetness (TCW), which is derived from a Tasseled Cap Transformation, was used as the spectral index because of its sensitivity to disturbance events and water content to map SL among other forest disturbance activities in Bago mountains, Myanmar (Shimizu *et al*., 2017).

Remote sensing (RS) is reliable in mapping forest disturbances such as SL (Andersen *et al*., 2014). In detecting SL, different algorithms have their own merits—often, different algorithms are mostly compared to get the best. While localized case studies often provide detailed maps of forest disturbance that are useful for forest management, large-scale mapping techniques could offer vital scientific insights (Senf *et al*., 2016). Presently, the major constraint is the ability to transfer methodologies designed for certain case studies to a wider geographic extent (Senf *et al*., 2016), and advancement on how to generalize and transfer current approaches will thus assist in carrying out local to regional analysis of forest disturbances such as SL (Miettinen *et al*., 2014). For example, some methods that were designed for the Amazon basin could perhaps be applied in other tropical regions like Africa or insular Southeast Asia. Examples of such methods are a segmentation-based automated statistical method for gap detection (Pithon *et al*., 2013) and a combination of spectral mixture analysis (SMA) information into one band, using the Normalized Difference Fraction Index (NDFI) to detect forest areas with canopy damage (Souza *et al*., 2005). Research of RS techniques applied in monitoring SL is still an active topic and innovative techniques are required to effectively apply the increasingly diverse and complex remotely sensed data acquired or projected to be soon acquired from space and airborne sensors.

High spatial resolution (5-10 m pixel size) images are viable for SL analysis since most of the features found in logging environments such as logging roads, log landings, and tree fall gaps can be easily recognized (Antropov *et al*., 2015). There are satellites (or UAVs or manned aircraft) now collecting data at very high spatial resolutions, up to 31 cm pixels for Worldview-3. But finer resolution images are expensive and contain computational constraints. Locating the boundaries of logged areas by applying visual interpretation on finer resolution images is not easy. Wulder *et al*. (2008) reported that automated detection of SL using very high-resolution satellite data has technical challenges:

- differences in view-illumination geometries between the pairs of images cause a problem with tree crown geo-location; and
- the appearance of artefacts due to canopy shadowing can induce misclassification of disturbance.

Synthetic aperture radar (SAR), operating at microwave frequencies, is one of the most promising RS methods for the mapping of forest disturbances (Mermoz *et al*., 2015). The amount of radiation backscattered to the sensor keeps on increasing as the number and/or size of trees present in an area increases, therefore, radar satellites have been used to map AGB, and degradation (Mitchard, 2016). Synthetic aperture radar (SAR) enables imaging in all weather conditions, at any time of day or night, and long-wavelength radar signals can penetrate canopies and have been related to forest structure and woody biomass up to a saturation limit, higher for longer wavelengths (Mermoz *et al*., 2015). However, despite these advantages, RADAR has various constraints limiting its application in studying SL, for several reasons (Whittle *et al*., 2012):

- the limited SAR data availability;
- the low data dimensionality for classification algorithms;
- the sensitivity of the SAR signal to surface moisture, and

- the absence of easy-to-use tools and methods for data interpretation.

LiDAR data utilizes laser light looking directly down to estimate tree height and structure (Mitchard, 2016). Repeat surveys can thus detect the removal of individual trees, and therefore, it is the only RS method that can guarantee to map SL with high accuracy, but the applications of present LiDAR RS sensors for the detection of SL, fall short in cost-efficiency, timing, and logistics (Ellis *et al.*, 2016). LiDAR sensor and mission specifications remain dissimilar, and this affects the level of confidence that is observable about SL from the same forest, therefore, the assumption behind model-based inference does not remain valid (Andersen *et al.*, 2014). No LiDAR satellite is currently collecting data, so LiDAR can only be obtained through aircraft or UAVs, at a high cost (Mitchard, 2016). This review has shown that optical data is the most widely used, followed by RADAR and LiDAR. This is the opposite of their data content, with LiDAR being the richest source of data since it provides full cross-sectional information on forests, while RADAR provides part of these data, and optical RS systems view only the top of the tree canopy (Mitchard, 2016). Data availability, cost, and the complexity of the analysis determine which type of data to use far more than which system would provide the most comprehensive information about a forest. The best method chosen to detect SL must strike a balance between accuracy, cost of data and analysis, and optimal monitoring frequency.

Fusing multiple RS sensor platforms helps address the limitations of some sensors (Mitchell *et al.*, 2017). The combination of optical/SAR can be used for the detection of canopy cover change, while that of SAR/LiDAR can be used in detecting sub-canopy structural change (Mitchell *et al.*, 2017). This can promote the extraction of detailed information. High spatial resolution sensors offer detailed textural information but have the problem of small area coverage. Moderate spatial resolution sensors have daily or near-daily repeat intervals, but they have less spectral and spatial information (Mitchell *et al.*, 2017). The integration of remotely sensed data from several sensors at a variety of spatial and temporal scales is very useful in estimating forest structure and structural change (Asner *et al.*, 2008). Thus, we expect more sophisticated data fusion techniques soon, which can aid in mapping forest disturbances such as SL. Obtaining near-coincident data is an uphill task due to little or complete lack of coordination of optical and SAR satellite observations by space agencies. Commitment by space agencies is a key requirement for systematic and coordinated observation of tropical forests on a sustainable basis and with an open data policy (Mitchell *et al.*, 2017). There is the need to have technological developments associated with unmanned aerial vehicles (UAVs) to collect and share UAV data at broader scales and minimal costs, with caution to current constraints such as space for taking off, landing, and piloting in thick forest environments. Also, the possibility of building and deploying a fleet of high-altitude imaging drones should be explored. Due to the challenges that hamper the acquisition of LiDAR data, the application of ground-based Global Positioning System (GPS) for mapping wall-to-wall infrastructure in similar remote tropical landscapes is recommended because it can be well integrated into the ground-based monitoring plan required for other impact parameters (e.g. felling inefficiencies), is less vulnerable to time lag errors (old skid trails can be detected on the ground), and remains considerably cheaper (Ellis *et al*., 2016). Thus, while the agreement with GPS-based logging infrastructure maps is promising, more research is required to assess the uncertainties of LiDAR-based logging infrastructure mapping. The integration of LiDAR RS and forest inventory schemes will reduce the total costs and the need for extensive field-based sampling. The innovation of accurate automated methods for processing LiDAR data could be of critical help because they would lower the data-processing costs, allowing for data acquisition of extensive areas while providing repeatable and consistent estimates of vital forest attributes (Melendy *et al*., 2018).

2.4.5 Accuracy assessment

In order to enable the combination of field and inventory data with RS data, researchers should publicly archive their field data upon publication, overtly including spatial metadata (Senf *et al*., 2016). For future work to guarantee appropriate interpretation of RS data calls for standardized field data collection, acquired over larger areas and in plots that consider forest disturbance dynamics and what temporal and spatial disturbances can or will be captured on these plots (Senf *et al*., 2016). Improved availability and accessibility of HR and VHR imagery notably will promote the creation of high-quality reference datasets for calibrating RS models of logging disturbances, however, the data must be made accessible to researchers at low costs (Olofsson *et al*., 2014).

2.5. Conclusion

Despite several initiatives for monitoring SL in tropical forests, mapping the areal extent or impacts of SL utilizing RS imagery remains a challenge. Selective logging (SL) is a key contributor to the economic, social, and ecological dynamics of tropical regions. The long history of SL in tropical forests, made them ideal study systems when assessing the long-term effects of selectively logged areas and notifying future management decisions. Nevertheless, logging impacts differed widely, making a generalization, as well as policy implementation, difficult.

While this review documents a significant increase in studies mapping SL in recent years, several challenges were also identified which require improvement:

- Forest monitoring programs to assess degradation, especially from selectively logged activities, are needed elsewhere within the tropical forests.
- Improved quality and accessibility of RS data applicable at different spatial scales would lead to more diversification of RS users.
- More integration of RS into scientific disciplines dealing with other types of research questions about degradation by SL to further the scientific involvement of this rapidly developing technology.
- More sophisticated data fusion techniques should be developed, that can be more effective in mapping degradation by SL.
- Commitment by space agencies for systematic and coordinated observation of tropical forests on a sustainable basis and with an open data policy.
- Technological developments associated with UAVs to collect and share UAV data at broader scales and minimal costs are needed.
- The innovation of accurate automated methods for processing LiDAR data could be of critical help because they would lower the data-processing costs, allowing for repeated data acquisition of extensive areas.
- Affordable and reliable methods are required for analysing dense time series of satellite image time series that provide high spatial details.
- Improved generalization and transfer of approaches designed for SL in certain case studies to a wider geographic extent—will assist in local to regional analysis.
- More research on RS techniques applied in monitoring SL is required to effectively apply the increasingly diverse and complex remotely sensed data acquired or projected to be acquired from space and airborne sensors.
- Researchers should standardize field data collection to guarantee appropriate interpretation of RS data.

Solving these concerns in future research can assist in integrating RS-based maps into operational forest management, and support the inclusion of SL disturbances as integrated processes into local, regional and global ecosystem models.

CHAPTER THREE

Machine learning classification of endangered tree species using WorldView-2 imagery and imbalanced dataset

This chapter is based on:

Jackson, C.M., and Adam, E. (2021). Machine learning classification of endangered tree species in a tropical submontane forest using WorldView-2 multispectral satellite imagery and imbalanced dataset. *Remote Sensing*, 13, 4970.

Abstract

Accurate maps of the spatial distribution of tropical tree species (TS) provide valuable insights for ecologists and forest management. The discrimination of TS for economic, ecological, and technical reasons is necessary for achieving promising results in TS mapping. Algorithm performance may be affected by imbalanced datasets. An imbalanced class distribution of a dataset is characterized by cases where some classes have more samples than others, leading to imbalanced training set. Most classifier learning algorithms that assume a fairly balanced distribution, such as random forest (RF) and support vector machine (SVM) are affected by imbalanced training data. Relatively rare classes will consist of smaller proportions of the training set, and vice versa. The original dataset was imbalanced despite efforts to add samples of rare species, such as *Prunus Africana* and *Albiz-zia gummifera*. This study aimed to assess the effects of imbalanced data in mapping trees species under threat in a selectively logged sub-montane heterogeneous tropical forest. The random forest (RF) and support vector machine with radial basis function (RBF-SVM) kernel classifiers, and the high resolution WorldView-2 (WV-2) multispectral imagery were used. For comparison purposes, the original imbalanced dataset was standardized using three data sampling techniques: oversampling, undersampling, and combined sampling techniques in R. For an imbalanced dataset, the F1-score—the harmonic mean of precision and recall—was a more appropriate metric. The combined sampling technique produced the best results; F1-scores of 68.56 ± 2.6% for RF and 64.64 ± 3.4% for SVM. The undersampled data reported the lowest F1-score, i.e. 45.8 ± 5% for RF, compared to SVM's 48 ± 6%, unlike the other datasets where RF performed better than SVM. The original dataset's F1 score was 63.8 ± 3.9% for RF and 62.7 ± 4.3% for SVM. The balanced dataset recorded improved classification accuracy compared to the original imbalanced dataset. This research observed that more separable classes recorded higher F1 scores. This research observed that classes that reported higher separability index values such as *Syzygium guineense* recorded higher F1 scores. *Syzygium guineense* reported the highest F1-score of 99 ± 0.8% and 98 ± 0.7%, for RF and SVM, respectively, while *Newtonia buchananii*, the least accurately mapped species achieved F1-score of 38 ± 7.1% for RF and 36 ± 7.9% for SVM. The most important spectral bands with the ability to detect and distinguish between tree species as measured by the RF's variable importance measurement were bands 5 (Red), 6 (Red Edge), 7 (Near Infrared 1), and 8 (Near Infrared 2).

Keywords: Tropical forests. Endangered tree species. Selective logging. Imbalanced data. Pixel-based classification. Machine learning algorithm.

3.1 Introduction

Tropical forests comprise woody, evergreen vegetation, and cover *c*. 47% of the world's total forest area (Solberg *et al*., 2008). Tropical forests host the highest proportion of global tree diversity, i.e. with more than 53,000 tree species (TS), compared to approximately 124 in temperate Europe (Slik *et al*., 2015). The height of the tree crowns forming the forest canopy is in the range of 30 to 50 m, but emergent trees may attain heights of approximately 70 m (Solberg *et al*., 2008). Tropical forests cover *c*. 7% of the globe, and they are home to more than half of all earth's biodiversity (Solberg *et al*., 2008). Vital environmental processes such as the water cycle, soil conservation, carbon sequestration, and habitat protection are immensely regulated by tropical TS (Solberg et al., 2008). Therefore, those forests maintain the ecosystem services and mitigate climate change (Wagner *et al*., 2018). Information on key parameters such as TS, tree diameter, height, crown size, and location are important for resource management, biodiversity assessment, ecosystem services assessment, and conservation (Wagner *et al*., 2018). Ecologists have long been interested in explaining species' distribution in ecosystems (Slik *et al*., 2015), as it can drive the exploitation and management policies of forests. In tropical forests, the maintenance of a canopy composed of emergent trees of fundamental species has been shown to provide conditions favourable to ecological processes, playing an essential role in the forest community's resilience and perpetuation (Ferreira *et al*., 2019).

Traditionally, detailed TS identification is obtained in relatively small areas with time-consuming, high levels of manpower and the associated high costs and often operationally prohibitive field inventories (Ferreira et al., 2019; Immitzer *et al*., 2012). Tropical rainforests are characterized by very high species richness and a lack of access to some parts of the forests (Immitzer *et al*., 2012). Therefore, obtaining exact information about the occurring species using field inventories is almost impossible (Immitzer *et al*., 2012), thus, they have difficulty resolving large geographic patterns. Remote sensing (RS) captures information over extensive areas in fine detail (Nagendra and Rocchini, 2008). Mapping trees at species level using RS is linked to an understanding that species have unique spectral signatures associated with characteristic biochemical and biophysical properties (Clark *et al*., 2005).

Several studies have assessed the potential of multispectral data in mapping TS in tropical forests. The earliest study by Clark *et al*. (2005) attempted a classification of seven species of emergent trees in a tropical rain forest in Costa Rica using hyperspectral digital collection experiment (HYDICE) imagery. Spectral-based species classification was performed using linear discriminant analysis (LDA), maximum likelihood, and spectral angle mapper (SAM) classifiers applied to combinations of bands from a stepwise-selection procedure. Zhang *et al*. (2006) selected five species to assess intra- and inter-class variability of TS using a high spectral and spatial resolution imagery acquired using the airborne sensor HYDICE data. Trichon and Julien (2006) used two sets of aerial photographs to identify TS through air photo interpretation. Somers and Asner (2012) performed a hyperspectral time series analysis of two native and two invasive species in Hawaiian rainforests, using the separability index (SI). Féret and Asner (2012) applied semi-supervised support vector machine (SVM) classification using tensor summation kernel to identify individual crowns using imaging spectroscopy and light detection and ranging (LiDAR). Clark and Roberts (2012) mapped seven tropical rainforest TS using hyperspectral data. Narrowband indices, derivative- and absorption-based techniques, and spectral mixture analysis (SMA) were used to derive metrics that respond to vegetation chemistry and structure. The random forest (RF) classifier was used to discriminate species with minimally-correlated, importance-ranked metrics. Papeş *et al*. (2013) used Earth Observing-1 Hyperion satellite hyperspectral imagery to spectrally discriminate between crowns of 42 individual trees of 5 taxa using LDA, and also evaluated seasonal variation in species discriminations related to phenology. Féret and Asner (2013) applied supervised classification to identify canopy species using airborne hyperspectral imagery acquired with the Carnegie Airborne Observatory-Alpha system. Singh *et al*. (2015) mapped and characterized selected TS using aerial data. To delineate individual tree crowns (ITCs) from very-high-resolution (VHR) aerial imagery and LiDAR data, the study used object-based image analysis (OBIA). Both maximum likelihood and spectral angle mapper (SAM) classifiers were applied to the aerial imagery. Other studies have combined hyperspectral and LiDAR sensors. For example, Baldeck *et al*. (2015) used airborne imaging spectroscopy to identify individuals of three focal canopy TS in a diverse tropical forest on Barro Colorado Island, Panama. The addition of co-registered LiDAR data further improved performance by identifying intra- and inter-canopy shadows that alter species signatures (Baldeck *et al*., 2015). Ferreira *et al*. (2016) used LDA, RF, and SVM classifiers on hyperspectral and multispectral data to discriminate and map tree species, at a pixel level. Simulated WorldView-3 (WV-3) data were used to assess the role of SWIR bands in species classification. A tree crown segmentation approach was used to hyperspectral data to map tree species. Graves *et al*. (2016) assessed the accuracy of a SVM model with a highly imbalanced dataset using a hyperspectral image mosaic. Wagner *et al*. (2018) applied automatic ITC delineation in a highly diverse tropical forest using WV-2 satellite images. Ferreira *et al*. (2019) applied visible to shortwave infrared WV-3 images and texture analysis to classify tropical TS in a semi-deciduous forest in different seasons.

Conventional multispectral sensors such as Landsat or MODIS lack both the spatial and spectral resolution to detect changes in TS composition (Nagendra and Rocchini, 2008; Immitzer *et al*., 2012). Other satellite-borne multispectral sensors such as IKONOS or QuickBird have a high spatial resolution but they lack the spectral resolution to map TS in tropical forests (Nagendra and Rocchini, 2008). Air-borne cameras that provide the highest spatial resolution do not offer the spectral resolution required (Immitzer *et al*., 2012). Moreover, aerial photos' wide field of view results in strong effects from bi-directional reflectance characteristics of most land cover types (Immitzer *et al*., 2012). Therefore, the spectral signature of an object can differ significantly. Tropical forests consist of trees of different species and ages, growing close to each other, with their crowns intertwined, and this made TS mapping a challenge. Their mapping requires remote sensing (RS) systems that can provide high spatial resolution as well as high spectral resolution. Airborne hyperspectral and LiDAR sensors enable mapping at very fine scales, but the high cost related to hyperspectral and LiDAR data acquisition and processing has hindered their application in mapping TS in tropical forests (Nagendra and Rocchini, 2008). However, some of the inherent features of hyperspectral data such as the carotenoids and chlorophyll-sensitive bands are preserved in WV-2 multispectral data (Mutanga *et al*., 2012). Thus, the high resolution (HR) multispectral sensor, WV-2, showed great potential to generate the information required for the identification of TS and canopy attributes in complex tropical forest environments (Ferreira et al., 2016).

Kenya's rainforest cover, mostly montane forests, are scattered and degraded patches (NEMA, 2010). This can be attributed to rapidly increasing population along forest edges, climate change, poor governance, changing societal needs and values, and politically motivated excisions (NEMA, 2010). The intense growth of the population around Mt. Kenya Forest Reserve (MKFR) since the early 1970s led to the illegal logging of important timber trees (KWS, 2010). This greatly reduced plant diversity and the regenerative capacity of such TS (Wass, 1995; Bussmann, 1996; Ng'eno, 1996; KWS, 2010). More demand than supply for wood in Kenya has led to over-exploitation of high-value tree species (KFS, 2010). Some of the targeted TS are African pencil-cedar (*Juniperus procera*), Wild olive (*Olea europaea*), East African rosewood (*Hagenia abyssinica*), East African camphor (*Ocotea usambarensis*), red stinkwood (*Prunus africana*), East African newtonia (*Newtonia buchananii*), East African yellow-wood (*Podocarpus spp*), East African olive (*Olea capensis*), Meru oak (*Vitex *keniensis*), and Peacock flower (*Albizzia gummifera*) (Wass, 1995; Bussmann, 1996; Ng'eno, 1996; KWS, 2010).

In addition to the number and quality of the training samples, algorithm performance may be affected by class imbalance (Chawla *et al*., 2004; Graves *et al*., 2016; Maxwell *et al*., 2018). Class imbalance occurs when the number of reference samples (Table 3.1), varies among the classes, leading to imbalanced training set (Chawla *et al*., 2004; Graves *et al*., 2016; Maxwell *et al*., 2018). Imbalanced data may occur in a deliberative sample, and it is also expected in random sampling (Chawla *et al*., 2004; Graves *et al*., 2016; Maxwell *et al*., 2018). In simple random sampling, the chance of choosing a class is related to the areal coverage of the class. Thus, relatively rare classes will consist of smaller proportions of the training set. Therefore, this study aimed to use imbalanced dataset to identify and map TS under threat in a selectively logged sub-montane heterogeneous tropical forest using RF and SVM classifiers and WV-2 multispectral imagery. For comparison purposes, the original imbalanced dataset was standardized using oversampling, undersampling, and combined sampling techniques. In addition, the explanatory power of the WV-2 spectral bands in discriminating the TS was evaluated. The maps showing the spatial distribution and abundance of the endangered TS in the study area formed the basis by which efforts can be made for the restoration of the TS, among other ecological applications.

3.2 Materials and methods

3.2.1 Acquisition and pre-processing of WorldView-2 satellite data

The study site was covered by one WV-2 scene, acquired on 30 January 2019. The image utilized in this study was pre-processed and orthorectified by the image distributor (DigitalGlobe, 2019). It was geo-referenced to World Geodetic System (WGS) 1984 datum and the Universal Transverse Mercator (UTM) zone 37S projection. Theoretically, the radiometric correction and calibration of image data are necessary for detailed feature extraction (Jensen, 2005). Atmospheric effects can complicate the spectral separability between landscape features with fine spectral differences (Jensen, 2005). Therefore, for optimal feature extraction, the WV-2 image was atmospherically calibrated by converting the digital numbers (DN) to the top-of-atmosphere reflectance. The conversion was done using the ENVI module (ENVI 5.3) FLAASH.

3.2.2 Field data collection

Because a published inventory of threatened plant taxa for Kenya does not exist, most botanists discovered a rising number of species requiring special protection (GoK, 2015). Different sources have tried to classify TS considered to be threatened or endangered. For example, a State of the Environment report by the National Environment Management Authority (NEMA) classified East African camphor (*Ocotea usambarensis*), Red stinkwood (*Prunus Africana*), African satinwood *(Zanthoxylum gilletii)*, East African sandalwood (*Osyris lanceolata*), and Meru oak (*Vitex keniensis*) as endangered (NEMA, 2010). Ng'eno (1996) listed *Prunus africana* as an endangered tree species in Kenya, and *Podocarpus spp.*, *Vetex keniensis*, *Newtonia buchananii*, and *Albizia gummifera* are threatened species. *Hagenia abyssinica*, *Juniperus procera*, and *Ocotea usambaraensis* are listed as restricted tree species. According to KWS (2010), threatened plant species in Kenya include *Prunus africana*, *Vitex keniensis*, and *Ocotea usambarensis*, among others. All these TS are commercial indigenous trees of Kenya (GoK, 2013). Therefore, field data collection involved the following TS: *Prunus Africana*, *Zanthoxylum gilletii*, *Albizia gummifera*, and *Newtonia buchananii*. Other species covered in this study were *Anthocleista grandiflora*, *Syzygium guineense*, and *Macaranga kilimandscharica*; they are not currently under any threat, however, an increase in the volume of extraction could endanger them. Rampant SL targeting especially *Ocotea usambarensis* made the species very rare in the study site, and only eight samples were identified in the field. Other landscape classes, i.e., other woody vegetation and shadow, were sampled. Other woody vegetation class consists of all species with very few samples, brought together to form a single mixed-species class. This allowed for their inclusion in the model so that they could be mapped, although individual species distinctions could not be made.

The ground truth points were collected in January and February 2020. Due to very steep mountainous terrain and dense forest cover, only a few areas were accessible, therefore, the field campaign randomly selected TS along trails traversing the study area. The field measurement involved the identification of TS, tree height, diameter at breast height (DBH), tree dominance, and size, which varied among stands. A handheld Global Positioning System (Garmin eTrex® 20 GPS Receiver) and an RGB false-colour composite (Near Infrared 2: Yellow: Coastal) of WorldView-2 image (pixel = 1.89 m), aided in locating tree crowns in the field. Using ArcGIS (ArcGIS v. 10.3®, ESRI, Redlands, CA, USA), points were set on all relevant crown pixels on the WorldView-2 imagery. By following the edges of the pixels, the points were converted into polygons. Pure pixels—without contamination from lianas or neighbouring trees—were carefully extracted (Clark et al., 2005). Also noted was any additional valuable information, such as tree flowering and leaves. Tree species (TS) information was assigned to each tree crown. The collected field data resulted in an imbalanced dataset, i.e. with common species having many samples and less common species contained few samples (Table 3.1). The dataset was randomly partitioned into 70% for training and 30% for testing.

Table 3.1: *List of tree species, codes, family, leafy phenology, diameter at breast height (DBH), number of individual tree crowns, and composition of training and test data for each species used to map trees in Mt. Kenya Forest Reserve.*

Botanical Name	Code	Family	Leaf Phenology	DBH (m)	Train Data	Test Data	Total
Macaranga k.	MK	Euphorbiaceae	Semi-deciduous	0.30–0.57	56	24	80
Zanthoxylum g.	ZG	Rutaceae	Semi-deciduous	0.97–3.00	56	24	80
Syzygium g.	SG	Myrtaceae	Evergreen	1.00–2.50	56	24	80
Newtonia b.	NB	Fabaceae	Deciduous	1.50–4.50	56	24	80
Anthocleista g.	AnG	Gentianaceae	Evergreen	1.01–2.50	36	15	51
Prunus a.	PA	Rosaceae	Evergreen	1.24–3.50	22	10	32
Albizzia g.	AlG	Fabaceae	Deciduous	1.37–4.00	20	8	28
Other woody vegetation	OWV	---	---	---	56	24	80
Shadow	SD	---	---	---	56	24	80

Shown in Figure 3.1 are samples of the TS in the study area—*Prunus Africana*, *Zanthoxylum gilletii*, *Albizia gummifera*, *Newtonia buchananii*, *Anthocleista grandiflora*, *Syzygium guineense*, and *Macaranga kilimandscharica*

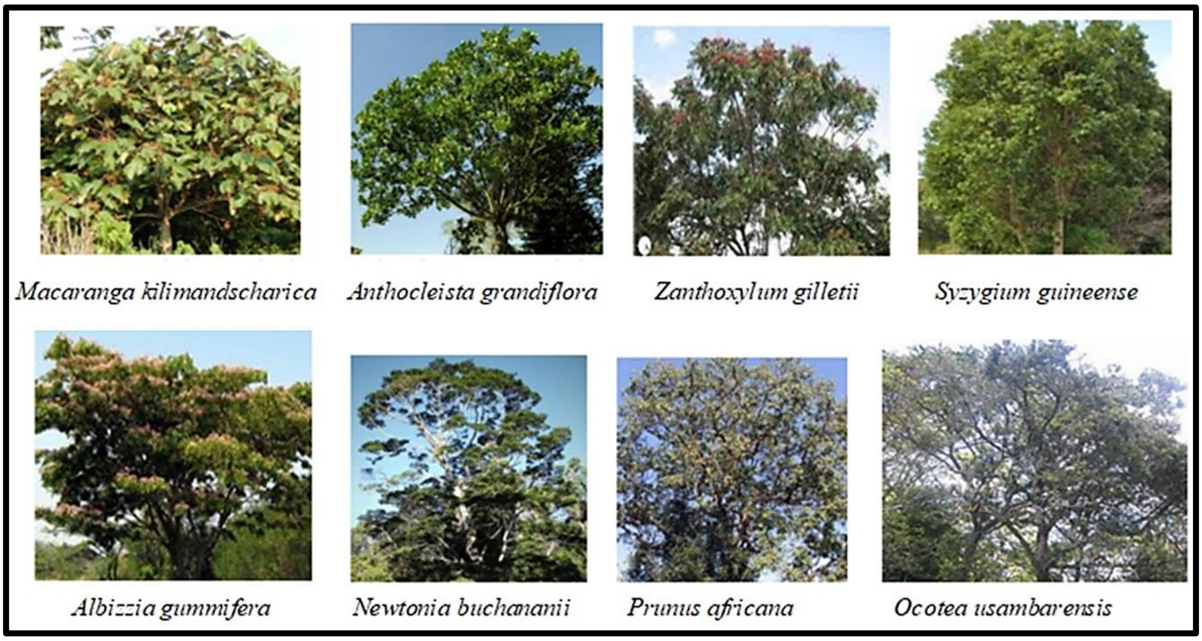


Figure 3.1. *Samples of endangered tree species in Mt. Kenya Forest Reserve in Chuka, Tharaka Nithi County, Kenya.*

3.2.3 Spectral separability

The first step in mapping tree species is to determine if the classes are spectrally different. On-screen digitizing of samples from the display of the WV-2 false-colour composite was implemented to generate the respective signature files. An assessment of the signatures was done to examine the spectral properties of the training sample classes and their separability over others. The spectral information was extracted using the central pixel within the crown polygon. This study used the Jeffries-Matusita (J-M) distance—a commonly used spectral separability measure in RS applications—has a good potential to be used for evaluation of the effectiveness of a feature in discriminating two classes (Richards and Jia, 1999). The J-M distance among the distributions of two classes ω_i and ω_j has been defined by Equation 3.1 as follows (Richards and Jia, 1999):

$$J\text{-}M_{ij} = 2(1 - e^{-B_{ij}}), \tag{3.1}$$

where B_{ij} is the Bhattacharyya distance calculated as (Kailath, 1967) (Equation (3.2)):

$$= \frac{1}{8}\left(\mu_i - \mu_j\right)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} \left(\mu_i - \mu_j\right) + \frac{1}{2} ln\left(\frac{1}{2}\frac{|\Sigma_i + \Sigma_j|}{\sqrt{|\Sigma_i||\Sigma_j|}}\right) \tag{3.2}$$

where μ_i and μ_j are the mean reflectances of species i and j, and Σ_i and Σ_j correspond to their covariance matrices, whereas $|\Sigma_i|$ and $|\Sigma_j|$ are the determinants of Σ_i and Σ_j, respectively. ln is the natural logarithm function, and T is the transposition function.

As a general rule, if the result is greater than 1.9, then the classes can be separated—between 1.7 and 1.9, the separation is fairly good—below 1.7, the separation is poor (Jensen, 2005). The J-M distance was calculated for all pairwise combinations between the mean spectral reflectance of the samples of a species and those of all other species, one at a time. The VNIR (450-1040 nm) WV-2 bands were used for computing the J-M distance.

3.2.4 Training of random forest and support vector machine classifiers

The RF and SVM algorithms were applied using RStoolbox in R software in a pixel-based classification setting approach. The RF algorithm groups decision trees and the splitting variables are randomly chosen feature subsets with bagged samples (Breiman, 2001). Normally, the number of decision trees (*ntree*) in the ensemble and the number of predictor variables (*mtry*) randomly selected at each node need to be defined before applying the RF model (Kuter, 2021). To extract the best-performing parameters to be utilized in training the algorithm, a 10-fold grid search method was applied (Kuter, 2021). RF brings together many weak learners, i.e. decision trees, into a stronger predictor by aggregating the predictions from all decision trees (Breiman, 2001). The majority 'vote' of all the trees is used to allocate a final class for unknown features. The mean decrease in accuracy (MDA) was used to evaluate the explanatory power of the input variables (Breiman, 2001). The MDA shows how much accuracy the model loses by excluding each variable (Breiman, 2001). The higher the MDA value, the more important the variable in the model. The classification models also evaluated the ability of the WV-2 spectral bands to detect the TS.

The SVM finds the position of the optimal separating hyperplane, i.e. decision boundary, which meets two ultimate goals at a go (Vapnik, 2000):

- separate the original data while maximizing the margin between classes, and
- minimize the misclassification error (Vapnik, 2000).

In the present study, the radial basis function (RBF) was selected as the kernel function because it has fewer parameter values to predefine, and it is as robust as other kernel types (Maxwell *et al*., 2018). The success of an SVM model depends on the C (penalty) and γ (Gamma) parameters in the kernel function. The goal is to identify the best C and γ for the TS mapping problem so that the classifier can accurately predict unknown data (Kuter, 2021). The

SVM parameters, i.e. gamma and cost, were optimized for the WV-2 dataset, also using a 10-fold grid search approach.

3.2.5 Class imbalance

A dataset's imbalanced class distribution is characterized by cases where some classes have more samples than others. Most classifier learning algorithms that assume a fairly balanced distribution have found it challenging (Chawla *et al*., 2004; Maxwell *et al*., 2018). Previous studies have shown that the *k*-NN (*k*-nearest neighbour), RF, and SVM algorithms are affected by imbalanced training data (Maxwell *et al*., 2018).

To evaluate the impact of training data imbalance in mapping endangered TS, this study applied data sampling techniques that aimed to modify the imbalanced dataset into balanced distribution, i.e., by altering the overall number of samples used in the training (Maxwell *et al*., 2018). In choosing the best model for the classification of TS in the study area, four datasets were compared:

- Category 1 was composed of the original imbalanced dataset (Table 3.1) from the field.
- In category 2, the training data for all classes were randomly oversampled to match that of the majority of classes, i.e. 56 data points.
- Category 3 consisted of the same number of samples for all classes, which was attained by randomly undersampling the training data for all classes to match that of the smallest class, i.e. 20.
- Finally, category 4 comprised a dataset created by combining both downsampling and oversampling techniques—the oversampling technique was applied to create a duplicate and artificial data points—the undersampling technique followed to eradicate noise data points, thus creating a robust balanced dataset suitable for model training.

Therefore, the last three categories are subsets of the original field acquired dataset. In all cases, the datasets were split into test and training data before running the models. Running models before splitting datasets can allow identical samples to be present in both the test and training data, leading to the models overfitting the training data (Graves *et al*., 2016). The sampling techniques were only applied to training data to avoid model overfitting. In all models, the 10-fold cross-validation method was repeated 10 times (Graves *et al*., 2016).

3.2.6 Measures of model performance

A multidimensional scaling (MDS) technique was applied in the analysis of (dis)similarities in the TS datasets. MDS calculates a (dis)similarity matrix among pairs of TS and then displays the data in a low-dimensional representation (Borg and Groenen, 2005). Therefore, MDS shows (dis)similarities among pairs of TS as distances between points in a low-dimensional space. The Euclidean distance formula (Equation (3.3)) can be used to calculate the distance between two points, e.g. *i* and *j*, as follows (NCSS, 2021):

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^{2}} \tag{3.3}$$

where p is the number of dimensions, d_{ij} is the distance, and x_{ik} is the data value of the *ith* row and *kth* column. The dissimilarity between points *i* and *j* is denoted d_{ij} and s_{ij} for similarity. Small d_{ij} values indicate points that are close together, thus, they belong to the same group and vice versa. The similarity values are the opposite, i.e. small s_{ij} values indicate points that are farther apart, hence they are not in the same group, and vice versa (Buja *et al*., 2007). The goodness-of-fit statistic, the stress measure (Equation (3.4)), which is based on the differences between the observed data and their predicted values, was used to express how well the datasets are represented by the model (NCSS, 2021).

$$stress = \sqrt{\frac{\sum (d_{ij} - đ_{ij})^{2}}{\sum d_{ij}^{\ 2}}} \tag{3.4}$$

where $đ_{ij}$ is the predicted distance based on the MDS model.

To fully evaluate model effectiveness, precision and recall performance metrics were used. Precision measures how accurate is the classifier's prediction of a class (Graves *et al*., 2016). Low precision indicates a high number of false positives. Recall, also known as sensitivity, measures the classifier's ability to identify a class (Graves *et al*., 2016). Low recall indicates a high number of false negatives. For an imbalanced dataset, the F1-score (Equation (3.5)) was a more appropriate metric. It is the harmonic mean of the precision and recall, and the expression is (Graves *et al*., 2016):

$$F1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.5}$$

To evaluate the performance of RF and SVM algorithm classifiers on mapping the selected TS using WV-2 data, 30% of the ground truth data were used to generate confusion matrices from which precision, recall, and F1-score were calculated. The overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) were also calculated. The PA is the ratio of the correctly detected trees to all the positive ground truth tree samples, whereas the UA is the ratio of the correctly detected trees to all the positive model-predicted trees samples (Graves *et al*., 2016).

The McNemar test (Equation (3.6)), a non-parametric test focused on the binary distinction between correctly classified and misclassified class allocations of two classification outputs, was used to compare and indicate the statistical significance of any difference in results (Foody and Mathur, 2004):

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{3.6}$$

where f_{12} is the total number of samples classified correctly by the first classification, but misclassified by the second, and f_{21} is the total samples classified correctly by the second classification but misclassified by the first one. A difference in accuracy between the confusion matrices of the different models is regarded as statistically significant ($p \leq 0.05$) if $Z \geq 1.96$ (Foody and Mathur, 2004).

3.3 Results

3.3.1 Spectral separability between tree species

Figure 3.2 shows the mean spectral signatures for the eight TS in the study area. In the visible bands (bands 1-5), *Syzygium guineense* had higher reflectance values compared to the other seven TS which reflected almost the same values. The Coastal band (band 1) shows the highest reflectance values for all TS. The difference in reflectance between *Syzygium guineense* and the other TS kept on increasing from the Coastal band (band 1) towards the Red band (band 5).

The reflectance values in the Red Edge (band 6), Near Infrared 1 (band 7), and Near Infrared 2 (band 8) were generally higher for the deciduous and semi-deciduous TS than for the evergreen TS. *Zanthoxylum gilletii* and *Albizzia gummifera* showed the highest reflectance values and differed significantly from the others. Among the evergreen TS, *Prunus africana* showed the highest reflectance values, followed by *Anthocleista grandiflora* (Figure 3.2). *Syzygium guineense* had notably low reflectance values. Among the three bands, Near Infrared

1 (band 7) showed the highest reflectance values among all TS. The Near Infrared 1 (band 7) and Near Infrared 2 (band 8) showed the largest differences in reflectance between the seven TS in the study area.

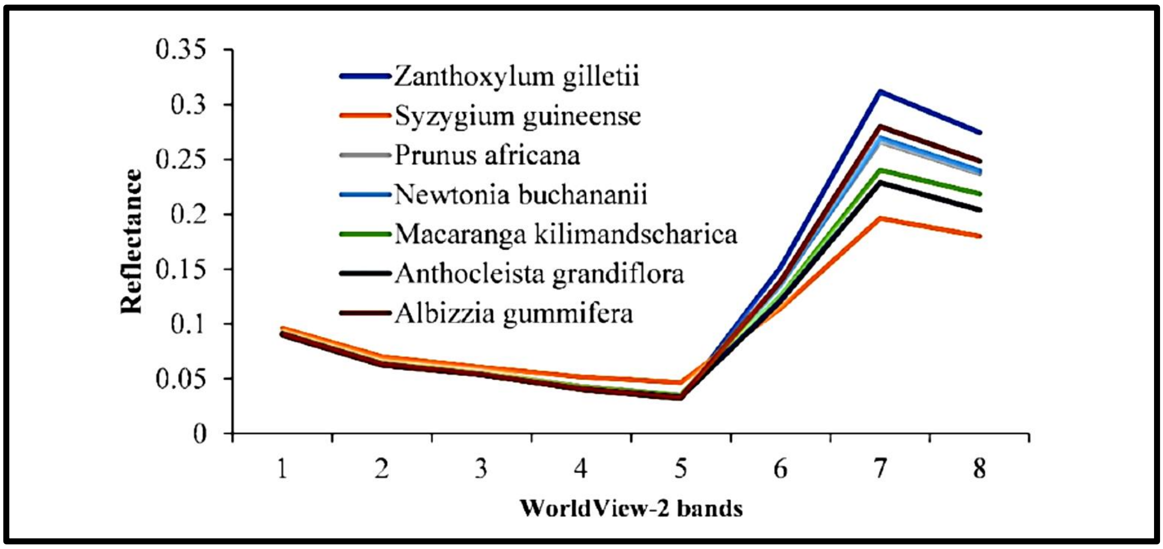


Figure 3.2. *Tree species mean spectral signatures derived from the WorldView-2 data.*

Figure 3.3 shows significant spectral overlaps between TS. Band-specific within-species variance was evident, e.g. *Syzygium guineense* within-species variance was bigger in Coastal band (band 1) and Blue band (band 2), and quite small in the Near Infrared 2 (band 8). *Zanthoxylum gilletii* had the smallest within-species variance in the Red band (band 5), whereas the Coastal band (band 1) and Green band (band 3) showed larger variances.

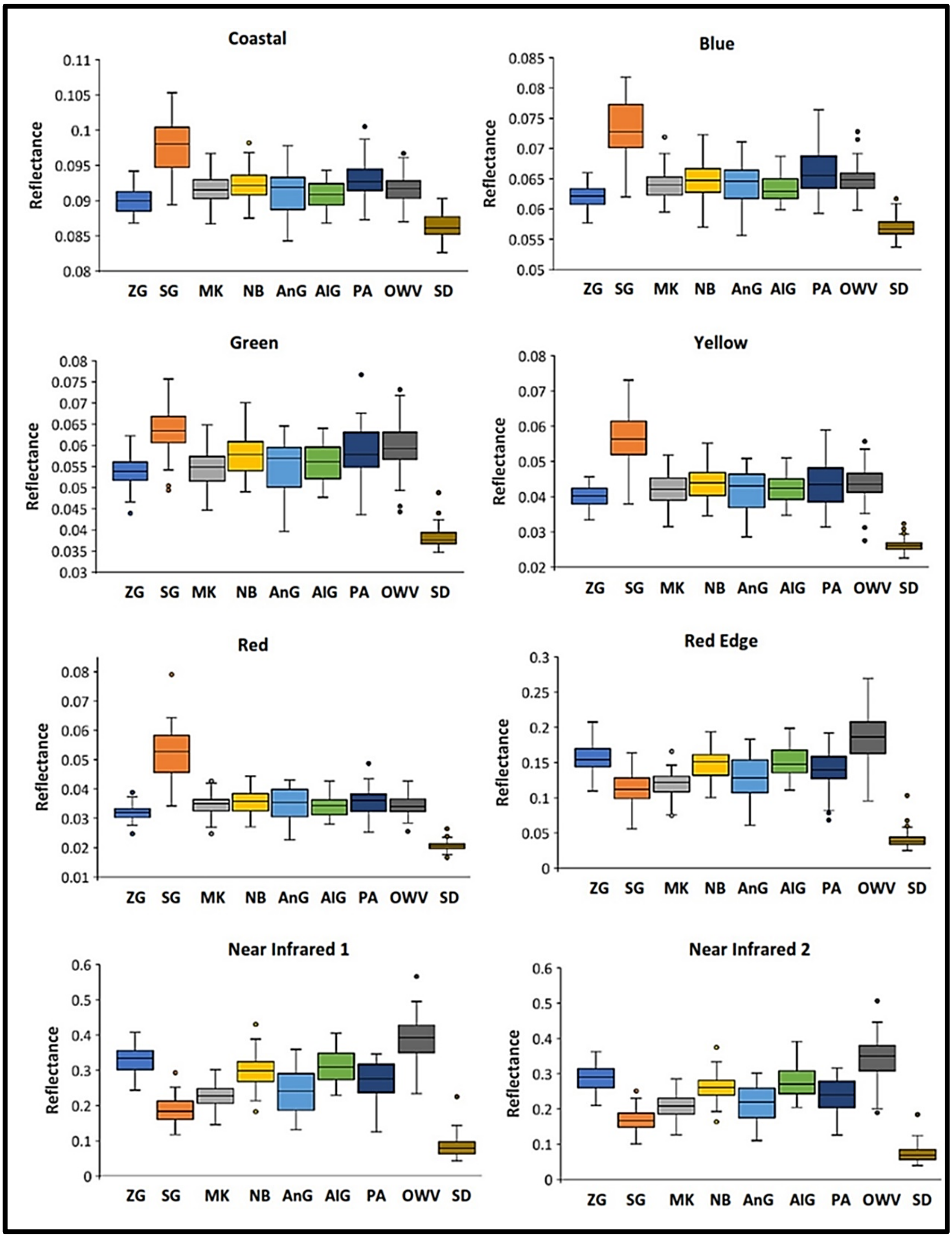


Figure 3.3. *Box-whisker plots showing tree species (TS) mean reflectance values derived from WorldView-2 (WV-2) data. AlG–Albizzia gummifera, AnG–Anthocleista grandiflora, MK–Macaranga kilimandscharica. NB–Newtonia buchananii, OWV–other woody vegetation, PA–Prunus Africana, SD–Shadow, SG–Syzygium guineense, ZG–Zanthoxylum gilletii.*

Table 3.2 is a matrix showing the inter-specific spectral separability of the TS calculated from the WV-2 VNIR bands using the J-M distance. The highest J-M distances were reported for *Syzygium guineense*. The lowest values were 0.59 for separability between *Newtonia buchananii* and *Albizzia gummifera*, and 0.61 for *Newtonia buchananii* and *Anthocleista grandiflora*. For separable TS, their values are shown in bold.

Table 3.2: *Tree species' inter-specific spectral separability as calculated by the Jeffries-Matusita (J-M) distance (Equation (3.1)). AlG–Albizzia gummifera, AnG–Anthocleista grandiflora, MK–Macaranga kilimandscharica. NB–Newtonia buchananii, OWV–other woody vegetation, PA–Prunus Africana, SD–Shadow, SG–Syzygium guineense, ZG–Zanthoxylum gilletii.*

	SG	ZG	AnG	AlG	PA	NB	MK
SG		**1.94**	**1.71**	**1.74**	**1.80**	**1.77**	1.64
ZG			1.43	1.06	1.49	1.26	**1.73**
AnG				0.72	0.84	0.61	0.93
AlG					0.92	0.59	1.23
PA						0.82	1.27
NB							1.05
MK							

3.3.2 Optimization of random forest and support vector machine

In all models, best performing *ntree* and *mtry* parameters used in training of the RF algorithm, were extracted using repeated 10-fold cross-validation (CV) technique. The SVM parameters were optimized using a similar approach for the WV-2 dataset. For each of the four models, the best parameter combinations are summarized in Table 3.3. The dataset produced by the combined sampling techniques performed better than the others. For the RF model, the default *mtry* value of 2 and *ntree* value of 3500 produced the lowest OOB error rate of 0.271 (Figure 3.4a). As seen in Figure 3.4b, a gamma value of 1 and a cost value of 100 yielded the best performance with a cross-validation error of 0.326. For the undersampled dataset, the best performance was an OOB error rate of 0.476 produced by an *mtry* value of 2 and an *ntree* value of 3500. The SVM yielded a gamma value of 0.01 and a cost value of 10, with a cross-validation error of 0.422.

Table 3.3: *The random forest (RF) and support vector machine (SVM) model optimization parameters.*

Model	Train Data per Species	RF			SVM		
		mtry	*ntree*	OOB Error	Gamma	Cost	CV Error
Original dataset	Refer to Table 3.1	3	1500	0.377	0.01	1000	0.387
Oversampling	56	4	4500	0.279	0.01	1000	0.327
Undersampling	20	2	3500	0.476	0.01	10	0.422
Combined sampling	varied	2	3500	0.271	1	100	0.326

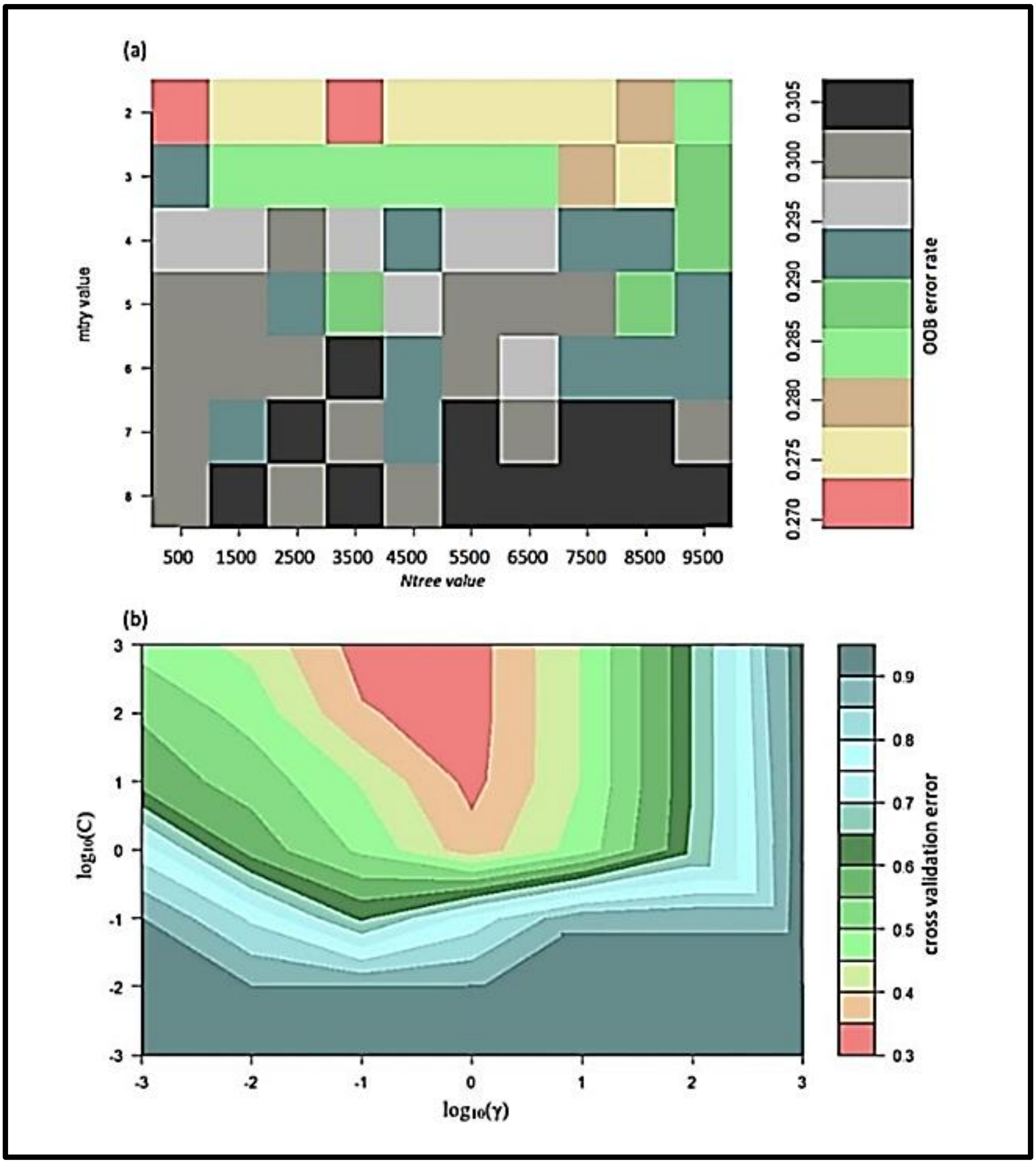


Figure 3.4. *Optimization by 10-fold cross-validation: (a) random forest parameters (mtry and ntree), and (b) support vector machine parameters (gamma and cost).*

3.3.3 Relative importance of variables

After running the classification models, the MDA provided the relative importance of each band (Figure 3.5). The most important spectral bands in all models, i.e. bands depicted by the highest MDA were the Red (band 5), Red Edge (band 6), Near Infrared 1 (band 7), and Near Infrared 2 (band 8).

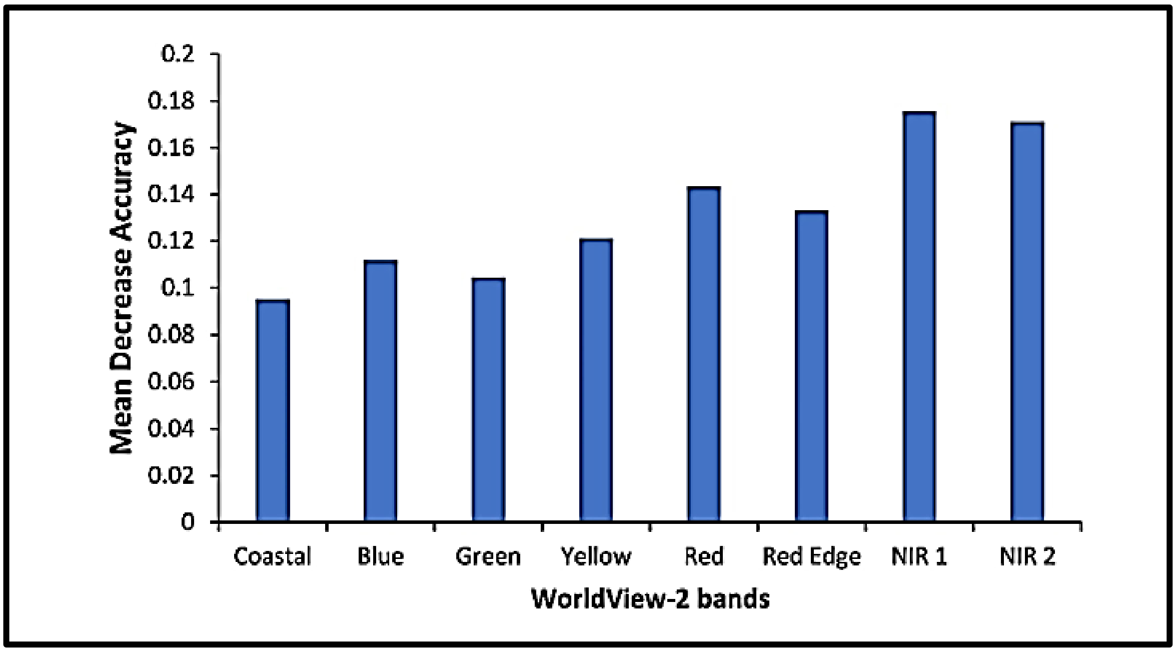


Figure 3.5. *The relative importance of WorldView-2 bands in discriminating between different tree species as measured by RF classifier.*

Figure 3.6 shows that the Red band (band 5) was critical in the identification of especially *Syzygium guineense*, and also involved to a greater extent in mapping *Albizzia gummifera*, *Anthocleista grandiflora*, *Prunus africana*, and *Zanthoxylum gilletii*. The Near Infrared 1 (band 7) contributed significantly to the mapping of *Albizzia gummifera* and *Prunus africana*, as well as *Anthocleista grandiflora*, *Macaranga kilimandscharica*, *Newtonia buchananii*, and *Zanthoxylum gilletii*. The Red Edge (band 6) majorly contributed to the identification of *Macaranga kilimandscharica* and other woody vegetation, and, to a greater extent, shadow. The Near Infrared 2 (band 8) was helpful in the mapping of all TS at different proportions, but it was very vital in the identification of *Albizzia gummifera*, *Anthocleista grandiflora*, and *Prunus africana*.

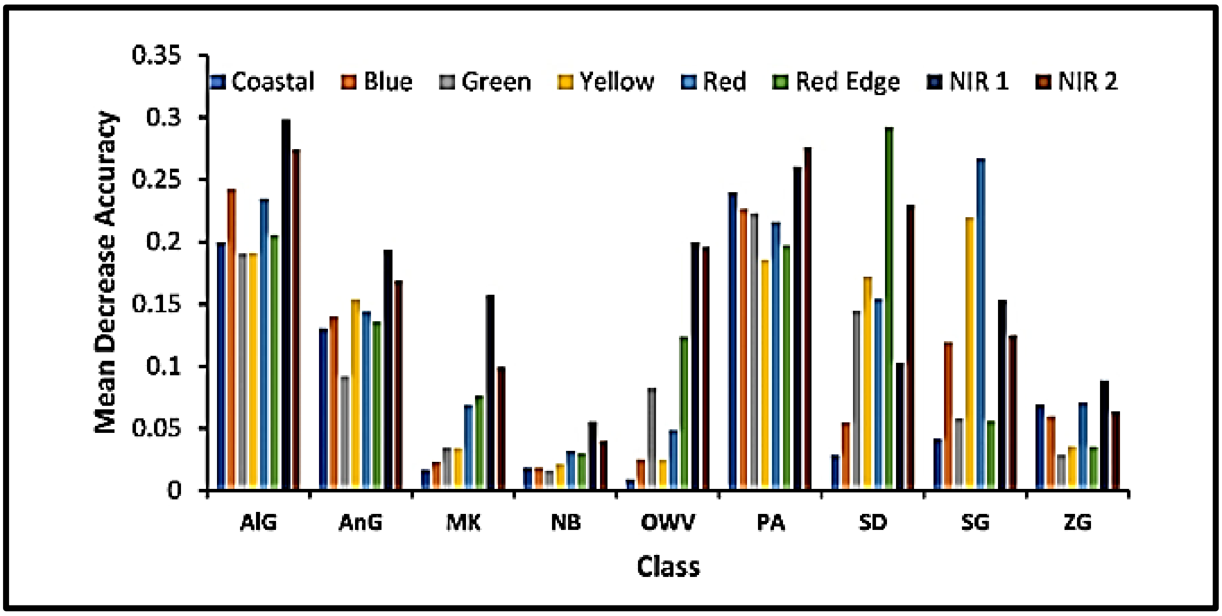


Figure 3.6. *The mean decrease accuracy (MDA) values show the relationship between each tree species and WorldView-2 spectral bands as measured by the RF classifier. AlG–Albizzia gummifera, AnG–Anthocleista grandiflora, MK–Macaranga kilimandscharica. NB–Newtonia buchananii, OWV–other woody vegetation, PA–Prunus Africana, SD–Shadow, SG–Syzygium guineense, ZG–Zanthoxylum gilletii.*

3.3.4 Model performance

The RF multidimensional scaling (MDS) is a nice way to visualize (dis)similarity within and among tree species, using the proximity matrix calculated from the variables in the training dataset (Figure 3.7(a)). SVM multidimensional scaling also visualizes (dis)similarity within and among tree species, using the test data, as seen in Figure 3.7(b). The stress measure (the goodness-of-fit statistic), was used to express how well datasets were represented by the models. The undersampled dataset performed poorly by attaining higher stress values, i.e. 2.41% for RF and 2.09% for SVM. The rest of the datasets performed fairly as their stress values were in the range of 1.41–1.89% for RF and SVM classifiers.

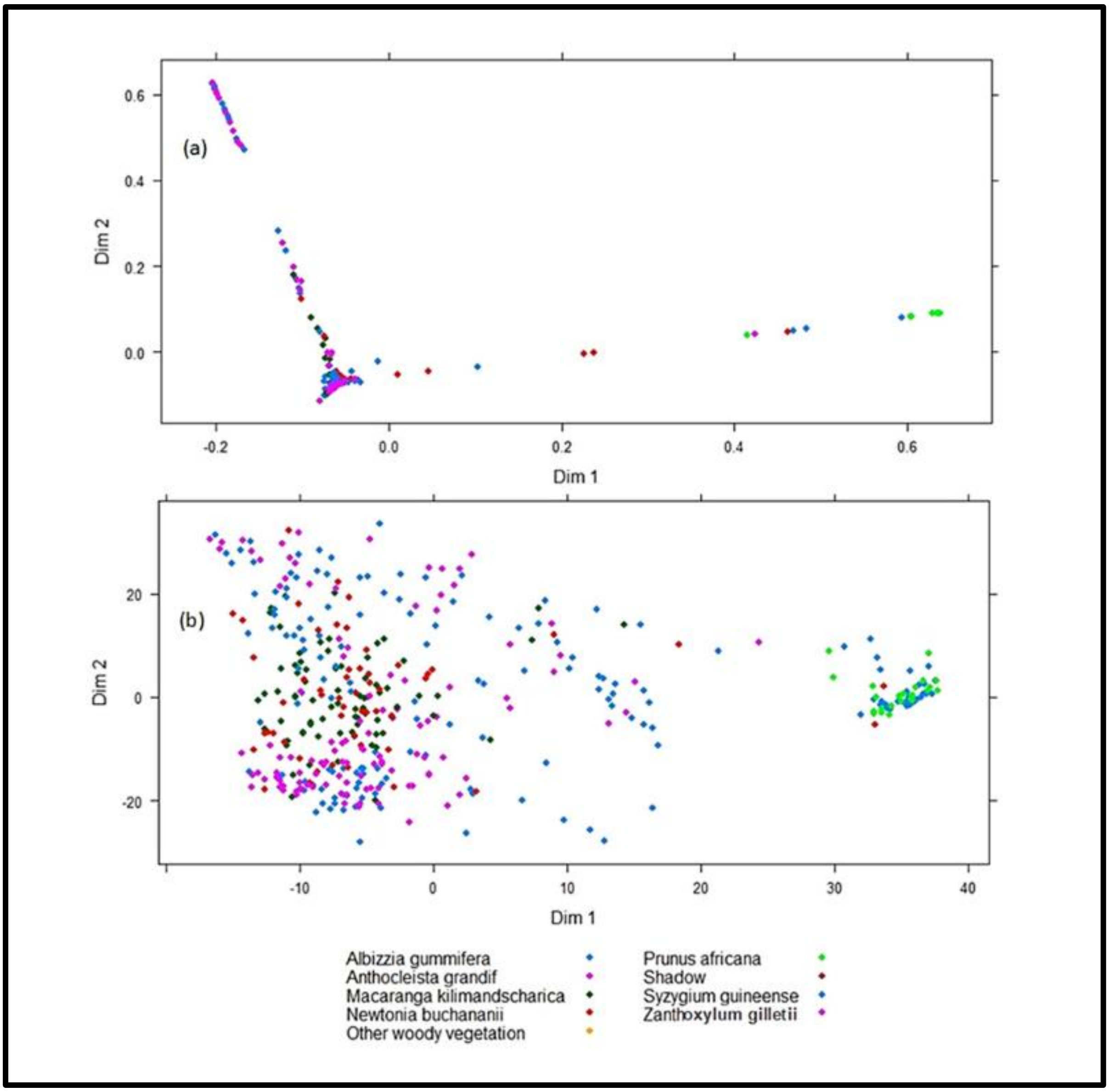


Figure 3.7. *Visualizing tree species (dis)similarity using multidimensional scaling: (a) random forest (b) support vector machine.*

The confusion matrices in Table 3.4 are for RF and SVM classifiers, for the combined technique, and F1-scores were closest to their model mean. The combined sampling technique dataset achieved the highest average OA for both RF and SVM models, i.e. 73.2 ± 2.5% and 70.9 ± 2.7%, respectively. The imbalanced dataset reported OA of 68.4 ± 3.6 for RF and 67.5 ± 3.9 for SVM. The oversampled dataset recorded OA score of 71.1 ± 2.8% for RF and 69.7 ± 4.1% for SVM. The undersampled data reported the lowest OA values—49.9 ± 5.7% and 53 ± 6.6% for RF and SVM, respectively. The user's accuracy (UA) values were in the range of 36-100% for RF and 33-100% for SVM. *Syzygium guineense* and *Zanthoxylum gilletii* achieved the highest UA values. *Albizzia gummifera*, *Anthocleista grandiflora*, *Newtonia buchananii*, and *Prunus Africana* produced low UA values across iterations. The producer's accuracy (PA) followed a similar pattern in all iterations. The PA scores were in the range of 32-95% and 23-100% for RF and SVM, respectively. *Syzygium guineense* and *Zanthoxylum gilletii* reported the highest PA scores, although the scores were relatively lower than the UA scores. *Albizzia gummifera*, *Anthocleista grandiflora*, *Newtonia buchananii*, and *Prunus Africana* recorded the lowest PA values in all models.

Table 3.4: *Confusion matrices for (a) random forest and (b) support vector machine algorithms, for the combined sampling technique. AlG–Albizzia gummifera, AnG–Anthocleista grandiflora, MK–Macaranga kilimandscharica. NB–Newtonia buchananii, OWV–other woody vegetation, PA–Prunus Africana, SD–Shadow, SG–Syzygium guineense, ZG–Zanthoxylum gilletii.*

(a) Random Forest

	AlG	AnG	MK	NB	OWV	PA	SD	SG	ZG	Total	UA (%)
AlG	4	1	1	3	0	0	1	0	0	10	40.0
AnG	0	9	4	1	2	2	0	0	0	18	50.0
MK	0	0	16	2	0	2	0	0	0	20	80.0
NB	3	4	2	9	1	0	0	0	1	20	45.0
OWV	0	0	0	2	17	0	0	0	1	20	85.0
PA	0	0	1	5	1	6	0	0	0	13	46.2
SD	0	0	0	0	0	0	23	0	0	23	100.0
SG	0	0	0	0	0	0	0	23	0	23	100.0
ZG	1	1	0	2	3	0	0	1	22	30	73.3
Total	8	15	24	24	24	10	24	24	24	177	
PA (%)	50.0	60.0	66.7	37.5	70.8	60.0	90.0	90.0	80.0		

Overall accuracy (OA) = 72.9%   F1-score = 68.0%

(b) Support Vector Machine algorithm

	AlG	AnG	MK	NB	OWV	PA	SD	SG	ZG	Total	UA (%)
AlG	3	1	1	2	1	0	0	0	0	8	37.5
AnG	1	8	3	2	2	2	1	0	1	20	40.0
MK	0	0	15	3	0	2	1	0	0	21	71.4
NB	3	4	2	7	1	0	0	0	1	18	38.9
OWV	0	0	0	2	17	0	0	0	0	19	89.5
PA	1	1	2	6	2	6	0	0	0	18	33.3
SD	0	0	0	0	0	0	22	0	0	22	100.0
SG	0	0	0	0	0	0	0	24	0	24	100.0
ZG	0	1	1	2	1	0	0	0	22	27	81.5
Total	8	15	24	24	24	10	24	24	24	177	
PA (%)	37.5	53.3	62.5	29.2	70.8	60.0	80.0	100.0	70.8		

Overall accuracy (OA) = 70.1%   F1-score = 64.2%

Syzygium guineense reported the highest F1-score of 99 ± 0.8% and 98 ± 0.7% for RF and SVM, respectively. *Zanthoxylum gilletii* achieved the F1-score of 78 ± 2.6% for RF and 76 ± 2.9% for SVM). Both classes were characterized by high accuracy and low variability, unlike *Albizzia gummifera*, *Anthocleista grandiflora,* and *Newtonia buchananii,* which had low variability across iterations and also low F1-score (Figure 3.8). The least accurately mapped species, Newtonia *buchananii*—recorded F1-score of 38 ± 7.1% for RF and 36 ± 7.9% for SVM. *Syzygium guineense recorded the highest UA values followed by Zanthoxylum gilletii.*

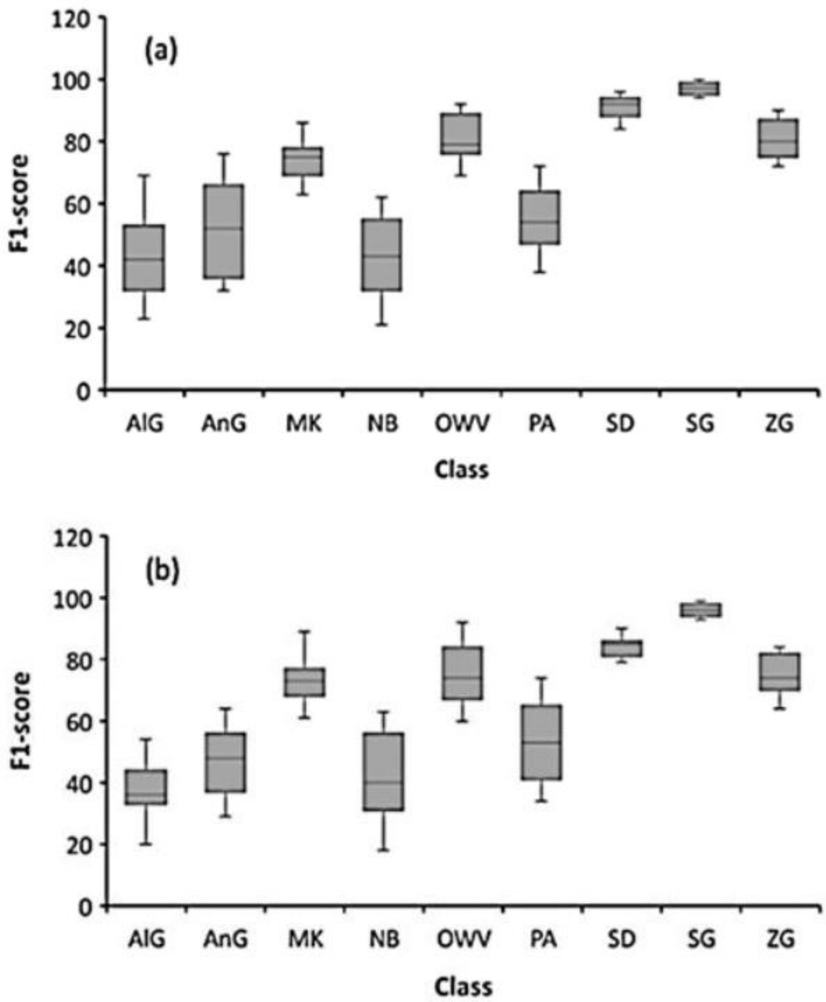


Figure 3.8. *Species-level F1-score for the combined sampling technique: (a) random forest; (b) support vector machine. AlG–Albizzia gummifera, AnG–Anthocleista grandiflora, MK–Macaranga kilimandscharica. NB–Newtonia buchananii, OWV–other woody vegetation, PA–Prunus Africana, SD–Shadow, SG–Syzygium guineense, ZG–Zanthoxylum gilletii.*

The combined sampling technique dataset produced average species F1-score of 68.56 ± 2.6% for RF and 64.64 ± 3.4% for SVM. Overall, the F1-score ranged between 18 and 100%. The undersampled data reported the lowest F1-score, i.e. 45.8 ± 5% for RF, compared to SVM's 48 ± 6%, unlike the other datasets where RF performed better than SVM. The original dataset's F1-score was 63.8 ± 3.9% for RF and 62.7 ± 4.3% for SVM. The oversampled dataset achieved F1-score values of 67.8 ± 4.1% and 63.6 ± 3.7% for RF and SVM, respectively (Figure 3.9).

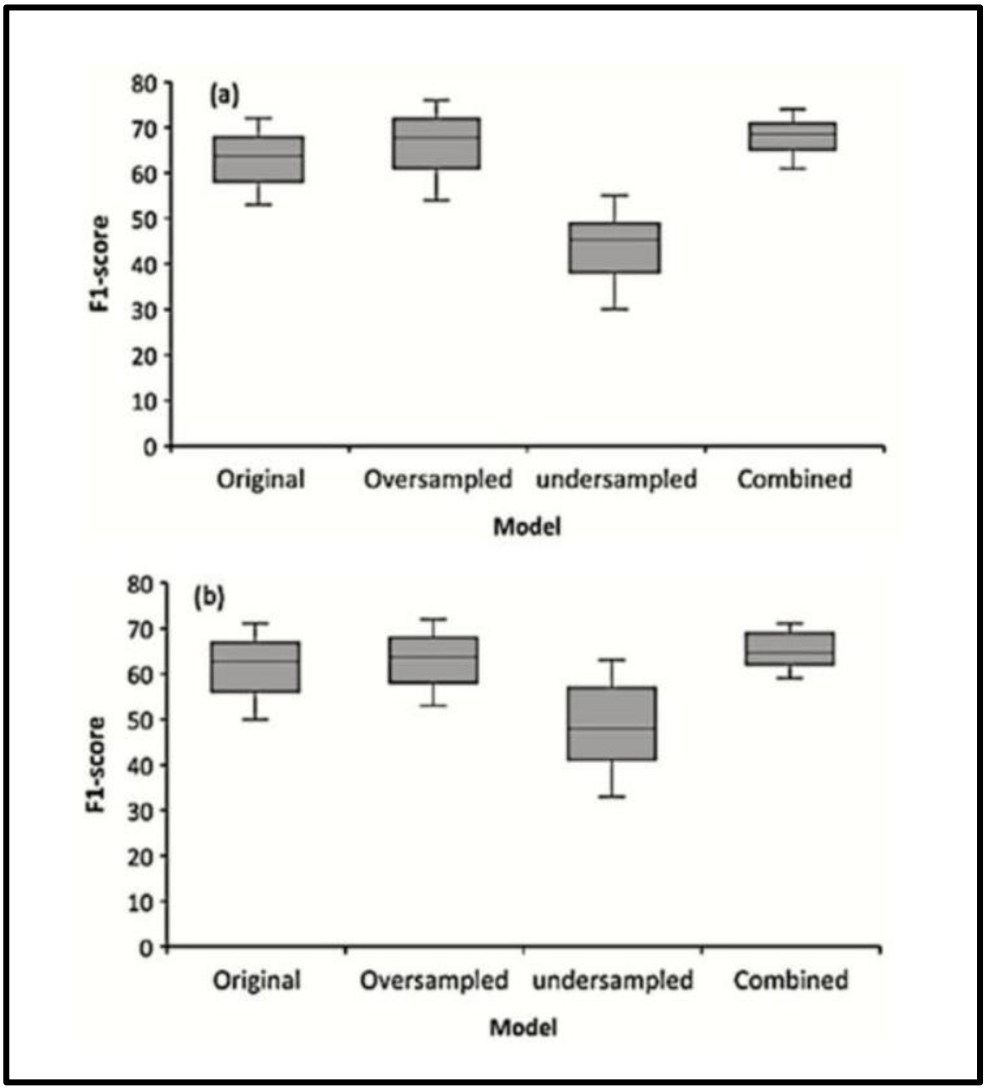


Figure 3.9. *Model-level F1-score for the datasets: (a) random forest; (b) support vector machine.*

The McNemar's test returned a Z value of 0.96 (Table 3.5). Thus, there were no significant differences ($Z \geq 1.96$) at a 95% confidence level, existing amongst the confusion matrices of the two classifiers.

Table 3.5: *McNemar's test results to compare random forest (RF) and support vector machine (SVM) classification models using the combined sampling technique.*

Classifier		RF		
		CC	IC	Total
SVM	CC	98	11	111
	IC	16	52	66
	Total	114	63	177

CC stands for correctly classified and IC is for incorrectly classified samples.

3.3.5 The spatial distribution of the endangered tree species

The distribution of the TS in submontane tropical forest reserve in Kenya was mapped using the RF and SVM classification algorithms (Figure 3.10). The maps were produced using the combined sampling technique dataset, whose F1-score was closest to their model mean. The maps showed that TS distribution is not haphazard. *Macaranga kilimandscharica* occupied the northwestern part of the study area. *Syzygium guineense* mostly predominantly occupied the lower and central parts of the study area. *Zanthoxylum gilletii* covered the central-eastern and western parts. A combination of other vegetation types dominated mostly the lower parts of the study area. The other TS, i.e. *Albizzia gummifera*, *Anthocleista grandiflora*, *Prunus africana*, and *Newtonia buchananii* were found throughout the study area.

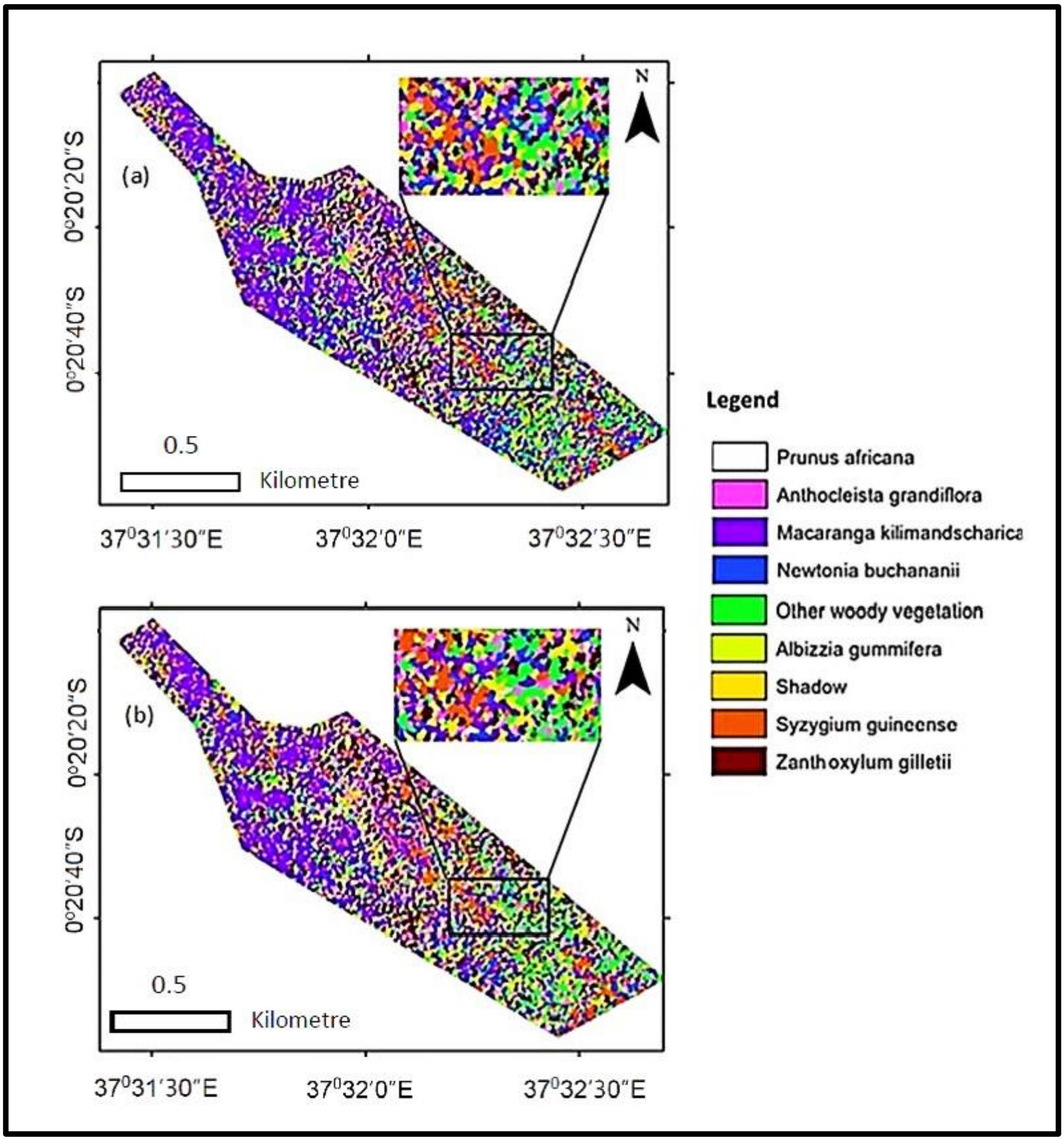


Figure 3.10. *Classification maps of tree species in Mt. Kenya Forest Reserve obtained using (a) random forest and (b) support vector machine.*

The RF and SVM classifiers showed differences in the size of each class. *Macaranga kilimandscharica* covered 45.2 ha and 44.1 ha for RF and SVM classifiers, respectively (Table 3.6). *Zanthoxylum gilletii* covered an area equivalent to 7.1 ha for RF and 7.2 ha for SVM. For *Syzygium guineense*, there was a 1% difference between the two classifiers: RF recorded 9.6 ha and SVM recorded 10.6 ha. *Newtonia buchananii* reported a difference of 0.4 ha between the two classifiers. *Anthocleista grandiflora* had the second-largest difference between the two classifiers, 1.8 ha, i.e. 13.4 ha for RF and 15.2 ha for SVM. Other woody vegetation showed a difference of 1.9 ha between RF and SVM classifiers. Shadow recorded 10.6 ha and 11.1 ha for the RF classifier and SVM classifier, respectively. The areal coverage of *Prunus Africana* and *Albizzia gummifera* was insignificant compared to that of the other classes.

Table 3.6: *Area coverage in hectares and percentage of the classes using random forest and support vector machine classifiers.*

Classes	RF Classified Area (ha)	Percentage (%)	SVM Classified Area (ha)	Percentage (%)
Macaranga kilimandscharica	45.2	34.8	44.1	33.9
Zanthoxylum gilletii	7.1	5.5	7.2	5.6
Syzygium guineense	9.6	7.4	10.6	8.2
Newtonia buchananii	23.6	18.1	23.2	17.8
Anthocleista grandiflora	13.4	10.3	15.2	11.7
Prunus africana	0.0	0.0	0.0	0.0
Albizzia gummifera	0.0	0.0	0.0	0.0
Other woody vegetation	20.5	15.8	18.6	14.3
Shadow	10.6	8.1	11.1	8.5
Total	130	100	130	100

3.4 Discussion

3.4.1 Spectral separability between the tree species

This study aimed to evaluate the impact of imbalanced data in mapping endangered TS in a selectively logged sub-montane heterogeneous tropical forest using RF and SVM classifiers, and WV-2 multispectral imagery. There were spectral overlaps within the TS classes (Figure 3.3). The overlaps could be explained by the very complex forest structure in the study site—most stands were multi-storied, and some tree crowns were relatively small, thus the presence of mixed pixels (Immitzer *et al*., 2012). This led to some misclassifications between species, which was in agreement with their inter-specific separability (Table 3.2). *Newtonia buchananii*, *Anthocleista grandiflora*, and *Albizzia gummifera* are examples of TS with significantly low F1-scores. The J-M distance values showed that the species were not as separable as *Syzygium guineense*, *Zanthoxylum gilletii*, and *Macaranga kilimandscharica*, which showed higher spectral separability values and were more separable, and therefore had higher F1-score values. Classes were separable when J-M distance values were greater than 1.9, but between 1.7 and

1.9 the separation was fairly good, while the separation was poor when the J-M distance values were below 1.7 (Jensen, 2005). In addition, very steep mountainous terrain and dense forest made it difficult to collect data in the field. This may have interfered with the quality of ground data used in the validation process. This research observed that higher spectral similarity within a class and higher spectral variability among classes led to higher classification accuracy. However, the contribution of the within- and among-species spectral variability to the classification accuracy of tropical tree species is poorly understood (Ferreira *et al*., 2016).

3.4.2 Relative importance of variables

In the classification process, the study was able to single out the most valuable bands, by making use of the variable importance measurement in the RF algorithm. RF uses the mean decrease in accuracy (MDA) and mean decrease in Gini (MDG) to evaluate the explanatory power of the input variables (Breiman, 2001). Generally, the best four bands in the classification process were the Red (band 5), Red Edge (band 6), Near Infrared 1 (band 7), and Near Infrared 2 (band 8). This shows that the VNIR portions of the electromagnetic spectrum are quite relevant in the mapping of endangered TS in the study area. The MDA was also used to show bands that were crucial in detecting particular TS (Figure 3.6). However, both MDA and MDG are biased in adapting to the variables in the tree structure, hence larger values are provided than the actual value. Strobl *et al*. (2008) pointed out that the two indicators are unable to determine the explanatory power of variables in the classification process because they cannot distinguish false correlations due to data characteristics. Strobl *et al*. (2008) developed a new, conditional permutation scheme to evaluate the influence of conditional variable classification to solve this problem, but the technique was inconsistent in grasping the explanatory power variables (Hur *et al*., 2017). Hur *et al*. (2017) developed another technique based on the Shapley Value method on RF regression. The determination of the relative importance of variables remains an active area of research.

3.4.3 Class imbalance

The effect of unequal training class sizes for species classification from WV-2 imagery was reported in Section 3.2.5. The original dataset was imbalanced despite efforts to add samples of rare species, such as *Prunus Africana* and *Albiz-zia gummifera*, whose density has diminished in MKFR. The variation in the sample size across classes in the original dataset increased the rate of classification errors (Maxwell *et al*., 2018). As expected, TS with larger sample sizes had positive prediction bias, i.e. commission errors were greater than omission errors, and *vice versa* (Chawla *et al*., 2004; Graves *et al*., 2016). In addition, species with small samples recorded high F1-score variability across iterations (Figure 3.8). The degree of bias in the model was minimized when the training data size was standardized across classes. The model that used the original dataset reported a reduced OA. Recent researches show that balanced data offers better overall classification performance (Chawla *et al*., 2004; Maxwell *et al*., 2018). The F1-score for the combined sampling technique model was 68.56 ± 2.6% and 64.64 ± 3.4%% for RF and SVM, respectively, whereas the unbalanced dataset reported F1-scores of 63.8 ± 3.9% for RF and 62.7 ± 4.3% for SVM. Although these results are consistent with other studies using similar classification algorithms, they may not be directly comparable because of factors such as differences in the topography of the study area, size of samples, species diversity in the study area, a continuous closed canopy, size of the study area, and spectral and spatial resolutions of the WV-2 image used.

The undersampling techniques reduce observation numbers from the majority class to make the dataset balanced. This technique may lead to the loss of important information in the training data. According to Graves *et al*. (2016), lowering the number of the sampled common TS while maintaining the species' full range spectral variability solves the reduced OA problem. Spectral variability in data can be retained when using SVM algorithms by selecting data points on the border between classes to delineate the separation between them (Foody and Mathur, 2004), or to iteratively prune the support vectors to attain the best separation between classes (Chen *et al*., 2005). The other method is the manipulation of the SVM algorithm by adjusting margin variable misclassification costs, e.g. the cost of misclassifying a feature in the minority class is set higher than that of misclassifying a feature in the majority class (Maxwell *et al*., 2018). However, the effects of these selection processes on bias in models and successive application to larger areas with many classes have not been quantified (Graves *et al*., 2016).

Other methods of reducing class imbalance such as the application of balanced class weight or the synthetic minority over-sampling technique (SMOTE) (Chawla *et al*., 2004) can be used. In this study, the idea was to combine the undersampling and oversampling techniques to create a robust balanced dataset fit for model training.

According to Krawczyk (2016), the imbalanced classification problem is not solved—at a time when we have such terms as big data, large neural network models, deep learning, and models such as the xgboost. Solutions should be identified and addressed specifically for each training dataset (Krawczyk, 2016).

3.4.4 Model Performance

In using an imbalanced dataset, the final classification under-predicts the classes with fewer samples, thus minority classes will have less effect on the accuracy compared to larger classes (He and Garcia, 2009). Therefore, test samples belonging to smaller classes are more often misclassified than those belonging to the dominant classes (Chawla *et al*., 2004; Maxwell *et al*., 2018). In such a case, a model may report a high accuracy level, but the map would not be useful. Accuracy is appropriate for balanced datasets but not for imbalanced ones (Graves *et al*., 2016). In most cases, test samples belonging to the small classes are misclassified more often than those belonging to the prevalent classes—the correct classification of samples in the small classes was equally important as the correct classification of samples in the prevalent classes (Chawla *et al*., 2004; Maxwell *et al*., 2018). The mapping of sparse TS, e.g. *Prunus Africana* and *Albizzia gummifera* was equally important.

The balancing of the original dataset slightly improved the F1-scores for RF and SVM classifiers. Omission errors were recorded for each species, but *Newtonia buchananii* had the highest omission errors in all models. *Newtonia buchananii* and *Anthocleista grandiflora* exhibited a higher level of spectral confusion. This further degraded the F1-scores in all models, because TS with low separability tend to have high misclassification rates. Furthermore, these TS exhibited different crown colours and densities (Adam *et al*., 2014). In addition, because random sampling was used to collect field data, tree crowns used to train and validate the classifications of TS are not distributed equally over the forest. The dense forest and rough mountainous terrain limited the places we could access within the forest. This may as well have influenced the outcome of the classification.

The F1-score was also skewed because it did not take into account the true negatives (TN). Again this depends on the high-level classification of positive and negative, thus it is as well relatively subjective. Researchers should try other metrics such as Matthew's Correlation Coefficient (introduced by biochemist Brian W. Matthews in 1975). The coefficient produces a more informative and truthful score than accuracy and F1-score—only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives) (Chicco and Jurman, 20202).

Finally, although error matrices are very crucial in comparing classification results, they only give an estimate of the accuracy of the classification, determined by the samples collected from the field. Thus, only biased conclusions can be made from such data (Foody and Mathur, 2004), therefore other metrics of model performance should be tried, e.g. balanced accuracy, and bias score, Matthew's Correlation Coefficient, among others (Graves *et al*., 2016).

3.4.5. The Spatial Distribution of Endangered Tree Species

This study found that *Macaranga kilimandscharica*, an invasive species, mostly occupied the north-western part of the study area. The TS is found in areas that have experienced intensive logging and forest disturbance. Logging activities took place in the area in the 1990s and early 2000s. The RF and SVM classified maps of the study area showed that *Newtonia buchananii* was more dominant than *Syzygium guineense*, *Zanthoxylum gilletii*, *Anthocleista grandiflora*, *Albizzia gummifera*, and *Prunus africana*. Both *Prunus africana* and *Albizzia gummifera* were hard to find in the study area. The same applies to *Ocotea usambarensis*—once dominant in the wet forests of Mount Kenya, the endangered TS was rare.

In cases where differences in accuracy are marginal, i.e. a few percentage points apart, Janssen and van der Wel (1994) proposed that a statistically rigorous way must be used to compare the accuracies, and the results should be expressed with confidence limits. In this study, the McNemar test was used to evaluate the RF and SVM classification outputs of the combined technique dataset. The McNemar test was meant to indicate whether a difference of 3.9% was statistically important. A difference in accuracy between the confusion matrices of different WorldView-2 spectral subsets is statistically significant ($p \leq 0.05$) if the Z value is more than 1.96 (Foody and Mathur, 2004). The Z value was 0.96, meaning there is no significant difference between the two maps. Therefore, either of the two maps can be considered for conservation purposes.

3.5 Conclusions

This study was aimed to identify and map TS under threat in a selectively logged sub-montane heterogeneous tropical forest using RF and SVM classifiers and WV-2 multispectral imagery. The study obtained average F1-scores of 68.56 ± 2.6% and 64.64 ± 3.4% for RF and SVM, respectively, for the best model, the combined sampling technique. This was an improvement from the original imbalanced dataset. The F1-scores reported were directly related to the differences between the spectral variability within and among species. The most important spectral bands identified to have played a major role in mapping the endangered TS in the study area were the Red (band 5), Red Edge (band 6), Near Infrared 1 (band 7), and Near Infrared 2 (band 8). The TS portrayed significant spectral overlaps, and this may have led to misclassification errors between classes. In addition, difficult mountainous terrain and dense forest made it hard to collect data in the field, and this may have interfered with the quality of data, and thus contributed to increased classification errors.

Given the results presented in this study, the approach used in this study may serve as the basis for forest recovery initiatives in MKFR, an ecosystem that is considered a biodiversity hot-spot for conservation priorities. However, these applications are based on models of species classification that are not perfect; therefore, more methods need to be developed to overcome the challenges caused by imbalanced data. Further research should target a larger study area and a higher number of TS using VHR satellite data and object-based image analysis (OBIA) techniques.

CHAPTER FOUR

A machine learning approach to map canopy gaps in tropical forests using WorldView-3 multispectral imagery

This chapter is based on:

Jackson, C.M., and Adam, E. (Accepted). A machine learning approach to map canopy gaps in an indigenous tropical submontane forest using WorldView-3 multispectral satellite imagery. *Environmental Conservation*, 49, 255–262.

Abstract

Selective logging (SL) in tropical forests may lead to deforestation and forest degradation, so accurate mapping of it will assist in forest restoration, among other ecological applications. This study was aimed to track canopy tree loss associated with illegal logging of *Ocotea usambarensis* (East African camphor) in a closed-canopy submontane tropical forest, by evaluating the potential of a very high resolution (VHR) WorldView-3 (WV-3) multispectral dataset using random forest (RF) and support vector machine (SVM) classifiers. A total of 55 spectral metrics were extracted from the WV-3 dataset, i.e. 23 means (of 15 vegetation indices–VIs and 8 visible-near-infrared–VNIR bands), and 23 standard deviations–SDs (of 15 VIs and 8 VNIR bands). Also extracted were 8 ratios (of 8 VNIR bands), and 1 brightness feature (average of the means of bands 1 to 8). The results showed average overall accuracies (OA) of 92.3 ± 2.6% and 94.0 ± 2.1%, for RF and SVM models, respectively. Average kappa coefficients (κ) were 0.88 ± 0.03 for RF and 0.90 ± 0.02 for SVM. The user's accuracy (UA) and producer's accuracy (PA) for both classifiers were in the range 84%–100%. This study further indicated that VIs derived from bands 5 and 6 helped detect canopy gaps in the study area. Both variable importance measurement in RF algorithm and pairwise feature selection proved useful in identifying the most pertinent variables in the classification of canopy gaps. These findings could allow forest managers to improve methods of detecting canopy gaps at larger scales, using remote sensing (RS) data and relatively little additional fieldwork.

Keywords: Tropical forests. WorldView-3. Canopy gaps. *Ocotea usambarensis*. Variable importance. Pairwise comparison. Morphological filters. Machine learning algorithm.

4.1 Introduction

Tropical rainforests cover *c*. 7% of the globe, and they perform vital environmental processes such as the water cycle, soil conservation, carbon sequestration, and habitat protection (Solberg *et al*., 2008). About three-quarters of Kenya's land mass is occupied by arid and semi-arid lands (ASALs), and *c*. 7% of Kenya is forested, which is below the UN recommended minimum of 10% (KWS, 2010). Kenya's rainforest cover, mostly montane forests, is fragmented into patches that are being degraded (NEMA, 2010). The Mt. Kenya Forest Reserve (MKFR) sustains a variety of biodiversity and affords vital ecosystem services, as well as being a major forested water catchment area in Kenya (NEMA, 2010). Intense growth of the human population around the MKFR and increasing poverty levels have led to forest degradation there due to illegal logging of important timber trees, especially *Ocotea usambarensis*, which is a hardwood tree with excellent decay and insect resistance (KFS, 2010). In 2000–2001, *Ooctea usambarensis* was the most highly priced timber in Kenya. The State of the Environment report by the National Environment Management Authority (NEMA) listed *Ooctea usambarensis* as endangered (NEMA, 2010). Normally, *Ocotea* trees mature in 60–70 years and are relatively large with spreading crowns and stem diameters in the range 3.75–9.5 m (Kigomo, 1987). The tree has low seed viability and poor regeneration, produces seeds only every 10 years and seed germination is sporadic (Kigomo, 1987). For decades now, the MKFR has undergone illegal logging for its commercially valuable reserves of indigenous timber, especially *Ocotea usambarensis* (KFS, 2010), resulting to canopy gaps in the forest landscape.

Tracking of selective logging (SL) in tropical forests is important due to its effects on biodiversity and other forest attributes, including ecosystem services, the microclimate and carbon pools (Dalagnol *et al*., 2019). Selective logging (SL) events can lead to deforestation, eventually contributing to the reduction of ecosystem services offered by tropical forests (Ota *et al*., 2019). Landscape-level spatial assessment of canopy gaps have primarily used ground-based methods, and due to the amount of effort and expense needed to acquire the measurements, the areas covered by these surveys are small and not contiguous spatially (Getzin *et al*., 2014). Remote sensing (RS) is a more cost-effective and less laborious method for modelling canopy gaps than ground-based methods (Vepakomma *et al*., 2008). However, SL can be challenging to detect and quantify because of its partial disturbance of the forest canopy and small scale of impact (Dalagnol *et al*., 2019). Three main types of very high resolution (VHR) earth observation (EO) data have been used to spot canopy gaps caused by SL in tropical forests—optical, LiDAR (light detection and ranging), and RADAR (radio detection and ranging) data (Zalazar, 2013). LiDAR technology has made the detection of small canopy gaps possible (Zalazar, 2013). For example, Asner *et al*. (2013) explored canopy height models (CHMs) from a single LiDAR data acquisition, while Andersen *et al*. (2014) used simple differencing of LiDAR CHMs to detect disappearing tree crowns with exceptional accuracy. Ellis et al. (2016) and Pearson *et al*. (2018) used LiDAR data from a single acquisition to detect SL by estimating aboveground biomass (AGB). Pinagé *et al*. (2019) and Rex *et al*. (2020) used LiDAR data acquired before and after logging to estimate the change in AGB due to logging. However, LiDAR covers relatively small spatial extents and data acquisition costs are high (Ellis *et al*., 2016; Silva *et al*., 2017). Therefore, researchers opted to integrate data from different RS systems; for example, Dalagnol *et al*. (2019) combined airborne LiDAR and VHR satellite data to quantitatively assess and validate canopy gaps. Automated mapping using time-series approaches applied to calibrated synthetic aperture radar (SAR) data have been successful in detecting SL (Baldauf and Köhl, 2009). Hethcoat *et al*. (2021) assessed the effectiveness of SAR data for monitoring tropical SL, but SAR-based biomass estimates have lower precision at the same resolution than optical data. Traditional aerial photography was successfully used for mapping canopy gaps before the introduction of high resolution optical data (Malahlela *et al*. 2014). However, aerial photography data acquisition is cost-intensive, although technological advances have revitalized the use of aerial photography through unmanned aerial vehicles (UAVs) (Otto *et al*., 2018). Spaias *et al*. (2016) used a hyperspectral camera on a UAV to detect and quantify small-scale canopy gaps in a tropical forest. Although the amount of spatial and spectral data gained made the data processing computationally demanding, especially where cloud-computing resources were lacking (Spaias *et al*., 2016). Ota *et al*. (2019) used digital aerial photographs (DAPs) acquired before and after logging to estimate the change in AGB linked to SL, while Kamarulzaman *et al*. (2022) used UAV data to detect forest canopy gaps attributed to SL. However, DAPs acquired using UAVs cover relatively small spatial extents (Dalagnol *et al*., 2019).

Machine learning (ML) for the classification of RS data has been applied in mapping SL with increasing success. Dalagnol *et al*. (2019) used a random forest (RF) model to detect tree loss with an average precision of 64%. Heathcoat *et al*. (2019) reported a detection rate of logged pixels of *c*. 90% using RF. Kamarulzaman *et al*. (2022) compared conventional and ML classifiers—the support vector machines (SVM) and artificial neural network (ANN) classifiers attained higher overall accuracy (OA) of 85%. Using ordinary least squares (OLS) regression and ML approaches (RF, *k*-nearest neighbour (*k*-NN), SVM, and ANN), Rex *et al*. (2020) monitored the change in AGB due to SL. Hethcoat *et al*. (2020, 2021) developed RF models and used logging records to detect SL. Therefore, the RF, SVM, and ANN approaches, which have superior image handling capabilities, have been mostly used in this area.

The development of VHR multispectral sensors such as WorldView-2 (WV-2) is critical for discriminating between tree canopies and vegetated gaps (Malahlela *et al.*, 2014), because some of the inherent features of hyperspectral data, such as carotenoids and chlorophyll sensitive bands, are preserved in WV-2/3 multispectral data (Mutanga *et al.*, 2012). Visual interpretation of VHR multi-date satellite imagery is a promising way to detect canopy gaps with fairly low uncertainty (Dalagnol *et al.*, 2019). Nonetheless, spatially accurate tree-scale validation data are not readily available, so automated approaches using VHR satellite data to accurately map canopy gaps over large and remote areas are not readily available (Dalagnol *et al.*, 2019). The primary focus of this study was to explore the potential of WorldView-3 (WV-3) imagery to develop a SL monitoring system capable of detecting canopy gaps over large spatial extents in a closed-canopy submontane tropical forest using random forest (RF) and support vector machine (SVM) models.

4.2 Materials and methods

4.2.1 Acquisition and pre-processing of satellite data

WorldView-3 (WV-3) data were acquired on 15 September 2019 for detecting canopy gaps in selectively logged sites. A WorldView-2 (WV-2) image acquired on 30 January 2014 and Google Earth were used for historical comparison. In order to cancel out the haze component caused by additive scattering from the RS data, the dark object subtraction method was applied (Chavez, 1988). The WV-2 satellite captures panchromatic images (450 to 800 nm) with a spatial resolution of 0.46 m and multispectral images with eight visible-near-infrared (VNIR) bands (400 to 1040 nm) at 1.84 m (DigitalGlobe, 2010). The WV-3 acquires panchromatic images with a spatial resolution of 0.30 m, multispectral imagery with eight VNIR bands at 1.2 m and eight shortwave-infrared (SWIR) bands (1195–2365 nm) at 3.7 m spatial resolution (DigitalGlobe, 2019). Additionally, there are 12 CAVIS (clouds, aerosol, vapour, ice, and snow) bands with a pixel size of 30 m (DigitalGlobe, 2019). Only WV-3 VNIR bands were used because SWIR bands covering the study area exhibited extensive cloud cover.

The satellite data was made available by Swift Geospatial, Pretoria, South Africa, geo-referenced to World Geodetic System (WGS) 1984 datum and the Universal Transverse Mercator (UTM) zone 37S projection. The WV-3 image was co-registered (i.e. aligned) with the WorldView-2 image to be able to match features between the two datasets—resulted in an average root mean square error (RMSE) of 3.41 m. It resulted in an average root mean square error (RMSE) of 3.41 m. The ENVI module (ENVI 5.3) FLAASH was used to atmospherically calibrate the images by converting the digital numbers (DN) to the top-of-atmosphere reflectance (Equation (4.1)):

$$p_{\lambda pixel,Band} = \frac{L_{\lambda Pixel,Band} \cdot {d_{ES}}^{2}.\pi}{Esun_{\lambda Band} \cdot Cos(\theta_s)} \tag{4.1}$$

where $L_{\lambda Pixel,Band}$ is the is the top-of-atmosphere band-averaged spectral radiance, d_{ES} is the Earth-Sun distance (AU) during the image acquisition, $\mathrm{Esun}_{\lambda\mathrm{Band}}$ is the band-averaged solar spectral irradiance ($Wm^{-2}\mu m^{-1}$) normal to the surface being illuminated, and θ_s is the solar zenith angle in degrees during the image acquisition (Omar, 2010). The Worldview-3 data was pansharpened to obtain a pixel size of 0.3 m. The satellite data were pansharpened to obtain new bands with a spectral resolution of the multispectral bands and a spatial resolution of the panchromatic band (Pu and Landry, 2012).

4.2.2 Acquisition of field data

Ground truth points were collected in February 2020 using a handheld Global Positioning System (Garmin eTrex® 20 GPS Receiver) and a pansharpened WV-3 image (pixel = 0.3 m). Examples of canopy gaps creating from logging of *Ocotea* trees are shown in Figure 4.1—(a) two uncut *Ocotea* trees in 2014 WV-2 image, pixel size = 1.89 m, (b) 2019 WV-3 multispectral image, pixel size = 1.20 m showing a fresh and an older canopy gaps created after cutting the two *Ocotea* trees, and (c) the two canopy gaps in pansharpened 2019 WV-3 image, pixel size = 0.3 m. In the WV-3 image, the canopy gaps were either partially/fully illuminated or not illuminated at all. Since the human-made canopy gaps shared similar reflectance characteristics with natural canopy gaps, GPS coordinates of 301 vegetated gaps and 301 shaded gaps were collected. A vegetated gap is a forest canopy gap with low vegetation inside it, which is the initial stage of vegetation recovery from forest disturbance. Coordinates of the approximate locations of gap centres were recorded and then overlaid on the pansharpened WV-3 image. Using a geographic information system (GIS—ArcGIS® v. 10.3; ESRI, Redlands, CA, USA), points were set on the vegetated and shaded canopy gap pixels on the WV-3 imagery, and by following the edges of the pixels, the points were made into polygons. Locations of closed forest canopy could be identified using the WV-3 image—thus, a total of 301 polygons were extracted. Canopy gaps formed after logging of *Ocotea* trees had their dimensions collected in the field, such as their dripline measurements, maximum length, and compass orientation, together with the maximum width perpendicular to the length. Using the GIS, a map of canopy gaps was generated from ground survey data. The accuracy of canopy gap delineation using RS was assessed by comparing the dimensions collected from the field.

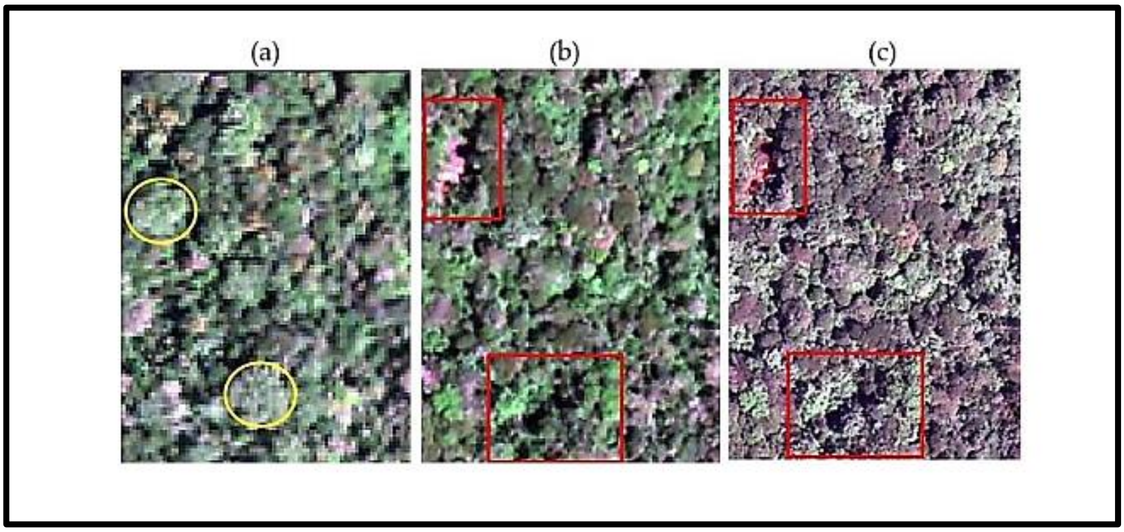


Figure 4.1. *A true-color image showing (a) two uncut Ocotea trees in 2014 WorldView-2 image (1.89 m); (b) fresh and older canopy gaps created after cutting the camphor trees in 2019 WorldView-3 multispectral image (1.20 m); (c) the two gaps in 2019 pansharpened WorldView-3 image (0.3 m).*

The reference data were randomly split into train and test datasets of 70% and 30%, respectively (Table 4.1).

Table 4.1: *Composition of train and test data for vegetated and shaded gaps, and forest canopy for the study area.*

Landscape feature	Code	Train data	Test data	Total
Vegetated gaps	VG	211	90	301
Shaded gaps	SG	211	90	301
Forest canopy	FC	211	90	301

4.2.3 Feature extraction and selection

Suitable techniques are key to extracting variables for modelling canopy gaps. In minimizing the effects of data saturation when mapping canopy gaps in dense forests, methods such as image transform algorithms, vegetation indices, spectral mixture analysis (SMA), and texture measures were utilized (Malahlela *et al*. 2014). Vegetation indices (VIs) derived from RS-based canopies can enhance the spectral information of RS data—VIs are quite straightforward simple and efficient algorithms for, among other applications, quantitative and qualitative evaluations of vegetation cover, vigour, and growth dynamics (Malahlela *et al*. 2014). Table 4.2 presents fifty-five (55) features/variables (23 means, 23 standard deviations—SD, 8 ratios, and 1 brightness feature) extracted from the pansharpened WV-3 image. The inclusion of the means and standard deviations of the vegetation indices was based on previous studies, such as Dalagnol *et al*. (2019) because these features could detect very small canopy gaps. Vegetation indices were chosen from the ones sensitive to greenness and plant senescence (Malahlela *et al*. 2014). To accurately extract shaded gaps and improve classification results, the shadow detection index (SDI) of Shahi *et al*. (2014) was used in the modelling. Shade is associated with canopy gaps as nearby trees appear on the edges of gaps (Spaias *et al*., 2016; Dalagnol *et al*., 2019). In reducing the redundancy and intercorrelation among the list of potential features, a subset of the best-performing features was extracted from the initial 55 features before classification. The features were used to discriminate between shaded gaps, vegetated gaps, and forest canopy. Approximate optimal thresholds were determined from the reference data using histograms and then adjusted to determine thresholds that resulted in the highest accuracy.

Table 4.2: *The features used to detect canopy gaps in Mt. Kenya Forest Reserve: 23 means (of the 15 vegetation indices and 8 visible–near-infrared bands), 23 standard deviations (of the 15 vegetation indices and 8 visible–near-infrared bands), 8 ratios (of the 8 visible–near-infrared bands) and 1 brightness feature.*

Feature	Equation[a]/Description	Reference
Bright	Brightness, an average of means of bands 1–8	Pu and Landry *et al*., 2012
Band 1–8	Means of pansharpened WorldView-3 bands 1–8	Pu and Landry *et al*., 2012; Dalagnol *et al*., 2019
SDB1–8	Standard deviations of WorldView-3 bands 1–8	Pu and Landry *et al*., 2012; Dalagnol *et al*., 2019
Ratio1–8	*i*th band mean divided by sum of band 1–8 means	Pu and Landry *et al*., 2012
Red edge position index (REPI)	(Max first derivative: 680–750 nm)	Horler *et al*., 1983
Chlorophyll absorption reflectance index (CARI)	$[(\rho700 - \rho672) - 0.2 * (\rho700 - \rho\rho553)]$	Kim, 1994
Structurally insensitive pigment index (SIPI)	$(\rho802 - \rho465)/(\rho802 + \rho681)$	Peñuelas *et al*., 1995
Normalized pigment chlorophyll index (NPCI)	$(\rho660 - \rho425)/(\rho660 + \rho425)$	Peñuelas *et al*., 1995
Red edge normalized difference vegetation index (RENDVI)	$(\rho753 - \rho700)/(\rho753 + \rho700)$	Gitelson *et al*., 1996
Photochemical reflectance index (PRI)	$(\rho534 - \rho572)/(\rho534 + \rho572)$	Gamon et al., 1997
Pigment-sensitive normalized difference (PSND)	$(\rho802 - \rho676)/(\rho800 + \rho676)$	Blackburn, 1998
Plant senescence reflectance index (PSRI)	$(\rho681 - \rho502)/\rho753$	Merzlyak *et al*., 1999
Simple ratio red/green (SRred/green)	RED/GREEN (i.e. mean reflectance of red bands (600–699 nm) and green bands (500–599 nm), respectively)	Gamon and Surfus, 1999
Modified CARI (MCARI)	$[(\rho700 - \rho672) - 0.2 * (\rho700 - \rho553)] * (\rho700/\rho672)$	Daughtry *et al*., 2000
Anthocyanin reflectance index (ARI)	$(1/\rho553) - (1/\rho700)$	Gitelson *et al*., 2001
Visible atmospherically resistant index (VARI)	$(\rho557 - \rho643)/(\rho557 + \rho643 - \rho465)$	Gitelson *et al*., 2002a; Gitelson *et al*., 2002b
Carotenoid reflectance index 1 (CRI-1)	$(1/\rho511) - (1/\rho553)$	Gitelson *et al*., 2002b
Carotenoid reflectance index 2 (CRI-2)	$(1/\rho511) - (1/\rho700)$	Gitelson *et al*., 2002b
Shadow detection index (SDI)	$[(\rho865 - \rho480)/(\rho865 + \rho480)] - (\rho835)$	Shahi *et al*., 2014

[a] ρ is the closest ProSpecTIR-VS band (the index number indicates the wavelength center in nanometers) to the original index formulation.

4.2.4 Spectral separability

Spectral separability measures the distance between two signatures, and separability between any combinations of variables can be used in the classification (Jensen, 2005). Only the subset from the best performing features was used in the analysis. The digitized vegetated gaps, shaded gaps, and forest canopy samples (Figure 4.2) were assigned spectral information, using the mean pixel values within their polygons, then used to generate respective signature files.

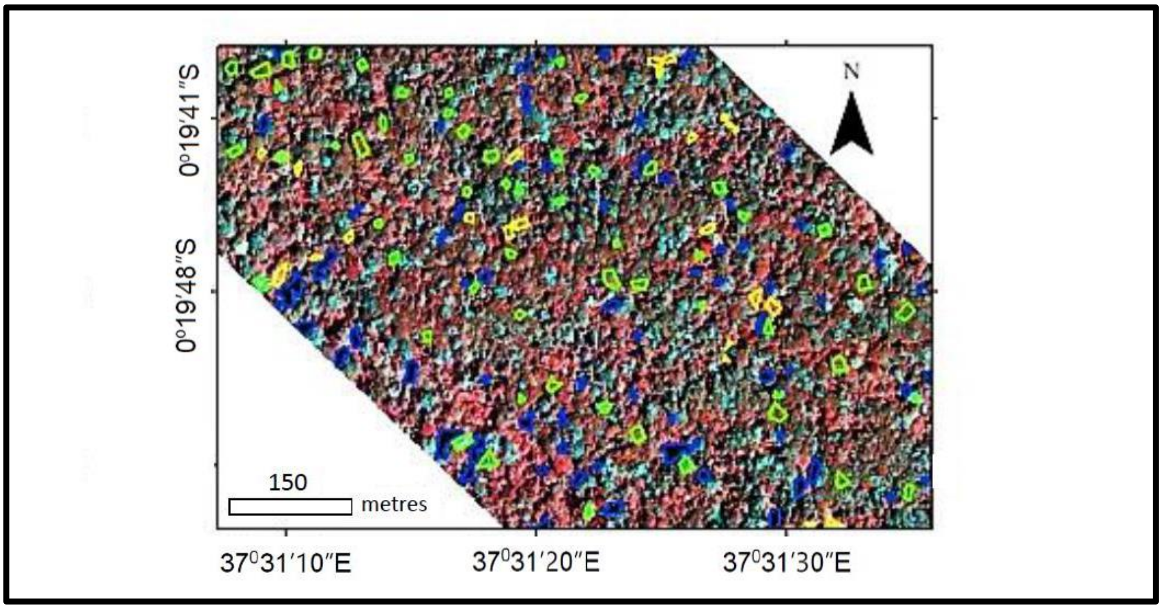


Figure 4.2. *False-color image (752-RGB) for the study area. Yellow, blue, and green polygons represent the reference vegetated gaps, shaded gaps, and forest canopy, respectively.*

The transformed divergence (TD) index and Jeffries-Matusita (J-M) distance separability measures were used. Divergence (D) is calculated from the mean and variance-covariance matrices of data of the feature classes (Kavzoglu and Mather, 2000) (Equation (4.2)):

$$D_{ij} = \frac{1}{2} tr\left[(\Sigma_i - \Sigma_j)(\Sigma_j^{-1} - \Sigma_i^{-1})\right] + \frac{1}{2} tr\left[(\Sigma_i^{-1} + \Sigma_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T\right] \qquad (4.2)$$

The TD (Equation (4.3)) is introduced to reduce the impact of well-separated classes that may raise the average divergence value and make the divergence measure misleading (Kavzoglu and Mather, 2000):

$$TD_{ij} = c\left[1 - e^{\frac{-D_{ij}}{8}}\right] \qquad (4.3)$$

where $tr[\cdot]$ is the trace of a matrix, which is the total of the matrix's diagonal elements, and Σ_i and Σ_j represent the variance-covariance matrices of classes i and j; μ_i and μ_j are the corresponding mean vectors; c is a value defining the range of TD values.

The J–M distance between distributions of two classes ω_i and ω_j has been defined as follows (Richards and Jia, 1999) (Equation (4.4)):

$$J\text{-}M_{ij} = 2(1 - e^{-B_{ij}}), \qquad (4.4)$$

where B_{ij} is the Bhattacharyya distance (Kailath, 1967) computed as (Equation (4.5)):

$$B_{ij} = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} ln\left(\frac{1}{2}\frac{|\Sigma_i + \Sigma_j|}{\sqrt{|\Sigma_i||\Sigma_j|}}\right) \qquad (4.5)$$

where μ_i and μ_j are the mean reflectances of species i and j, and Σ_i and Σ_j correspond to their covariance matrices, while $|\Sigma_i|$ and $|\Sigma_j|$ are the determinants of Σ_i and Σ_j, respectively. ln is the natural logarithm function, and T is the transposition function. The J-M distance improves the Bhattacharya distance by normalizing it to between 0 and 2.

4.2.5 Pairwise feature comparison

A correlation matrix was computed from the reference samples (Table 4.1). Similarity scores were calculated between each pair of features, to determine whether or not two features/variables share the same characteristics (Li and Wang, 2013). If the similarity score is more than a predefined threshold, then the two features are co-referent. Pairwise comparisons are in form of a matrix: $\boldsymbol{C} = [c_{kp}]_{n \times n}$; where c_{kp} is the pairwise comparison rating for *k*th and *p*th criteria. (Malczewki, 2016).

4.2.6 Morphological filtering

In the post-processing stage, it was necessary to transform the classification map so that it has only two classes, i.e. canopy gaps and forest canopy. Therefore, shaded and vegetated gap classes were merged into one class—canopy gaps. ArcGIS tools were used to implement morphological filters, i.e. the Shrink tool was used to remove narrow passageways and tiny gaps/holes in tree crowns. The classified raster gaps were enlarged with the Expand tool, and the remaining gaps were resized before being shrunk. Pixelation was reduced and any remaining trees in the gaps were removed using the Focal Statistics tool. The most frequent value was retained by employing the Major filter. Less than 100 m^2 in size gap polygons were eliminated using the Select function.

4.2.7 Image classification

The results in this study were obtained by training random forest (RF) and support vector machine (SVM) classifiers in R software—to replicate a target set of canopy gap data for the same study site that was obtained from ground survey. Using univariate splits at internal nodes, decision trees in RF recursively divide the source set into subsets with bagged samples (Maxwell *et al*., 2018). Before running the RF model, the number of decision trees (*ntree*) and the number of predictor variables (*mtry*) randomly selected at each node need to be defined. RF classification aggregates predictions from all DTs, then the majority vote of all trees assigns a final class for unknown features (Maxwell *et al*., 2018). Two metrics in RF—mean decrease in accuracy (MDA) and mean decrease in Gini (MDG)—were used to evaluate the explanatory power of the input variables (Han *et al*., 2016), in detecting vegetated gaps, shaded gaps, and tree crowns. The MDA coefficient shows how much accuracy is lost when each variable is removed from the model (Han *et al*., 2016). The MDG quantifies the contribution of variables to the homogeneity of the nodes and leaves in the RF (Han *et al*., 2016). The significance of variables in the model increases with the MDA and MDG values (Han *et al*., 2016).

After the training stage, the SVM selects one of the two possible labels to identify a class (Vapnik, 2000). SVM finds the position of the optimal separating hyperplane, which separates the original data while maximizing the margin between classes and minimizing the misclassification error (Vapnik, 2000). Any two distinguishable classes that have k samples symbolized by $(x_1, y_1),\ldots, (x_k, y_k)$, where $x \in R^n$ is an n-dimensional space, and y, a class label whose values are +1 or −1, then SVM will look for a desirable hyperplane designated by w = $(w_1,..., w_n)$ and b (Huang *et al*., 2008) (Equation (4.6)):

$$y_i = [w \cdot x_i + b] \geq 1 i \ = \ 1, \ldots, k \qquad (4.6)$$

The hyperplane's location is made possible by reducing the norm of w, or the following function under the above inequality function (Huang *et al*., 2008) (Equation (4.7)):

$$F(w) = 0.5(w \cdot w) \qquad (4.7)$$

Using kernel functions, SVM applies decision boundaries that are not linear, and a cost parameter C and gamma parameter γ are introduced to determine the cost of incorrect classification, and to give curvature weight of the decision boundary, respectively (Huang *et al*., 2008). The robust radial basis function (RBF) was selected since it requires fewer parameter values to be predefined. A parameter search must be done to select the best C and γ for a certain classification problem (Huang *et al*., 2008). Therefore, the γ parameter needs to be predefined (Huang *et al*., 2008) (Equation (4.8)):

$$K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2} \quad (4.8)$$

The cost parameter *C* also needs to be predetermined for the canopy gap mapping problem. A cross-validation quantitative analysis of pairs of values for the *C* and γ parameters was carried out. The set of parameters with the least amount of error was chosen to train the algorithm. Both RF and SVM classifiers were trained using 70% of the ground reference data, and for robust classification results, the 10-fold cross-validation method was repeated 10 times.

4.2.8 Measures of model performance

Accuracy assessment is used to determine whether pixels identified in the field are classified as they should (Adam *et al*., 2014). The performance of the two machine learning classifiers was evaluated using 30% of the ground truth data. Confusion matrices were computed and averaged over 10 repetitions. The overall accuracy (OA) is computed by dividing the number of correctly classified pixels by the total pixel number—typically expressed as a percentage (Rodriguez-Galiano *et al*., 2012). The producer's accuracy (PA) is the proportion of particular classes on the ground, referred to as such by the classification map (Rodriguez-Galiano *et al*., 2012). The user's accuracy displays the likelihood that a pixel labeled as a particular feature will be placed on the classified map as such (Mutanga *et al*., 2012). The kappa coefficient (κ) is the difference between the accuracy that was observed and the expected. The statistical significance between the two classifiers was compared using the McNemar test (Foody and Mathur, 2004) (Equation (4.9)):

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (4.9)$$

where f_{12} is the total number of samples classified correctly by the first classification, but misclassified by the second, and f_{21} is the total samples classified correctly by the second classification but misclassified by the first one. The confusion matrices of the various models are compared for accuracy, and a difference in accuracy is deemed statistically significant ($p \leq 0.05$) if $Z \geq 1.96$ (Foody and Mathur, 2004).

4.3. Results

4.3.1. Explanatory Power of the features extracted from WorldView-3 VNIR Bands

The most important features in all models, i.e. variables with the highest values of MDA were the brightness feature, the means of the pansharpened WorldView-3 (WV-3) visible-near-infrared (VNIR) bands, and those of the Chlorophyll Absorption Ratio Index (CARI), and the Modified Chlorophyll Absorption Ratio Index (MCARI). Others were the means of the Carotenoid reflectance index 2 (CRI-2), the normalized pigment chlorophyll index (NPCI), the plant senescence reflectance index (PSRI), the Red edge position index (REPI), and the Shadow detection index (SDI) (Figure 4.3a). The same features/variables were identified by the MDG (Figure 4.3b). The MDA and MDG values of all standard deviations (SD) of the pansharpened WorldView-3 VNIR bands, the vegetation indices except the anthocyanin reflectance index (ARI), the shadow detection index, and ratios of the pansharpened WorldView-3 VNIR bands were low.

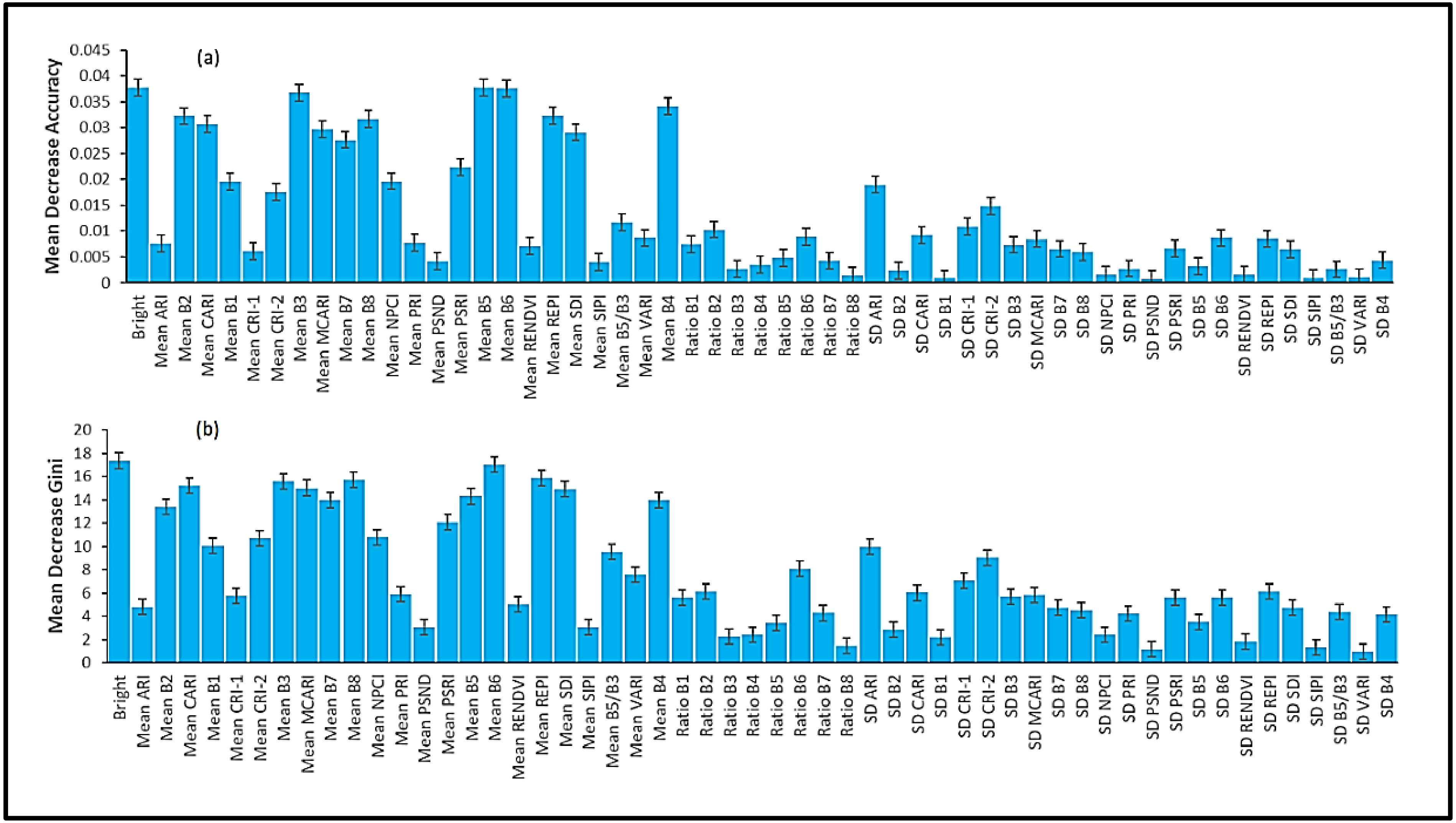


Figure 4.3. *The relative importance of metrics derived from pansharpened Worldview-3 bands in discriminating newly created, shadowed, and vegetated gaps, and forest canopy as measured by random forest classifier using (a) mean decrease analysis; and (b) mean decrease in Gini.*

4.3.2. Optimization of Random Forest and Support Vector Machine

The RF model whose iteration was closest to the model mean produced default *mtry* and *ntree* values of 14 and 500, respectively, with an OOB error rate of 0.074 (Figure 4.4a). The SVM model produced 0.1 and 10 for gamma and cost, respectively, yielding a cross-validation error of 0.060 (Figure 4.4b).

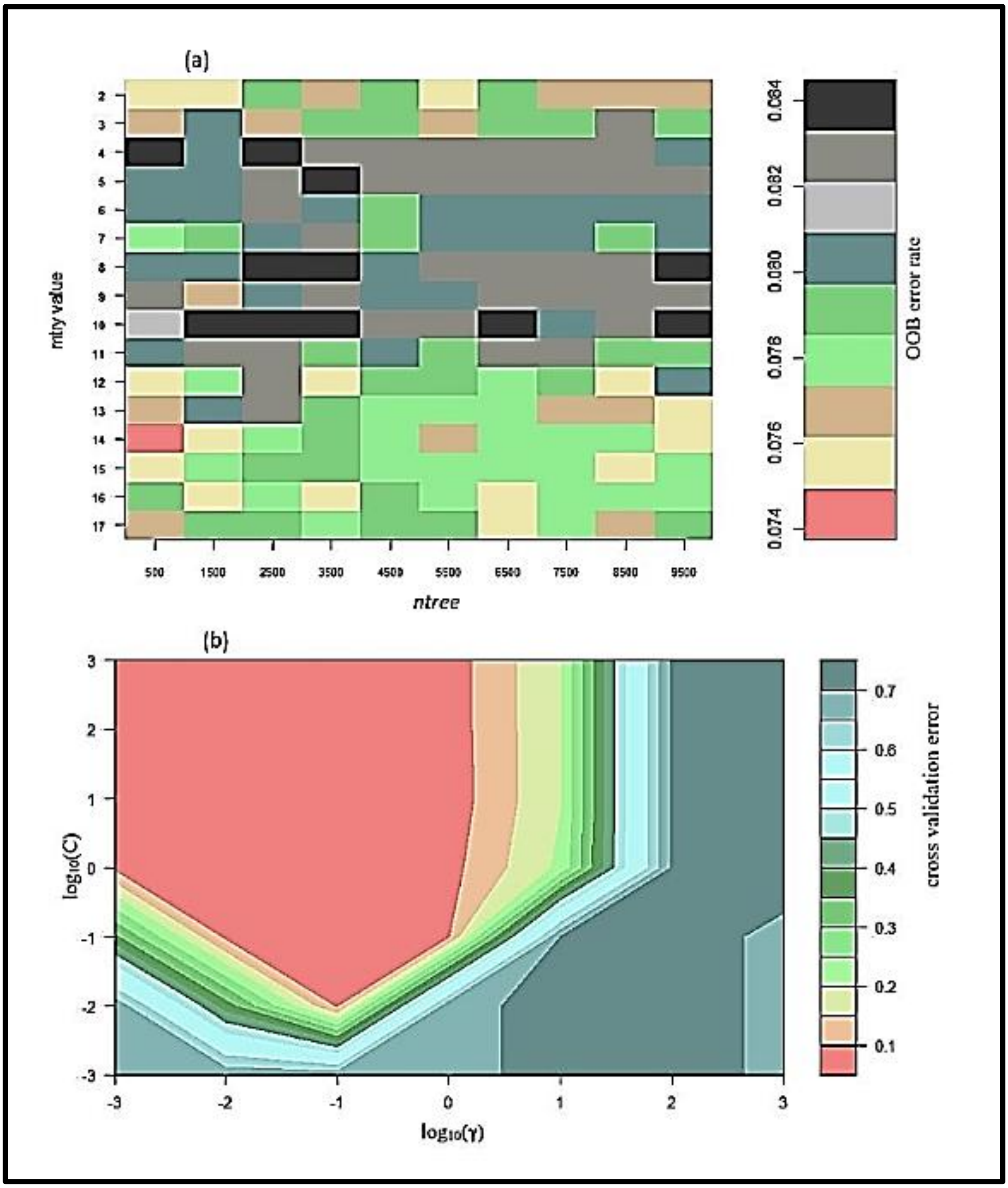


Figure 4.4. *Optimization by 10-fold cross-validation: (a) random forest parameters (mtry and ntree), and (b) support vector machine parameters (gamma and cost).*

4.3.3 Spectral separability

The mean spectral reflectance curves of the training data were extracted from pixels of the 17 best-performing variables and plotted with their standard deviations, as shown in Figure 4.5. Some of the variables, such as the means of the five visible spectrum bands, NPCI, PSRI, and SD of ARI exhibit considerable spectral overlaps by the three classes. Only the means of bands 7 (Near Infrared 1) and 8 (Near Infrared 2), and the SDI helped separate the three classes beyond 1 standard deviation of uncertainty. The brightness feature and means of band 6 (Red Edge), the CARI, and the MCARI showed considerable overlaps between the forest canopy and vegetated gaps.

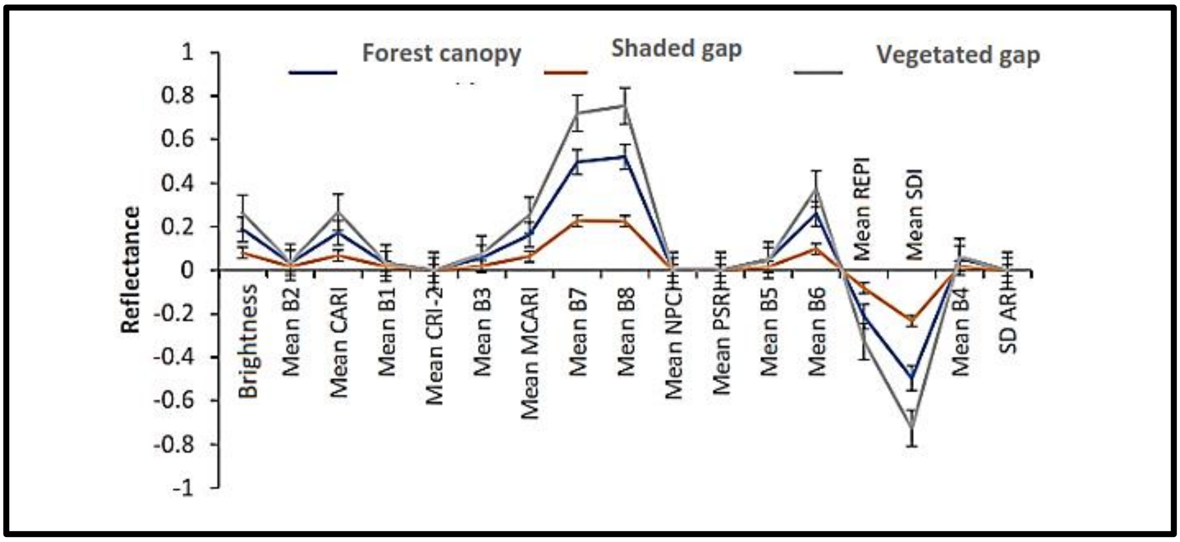


Figure 4.5. *Mean (spectral reflectance curves) and standard deviations (short lines) of shaded gaps, vegetated gaps, and forest canopy extracted from the best performing variables of WorldView-3 imagery pixels.*

According to the results of spectral separability (Table 4.3), the transformed divergence (TD) index recorded 1.91, 1.48, and 2.00 as average, minimum, and maximum values, respectively, for the means of the WorldView-3 VNIR bands. The Jeffries-Matusita (J-M) distance reported 1.90, 1.63 and 1.99, respectively. The average, minimum and maximum values for means of the VIs were 2.00, 2.00, and 2.00, respectively, for the TD index, while J-M distance reported 1.98, 1.92, and 2.00, respectively.

Table 4.3: *Spectral separability as calculated by: (a) transformed divergence index (Equation (4.3)), and (b) Jeffries-Matusita distance (Equation (4.4)).*

(a)

Means of vegetation indices	Means of pansharpened WorldView-3 VNIR bands				
		NG	SG	VG	FC
	NG		2.00	2.00	2.00
	SG	2.00		1.99	1.98
	VG	2.00	2.00		1.48
	FC	2.00	2.00	2.00	

(b)

Means of vegetation indices	Means of pansharpened WorldView-3 VNIR bands				
		NG	SG	VG	FC
	NG		1.99	1.96	1.90
	SG	2.00		1.98	1.94
	VG	2.00	1.99		1.63
	FC	1.92	1.99	1.95	

The mean of band 4 (Yellow) played a crucial role in the identification of the three classes compared to the other variables (Figure 4.6). The mean of band 5 (Red) was crucial in detecting vegetated gaps, as also means of the CARI, the CRI-2, the MCARI, the PSRI, the REPI, and bands 1 (Coastal), 3 (Green), and 6 (Red Edge). The mean of band 5 (Red) also helped detect shaded gaps. The means of brightness, bands 2 (Blue) and 7 (Near Infrared 1), the SDI, the NPCI, and the SD of ARI helped to detect the forest canopy.

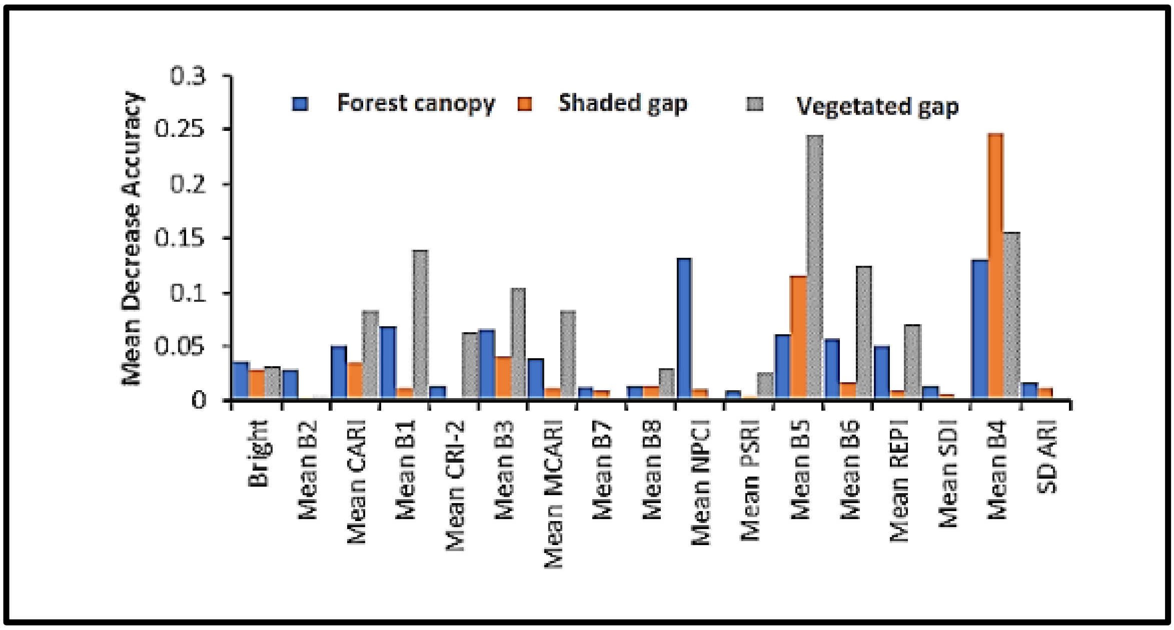


Figure 4.6. *The mean decrease accuracy values showed the relationship between each forest landscape feature and the metrics extracted from pansharpened WorldView-3 bands as measured by the random forest classifier.*

4.3.4 Pairwise feature comparison

Generally, the means of the WorldView-3 VNIR bands are partly highly correlated, providing redundant information (Table 4.4). The same applied to the means of the vegetation indices, such as the NPCI, the CARI, the MCARI, the CRI-2, and the PSRI. The highest negative correlations were reported between means of the REPI and the CRI-2, the MCARI, the NPCI, and the PSRI. High negative correlations were also recorded between the mean of SDI and those of the CARI and the MCARI. The lowest positive correlation, 0.001, was between the MCARI and bands 1 (Coastal) and 2 (Blue). The means of the SDI and the MCARI were perfectly positively correlated, while the means of the REPI and the CARI were perfectly negatively correlated. A high correlation was found between the standard deviation (SD) of ARI and the means of the MCARI and the SDI.

Table 4.4: *Pairwise correlations between the best-performing variables extracted from the WorldView-3 visible–near-infrared bands (B1-8), computed from the reference samples. CARI - chlorophyll absorption ratio index, MCARI - modified chlorophyll absorption ratio index, CRI 1 - carotenoid reflectance index 1, CRI 2 - carotenoid reflectance index 2, NPCI - normalized pigment chlorophyll index, PSRI - plant senescence reflectance index, REPI - red edge position index, and the Shadow detection index (SDI), and ARI - anthocyanin reflectance index.*

	Bright - ness	Mean B2	Mean CARI	Mean B1	Mean CRI-2	Mean B3	Mean MCARI	Mean B7	Mean B8	Mean NPCI	Mean PSRI	Mean B5	Mean B6	Mean REPI	Mean SDI	Mean B4	SD ARI
Brightness	-----	0.611	0.332	0.531	0.355	0.839	0.004	0.985	0.988	0.320	0.203	0.599	0.992	-0.330	-0.002	0.724	-0.403
Mean B2		-----	0.162	0.923	0.298	0.886	0.001	0.485	0.505	0.352	0.459	0.966	0.611	-0.163	-0.002	0.943	-0.437
Mean CARI			-----	0.137	0.717	0.262	0.942	0.333	0.331	0.875	0.747	0.158	0.334	-1.000	0.940	0.213	0.564
Mean B1				-----	0.290	0.810	0.001	0.405	0.426	0.227	0.446	0.900	0.532	-0.138	-0.002	0.864	-0.411
Mean CRI-2					-----	0.378	0.649	0.321	0.327	0.852	0.872	0.400	0.361	-0.718	0.647	0.393	0.110
Mean B3						-----	0.003	0.745	0.755	0.378	0.396	0.884	0.857	-0.260	-0.002	0.958	-0.460
Mean MCARI							-----	0.004	0.004	0.832	0.752	0.002	0.005	-0.943	1.000	0.003	0.725
Mean B7								-----	0.991	0.281	0.125	0.469	0.971	-0.330	-0.002	0.605	-0.357
Mean B8									------	0.287	0.138	0.489	0.971	-0.329	-0.002	0.622	-0.365
Mean NPCI										-----	0.902	0.405	0.324	-0.877	0.830	0.402	0.350
Mean PSRI											------	0.545	0.208	-0.749	0.752	0.479	0.256
Mean B5												------	0.605	-0.159	-0.001	0.960	-0.464
Mean B6													------	-0.332	-0.002	0.743	-0.402
Mean REPI														------	-0.941	-0.213	-0.564
Mean SDI															------	-0.001	0.727
Mean B4																------	-0.465
SD ARI																	------

4.3.5 Logging feature detectability

Seventy (70) canopy gaps formed after illegal logging of *Ocotea* trees were located in the field. In the WV-2 image and Google Earth covering the study area, 66 camphor trees could be identified, and with high confidence, their respective canopy gaps created after these trees were logged were identified in the WV-3 image. High confidence for gap identification meant a marked change in image pixels. Figure 4.7 shows the means of the CARI, the REPI, and the MCARI extracted from WV-2 (a) and WV-3 (b). The circles compare VI values before and after the SL events.

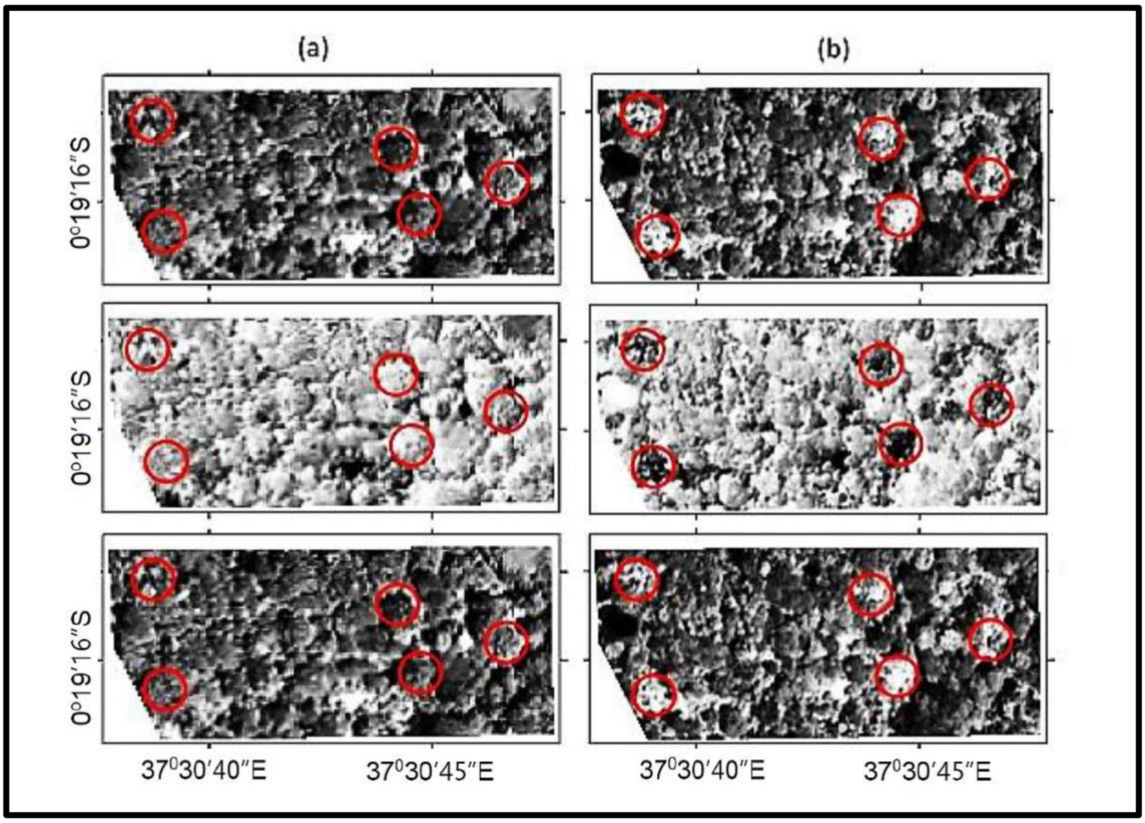


Figure 4.7. *The means of the chlorophyll absorption ratio index, the red edge position index, and the modified chlorophyll absorption ratio index extracted from 2014 pansharpened WorldView-2 data showing sites before logging events (a); and the means of the chlorophyll absorption ratio index, the red edge position index, and the modified chlorophyll absorption ratio index extracted from 2019 pansharpened WorldView-3 data after known logging events (b).*

4.3.6 Model Performance

Table 4.5 shows confusion matrices for RF and SVM classifiers, for the respective models whose overall accuracy (OA) was closest to the average OA. The average OA for the RF and SVM models performed relatively the same, 92.3 ± 2.6% and 94.0 ± 2.1%, respectively. The average kappa coefficient (κ) was 0.88 ± 0.03 and 0.90 ± 0.02 for the RF and SVM models, respectively. The user's accuracy (UA) ranges were between 82-100% for the RF model, and 86-100% for the SVM model. The producer's accuracy (PA) ranges were 83-100% for the RF model and 86-100% for the SVM model. Generally, the shaded gaps class showed the highest UA and PA, primarily because the other two classes represent vegetation. Therefore, the non-vegetation class was spectrally distinguished from the vegetation classes. In general, forest canopy had lower UA and PA, and this could be attributed to the range of reflectance characteristics by tree crowns in the WV-3 imagery.

Table 4.5: *Confusion matrices for (a) random forest algorithm, and (b) support vector machine algorithm for the best-performing model.*

(a) Random Forest algorithm						(b) Support Vector Machine algorithm					
	FC	SG	VG	Total	UA (%)		FC	SG	VG	Total	UA (%)
FC	83	I	13	97	86	FC	81	1	6	88	92
SG	0	89	0	89	100	SG	0	89	0	89	100
VG	7	0	77	84	92	VG	9	0	84	93	90
Total	90	90	90	270		Total	90	90	90	270	
PA (%)	92	99	86			PA (%)	90	99	93		
Overall accuracy = 92.2%; Kappa = 0.88						Overall accuracy = 94.1%; Kappa = 0.91					

4.3.7 Classification maps

Generally, the classified maps indicated that SL of *Ocotea* trees in MKFR, although carried out on a smaller scale caused widespread disturbance (Figure 4.8).

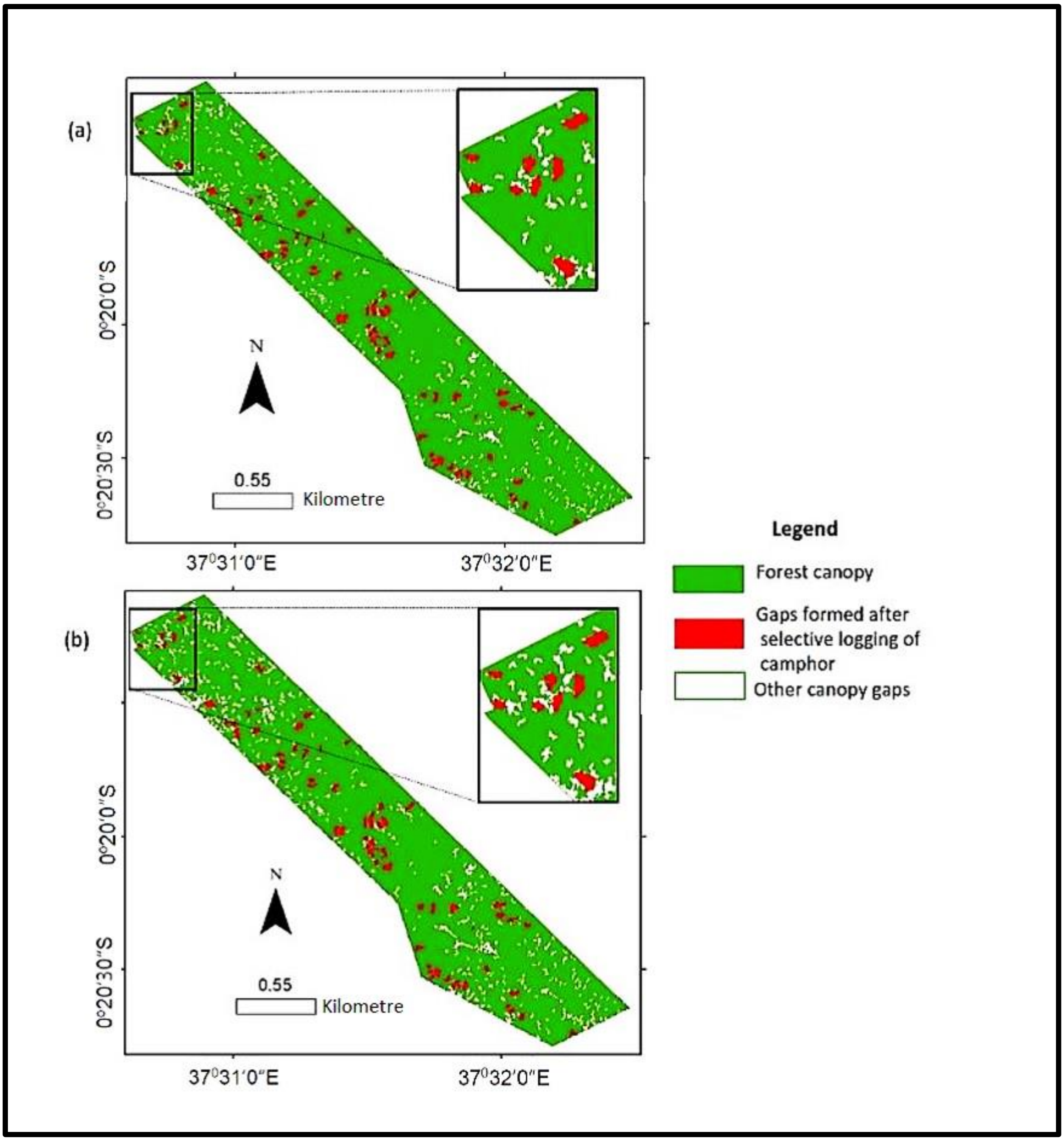


Figure 4.8. *Final thematic maps showing results of supervised pixel-based classification for (a) random forest, and (b) support vector machine.*

The McNemar's test return a Z value of 0.88 (Table 4.6). As a result, the confusion matrices of the two classifiers did not differ significantly ($Z \geq 1.96$) at a 95% confidence level.

Table 4.6: *The McNemar's test results comparing random forest and support vector machine classification models.*

Classifier		RF		
		CC	IC	Total
	CC	213	18	231
SVM	IC	13	26	39
	Total	226	44	270

CC stands for correctly classified and IC is for incorrectly classified samples.

4.4. Discussion

Using VHR multispectral satellite data, field data, and machine learning (ML) algorithms showed great potential for monitoring SL in Mt. Kenya Forest Reserve (MKFR)—the study portrayed great potential to monitor SL in similar tropical forests.

4.4.1 Relative importance of variables

The major objective of variable selection is to select a small number of variables—of high importance—enough for a good prediction, i.e. to make it simpler and faster to fit and predict. Generally, the brightness feature, together with individual means of the WV-3 VNIR bands, and those of VIs such as the CARI, the NPCI, the REPI, the SDI, and the standard deviation of ARI were critical in the detection of canopy gaps (Figure 4.3). Figure 4.6 shows that means of Bands 4 (Yellow) and 5 (Red) played a major role in the identification of shaded gaps. The means of the CARI, the CRI-2, the MCARI, the PSRI, the REPI, and Bands 1 (Coastal) 3 (Green), 5 (Red), and 6 (Red Edge) were crucial in the identification of vegetated gaps. The brightness feature and the means of bands 2 (Blue) and 7 (Near Infrared 1), the SDI and the NPCI, and the standard of ARI were significant in the identification of forest canopy class. Therefore, the index means outperformed the SD and ratio variables in terms of the detection of canopy gaps because spectral variability in areas of canopy gaps was only due to shadows from surrounding trees and/or low vegetation. Dalagnol *et al*. (2019) reported that the most important variables for tree loss detection were the SDs of the reflectance VNIR bands (especially band 5 - Red) and the shadow fraction. Therefore, marked increases in spectral variability in tree loss areas were due to shadows cast by nearby trees, the non-photosynthetic vegetation and exposed soil. In Malahlela *et al*. (2014), the observed improved results were associated with the use of the red-edge band of the WorldView-2 sensor. The vegetation indices were mostly derived from bands 5 (Red) and 6 (Red Edge) of the multispectral WV-3 imagery. They are, therefore, transferrable in other tropical, closed-canopy ecosystems with different species compositions where ground-truth data are not available. The WV-3 satellite SWIR sensor provides rich data for precisely identifying and characterizing forest landscape features, further enhancing WV-3's capability to monitor canopy gaps.

4.4.2 Pairwise feature comparison

Only correlation values >0.70 are significant (Table 4.4), thus, pairs of variables with such values are known to be correlated, providing redundant information, although ML algorithms can effectively handle this collinearity (Immitzer *et al*., 2012). Examples are the means of the visible bands, the NPCI, the CARI, the MCARI, the CRI-2, and the PSRI. The mean of REPI and those of CARI, CRI-2, MCARI, NPCI, and PSRI reported the highest negative correlations, i.e. –1.000, –0.718, –0.943, –0.877, and –0.749, respectively. The REPI and CARI exhibited a perfect negative correlation; a negative correlation means one variable is the opposite of the other. Therefore, the higher the absolute correlation coefficient, the more the variables might be critical during classification.

4.4.3 Spectral Separability

Figure 4.5 shows that, within one standard deviation, variables showed varying degrees of overlap between shaded and vegetated gaps, and forest canopy, e.g. the means of the visible bands, the CRI-2, the NPCI, the PSRI, and the standard deviation of ARI. It can also be noted that the brightness feature, the means of the CARI, the MCARI, the REPI, and band 6 (Red Edge) did not depict any spectral overlap between shaded gaps and the other two classes which portrayed significant spectral overlaps. Only the means of bands 7 (Near Infrared 1) and 8 (Near Infrared 2), and the SDI fully discriminated between the three classes. The transformed divergence (TD) index and Jeffries-Matusita (J-M) distance results showed the forest canopy and vegetated gaps yielded the lowest value of separation (Table 4.3). Generally, a value >1.9 indicated that the classes were separable, while 1.7 to 1.9, showed a fairly good separation, but <1.7 indicated a poor separation (Jensen, 1996). Higher spectral separability values translated to higher overall accuracy (OA). The TD index recorded higher values compared to the J-M distance, although the former recorded the lowest value of 1.48 compared to J-M's 1.63. The means of the VIs produced higher values compared to the means of VNIR bands. The separability between the forest canopy and vegetated gaps is low in both TD index and J-M distance by the means of the WorldView-3 VNIR bands—the VIs indicated that forest canopy and vegetated gaps were separable. Furthermore, one advantage of machine learning classifiers, *e.g.*, SVM is that they are suited for extreme case binary classification (Kuter, 2021).

Both bands 5 (Red) and 6 (Red Edge) were used to compute most of the vegetation indices—this indicated the role of the two bands in canopy gap mapping. The saturation problem is minimized whenever band 6 is used in computing vegetation indices. This improved detection of canopy gaps since separability between canopy gaps and forest canopy was increased—minimized the confusion between vegetated gaps and forest canopy (Mutanga and Skidmore 2004; Malahlela *et al*. 2014). This study found that greater spectral similarity within a class and greater spectral variability between classes increased classification accuracy (Jackson and Adam, 2021).

4.4.4 Model performance

All classes reported very high UA and PA, thus, high OA for the RF and SVM models (Table 4.5). This was attributed to the features extracted from the pansharpened WV-3 multispectral data. As well, the large sample size for the classes could have offered more accurate mean values, leading to a positive impact on classification results. Furthermore, manually delineating the forest landscape classes meant only relevant pixels were obtained. Generally, the shaded gaps class showed the highest UA and PA—was expected because the other two classes represent vegetation, therefore, their spectral characteristics are different from the non-vegetation class. The results showed that the SDI accurately extracted shaded gaps. In general, forest canopy recorded lower UA and PA—this could be attributed to a wide range of reflectance characteristics by tree crowns in the WV-3 imagery. The OA values of 92.3 ± 2.6%, and 94.0 ± 2.1%, for RF and SVM, respectively surpass the minimum threshold of 85% used as a benchmark for classification with high accuracy (Pringle *et al*., 2009). This indicated a good agreement between the classification of SL of *Ocotea* trees and field data. Therefore, the approach detailed in this study could be applied to any other VHR satellite data.

4.4.5 Spatial Distribution of canopy gaps

Knowledge about where canopy gaps are located and how they are distributed is critical in understanding their importance in the global carbon cycle, and also for managing the proliferation of invasive species. This study found that canopy gaps formed after logging of *Ocotea* trees are widespread in the study area. Although they dominated the wet forests of Mt. Kenya, the endangered tree species have become scarce. The canopy gaps stimulate the regrowth of secondary forests dominated by fast-growing species, especially *Macaranga kilimandscharica* and *Neoboutonia macrocalyx*, significantly reducing the presence of shade-tolerant species. In this study, the McNemar test was meant to indicate whether a difference of 1.7% was statistically important. The Z value was 0.88, meaning there are no significant differences ($Z \geq 1.96$) at a 95% confidence level, existing between the two maps. Therefore, forest managers and policy makers can use the maps to make better-informed decisions to implement priority on-the-ground conservation and restoration projects in MKFR.

4.4.6 Challenges, limitations, and the way forward

First, mapping canopy gaps in tropical forests using optical RS systems remain a challenge because of persistent cloud cover. Dense clouds obscure tree canopies, and reflectance values of forest landscape features are altered by light clouds and cloud shadows. The problem is further compounded by unreliable cloud and cloud shadow detection algorithms (Masiliūnas 2017). The rough terrain in MKFR and extensive cloud cover exhibited by the WV-3 image limited size of the study area. Other data sources, such as synthetic aperture radar (SAR) which can capture images of the earth's surface regardless of smoke, darkness, or cloud cover can be explored. Although SAR systems suitable for small UAVs are still in the development stage (Whitehead and Hugenholtz, 2014). Second, trees that had dropped their leaves or lianas on top of tree crowns that experience sudden dieback may wrongly be interpreted as canopy gaps. In obtaining reliable quantitative data for verification, Immitzer *et al*. (2012) recommend stereoscopic photo interpretation rather than measurements in the field. In their absence, the application of time series methods that use seasonal models should be explored. Third, the method used in this study might not be applicable in sparser forests, because the only notable change is the disappearance of the shadow of the logged trees. Also, a tree can be felled in the direction of an existing gap, where the existing gap may have an insignificant increase in size. Researchers should aim to develop accurate methods to detect such canopy gaps. Fourth, higher classification accuracies can be achieved using LiDAR data (Dalagnol *et al*., 2019)—short flight sessions and the high cost of these surveys with experienced personnel, and high data dimensionality prevent continuous studies in tropical forests in developing countries (Mutanga *et al*., 2012). Nevertheless, data integration, which is heavily dependent on the compatibility of multi-source RS data, specifically the consistency in spatial, spectral, temporal, and radiometric resolutions accompanied by more sophisticated data fusion techniques, is key to the measurement and mapping of forest attributes (Jackson and Adam, 2020). Furthermore, LiDAR may serve as a sampling technique when trying to scale-up SL events to broader regions. In its absence, additional sources of publicly accessible training data can be used. Still, the development of advanced automated methods for processing LiDAR data would lower data-processing costs, allowing for data acquisition in extensive areas (Jackson and Adam, 2020). Further studies should make use of UAVs to cover larger areas at reduced costs, with caution taken regarding the current challenges posed by UAVs. Furthermore, airborne/spaceborne hyperspectral imagery covering extensive geographic areas may be obtained for such research through the use of hyperspectral imaging satellites which are to be launched in the next few years—*e.g.*, NASA's Surface Biology and Geology (SBG) mission and the Carbon Mapper constellation of satellites.

While the metrics extracted from WV-3 VNIR bands identified canopy gaps, there is a need to try other methods in combination with ML or statistical classifiers. Spatial features such as various forms of texture analysis and normalization based on neighbouring pixels can also be tested, to see whether making use of information in the surrounding pixels in time series analysis, rather than looking at pixels individually, improves detectability.

4.5 Conclusions

This study used very high resolution (VHR) multispectral satellite data and machine learning (ML) algorithms, with the help of historical data from WorldView-2 (WV-2) and Google Earth to detect canopy gaps resulting from illegal logging of *Ocotea* trees. Vegetation indices majorly derived from bands 5 (Red) and 6 (Red Edge) of the WorldView-3 (WV-3) satellite imagery—the CARI, the CRI-2, the MCARI, the PSRI, the REPI and the NPCI—detected canopy gaps. The SDI significantly played a key role in mapping the shaded gaps. The manual delineation pixels of the samples of the three classes may have as well contributed to higher classification accuracies because only relevant pixels were extracted. The WV-3 sensor's spectral and spatial characteristics and accurate tree-scale validation data from WV-2 and Google Earth resulted in high classification accuracies. In order to achieve higher classification accuracy, future research will explore the data integration approach—LiDAR and WV-3 multispectral data. The results reported in this study could allow forest managers to improve methods of detecting canopy gaps at larger scales using remotely sensed data and relatively little additional fieldwork. The classification maps in the study can be used to implement on-the-ground conservation and restoration projects in MKFR.

CHAPTER FIVE

Feature extraction and classification of canopy gaps using GLCM - and MLBP-based rotation-invariant feature descriptors derived from WorldView-3 imagery

This chapter is based on:

Jackson, C.M., and Adam, E. (in review). Feature extraction and classification of WorldView-3 imagery using GLCM - and MLBP-based rotation-invariant feature descriptors for detection of canopy gaps. *Tropical Conservation Science*.

Abstract

Selective logging (SL) in tropical forests has the potential to reduce the ecosystem services provided by those forests. Accurate mapping of SL is critical because it serves as the foundation for additional research on forest restoration and regeneration, species diversification and distribution, and ecosystem dynamics, among other topics. In order to model canopy gaps created by illegal logging of *Ocotea usambarensis* in Mt. Kenya Forest Reserve (MKFR), the study used a texture-spectral analysis framework to exploit the potential of the very high resolution (VHR) WorldView-3 (WV-3) multispectral imagery. For comparison purposes, two classification schemes were used— first, texture properties are explored in the sub-band images using fused grey-level co-occurrence matrix (GLCM) and local binary pattern- (LBP-) based texture feature extraction. Second, the GLCM statistical model of texture analysis was combined with the multivariate local binary pattern (MLBP) model—an extension of the basic LBP model. Two advanced machine learning (ML) classification models—random forest (RF) and support vector machine (SVM) models were used to identify and classify distinctive features in the spectral and texture domains of the WV-3 multispectral imagery. The results showed individual overall accuracy (OA) scores for the respective MLBP models were in the range of 80-95.1%. The average OA scores were 84.68 ±3.1, 84.54 ±2.5, 84.86 ±3.0, 86.46 ±3.9, 87 ±4.0, and 85.44 ±3.7 for image blocks A, B, C, D, E, and F, respectively, for the RF model, and 85.44 ±3.6, 87.2 ±1.8, 86.3 ±4.3, 89.84 ±2.8, 87.28 ±4.5 and 86.12 ±3.6 for the SVM model. The individual user's accuracy (UA) and producer's accuracy (PA) for the MLBP models were in the range of 77-100%. Classification results from univariate LBP models showed lower UA and PA—67-75%. The spectral enhancement of texture features improved the classification results significantly.

Keywords: Tropical forests. WorldView-3. Canopy gaps. *Ocotea usambarensis*. Texture analysis. Grey level co-occurrence matrix. Multivariate local binary pattern. Machine learning algorithms.

5.1 Introduction

Forests of the World serve various economic, social, and environmental functions (Solberg *et al*., 2008). Tropical forests are *c*. 7% of the earth's terrestrial environment, and house about half of all biodiversity (Solberg *et al*., 2008). Those forests are decreasing both quantitatively and qualitatively because of various problems, such as selective logging (SL). Selective logging (SL) reduces forest density when sparsely distributed, and most valuable trees are cut, creating canopy gaps (Jackson and Adam, 2020). Forest gaps from SL can alter the intercepted photosynthetically-active radiation (IPAR), sensible and latent heat fluxes, water stress, and eventually the productivity of plants (Healey *et al*., 2000). According to Gatti *et al*. (2014), canopy gaps are the legacy of disturbance events in forest landscapes and they significantly affect the composition of species, reduce above-ground biomass (AGB), alter forest structure, and carbon budgets in the short- and mid-term. Also, SL may introduce regeneration and growth of secondary forests consisting of light-demanding species, replacing the shade-tolerant species which form most of the commercial species in some tropical forests (Jackson and Adam, 2020). According to Putz *et al*. (2000), among all forest uses, logging has the worst effects on the environment, both directly and indirectly, e.g. increased soil erosion and runoff lead to reduced nutrients and organic matter in the soil. Previous studies on the estimation of deforestation rates in tropical forests generally ignored the effects of SL (Achard *et al*., 2002). However, recently researchers have emphasized the contribution of illegal and selective logging to the rates of deforestation (Edwards *et al*., 2014).

Adequate field sampling of remote selectively logged forests is costly and difficult (d'Oliveira *et al*., 2012), thus the potential for better management of tropical SL remains uncertain. In SL, the disturbance in the form of a logging gap of a large tree cut down can be visible from above (Masiliūnas, 2017). Remote sensing (RS) can monitor forests, even in the remote places of the vast forested tropical areas (Masiliūnas, 2017). However, canopy gaps formed from SL are hard to detect especially in images whose pixels are bigger than the tree crowns (Descals *et al*., 2017). Yuan (2014) noted that improved RS data acquisition, i.e. in terms of spatial, spectral, and temporal resolution capabilities, is not in tandem with the capability for data analysis and understanding. Accordingly, the mixed pixels' complex spectral reflectance in disturbed tropical forests calls for enhanced geospatial methods for detecting SL (Graça *et al*., 2015).

The application of texture analysis in RS studies reported great achievements (Lucieer *et al*., 2005). Texture is the frequency change in the tone of pixels in images (Lucieer *et al*., 2005). Unlike pixel-based techniques, texture-based classification techniques considered how a pixel related to its neighbourhood (Kim *et al*., 2002; Suruliandi and Jenicka, 2015). A successful classification of texture requires (Suruliandi and Jenicka, 2015):

- the selection of features that discriminate between image textures;
- an adequate number of training samples;
- an appropriate neighbourhood with a pattern unit;
- the right window size, and
- the development of classification models that distinguish between texture features related to different texture classes.

In RS, a variety of generally categorized approaches have been applied in extracting textural features from images (Jenicka and Suruliandi, 2015; Suruliandi and Jenicka, 2015). Originally by Haralick *et al*. (1973), texture measures, e.g. statistical metrics are extractable from GLCM. Model-based approaches classify textured images according to probability distributions in random fields, e.g. Cross and Jain (1983), and the local linear transformations, e.g. Unser (1995). He and Wang (1990) relied on the texture spectrum. Overall, all these approaches have limited application because of their high computational and time complexities (Kim *et al*., 2002; Arivazhagan and Ganesan, 2003). Therefore, their spatial analysis is mostly applied to small neighbourhoods, on a single scale (Arivazhagan and Ganesan, 2003). This difficulty has been solved through the development of multi-channel-based image analysis (Arivazhagan and Ganesan, 2003; Jenicka and Suruliandi, 2015; Suruliandi and Jenicka, 2015). Therefore, the multichannel filtering methods, i.e. the signal processing methods are widely used because of their simplicity (Kim *et al*., 2002). A textured image is normally reduced into characteristic feature images by applying, e.g. wavelet, Gabor, or neural network-based filters (Jenicka and Suruliandi, 2015; Suruliandi and Jenicka, 2015). Thus, with just a few feature statistics, a high-dimensional textural pattern can be modelled (Jenicka and Suruliandi, 2015; Suruliandi and Jenicka, 2015). The local binary pattern (LBP) model was developed by Ojala *et al*. (2001) for gray-level images. A multivariate local binary pattern (MLBP) model was done by Lucieer *et al*. (2005) for RS images. Arivazhagan and Ganesan (2003) and Arivazhagan *et al*. (2006) used wavelet transform and rotation invariant features computed from Gabor coefficients of an image, respectively. Liao and Chung (2007) proposed the advanced local binary pattern (ALBP), while Suruliandi and Ramar (2008) used a local texture pattern (LTP) for gray level images and later the multivariate local texture pattern (MLTP). Liao *et al*. (2009) applied the dominant LBP to find the dominant patterns contributing 80% of pattern units. Support vector machine (SVM), self-organizing map, and fuzzy *k*-nearest neighbor (*k*-NN) classification algorithms were used with a modified MLBP texture model on RS images by Jenicka and Suruliandi (2011). Yuan (2014) performed RS image segmentation and object extraction using locally excitatory globally inhibitory oscillator network (LEGION) framework. Jenicka and Suruliandi (2015) proposed the multivariate ternary pattern (MTP) for feature extraction. Suruliandi and Jenicka (2015) used MLBP, multivariate local texture pattern (MLTP), multivariate advanced local binary pattern (MALBP), wavelet, and Gabor wavelet models to extract textural features. Jenicka and Suruliandi (2014a) applied a multivariate discrete local texture pattern (MDLTP) model for land cover classification. Shao *et al*. (2014) proposed the color Gabor wavelet texture for RS image retrieval. Jenicka and Suruliandi (2014b) used a fuzzy local texture pattern (FLTP) and the multivariate fuzzy local texture pattern (MLFTP) models for land cover classification. Among the texture models, the variants of LBP (Ojala *et al*., 2001) like MLBP (Luceer *et al*., 2005), MLTP, and MALBP (Suruliandi and Jenicka, 2015) are computationally convenient (Lucieer *et al*., 2005).

The spectral information contained in multispectral RS images is, in essence, limited (Ghasemi *et al*., 2013). Therefore, texture features of a RS image can reveal the spatial correlation of pixels, showing the change of vegetation structure (Zhou *et al*., 2017). Very high resolution (VHR) RS data and grey-level co-occurrence matrix (GLCM) analysis have been successfully used in mapping tropical forests (Rakwatin *et al*., 2010; Wang *et al*., 2016; Marissiaux, 2018; Burnett *et al*., 2019). Among the texture models, the variants of LBP (Ojala *et al*., 2001) like MLBP (Luceer *et al*., 2005), MLTP, and MALBP (Suruliandi and Jenicka, 2015) were computationally convenient (Lucieer *et al*., 2005). Therefore, using a study area in a very dense heterogeneous tropical forest, this research work was aimed to demonstrate the suitability of a fused texture-spectral analysis to model selectively logged trees in Mt. Kenya Forest Reserve (MKFR). Therefore, two models were investigated— first, a fused grey-level co-occurrence matrix (GLCM) and local binary pattern (LBP) based texture feature extraction—to extract GLCM features from texture images—individual WorldView-3 (WV-3) bands. Second, the GLCM statistical model of texture analysis was combined with the multivariate local binary pattern (MLBP) model—an extension of the basic LBP model—to extract the GLCM features from the WV-3 composite images. The random forest (RF) and support vector machine (SVM) classification models were used to identify and classify canopy gaps. The two were regarded superior to other classifiers (Kamarulzaman *et al*., 2022; Hethcoat *et al*., 2021, 2020, 2019; Rex *et al*., 2020; Dalagnol *et al*., 2019). The RF and SVM have low sensitivity to problems like noisy data and over-training on large forests (Rodriguez-Galiano *et al*. 2012).

5.2 Materials and methods

5.2.1 Acquisition and pre-processing of satellite data

A WorldView-3 (WV-3) multispectral image acquired on 15 September 2019 was used for detecting canopy gaps in selectively logged sites, and WorldView-2 data acquired on 30 January 2014 and Google Earth were sources of historical data. The satellite data of Mt. Kenya Forest Reserve (MKFR) was provided by Swift Geospatial, Pretoria, South Africa. WorldView-2 (WV-2) satellite captures panchromatic images (450-800 nm) at 0.46 m pixel size and multispectral images with eight visible-near-infrared (VNIR) bands (400 to 1040 nm) at 1.84 m (DigitalGlobe, 2010). WorldView-3 (WV-3) captures panchromatic images at a pixel size of 0.30 m (DigitalGlobe, 2019). It also captures multispectral imagery with eight VNIR bands at 1.2 m and eight shortwave-infrared (SWIR) bands (1195–2365 nm) at 3.7 m (DigitalGlobe, 2019). The CAVIS (clouds, aerosol, vapour, ice, and snow) sensor acquires 12 CAVIS bands with a pixel size of 30 m (DigitalGlobe, 2019).

The WV-3 data was co-registered with the WV-2 image to be able to match features between the two datasets—resulted in an average root mean square error (RMSE) of 3.41 m. Before calculating textural features, the VNIR bands were down-scaled to 0.3 meter pixels, with the 1.2 m pixels sub-divided into 16 pixels with a conserved value (Burnett *et al*., 2019). A major challenge with LBP is that in VHR RS data it may interpret small texture variation as a different texture feature. The method necessitated the extraction of rich textural information without the inclusion of uncertainties of pansharped VNIR bands (Li *et al*., 2017).

5.2.2 Acquisition of field data

A Global Positioning System (Garmin eTrex® 20 GPS Receiver) and false-colour composite (853-RGB) of WV-3 image, were used to locate canopy gaps in the field. The three bands were among the best performing bands in the previous chapter, therefore, they were used in this study's analysis. In the WV-3 image, the canopy gaps were either partially/fully illuminated or not illuminated at all. In the study area, the human-made canopy gaps shared similar reflectance characteristics with natural canopy gaps. GPS coordinates of 100 vegetated gaps and 100 shaded gaps per image block were collected. A vegetated gap is a forest canopy gap with low vegetation inside it, which is the initial stage of vegetation recovery from forest disturbance.

Approximate locations of gap centres were recorded—then the GPS points collected were overlaid on the WV-3 image. Using a geographic information system (GIS—ArcGIS® v. 10.3; ESRI, Redlands, CA, USA), vegetated and shaded canopy gap pixels were extracted from the WV-3 imagery. The spatial resolution of the WV-3 data enabled the derivation of forest canopies as references—thus 100 samples of tree crowns were extracted.

Appropriate image block sizes were selected to calculate texture features. Regions in large blocks show a mixture of textures, while small blocks may reduce the probability of computing a texture measure (Lucieer *et al*., 2005). In this study, six non-overlapping subset images (each 1400 × 1400 pixels) were extracted from the WV-3 imagery covering the study area. As seen in Figure 5.1, the three (vegetated and shaded forest canopy gaps, and forest canopy) classes were easily differentiated in the WV-3 imagery.

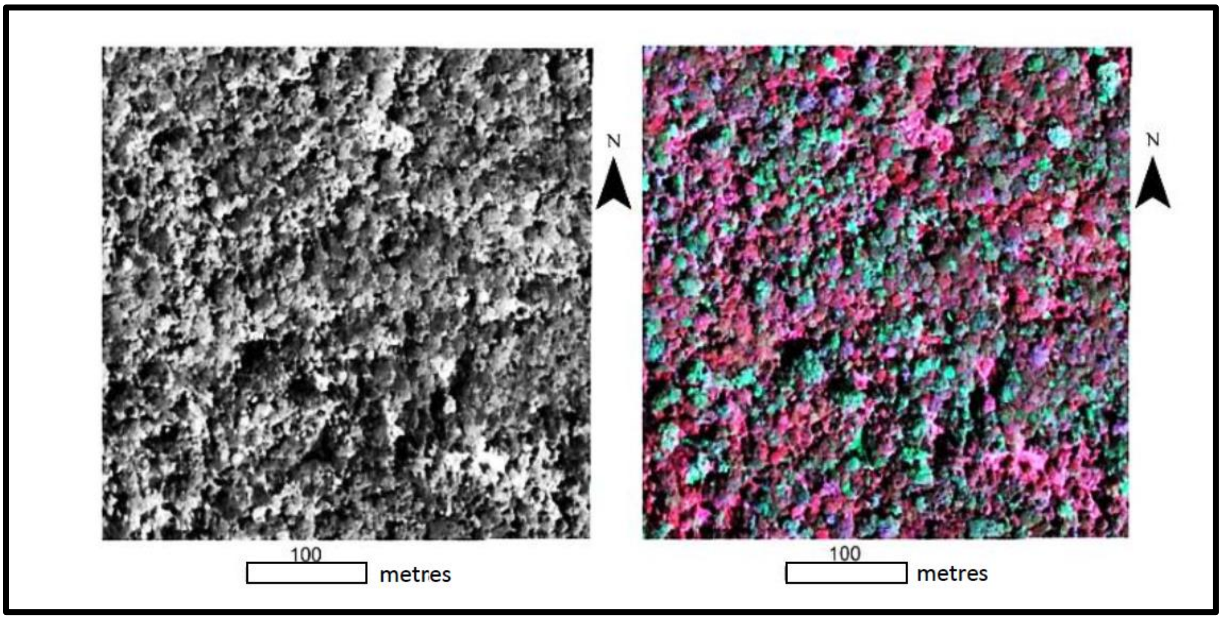
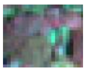


Figure 5.1. *A subset of the WorldView-3 image of the study area: (a) band 8; (b) colour composite of bands 8, 5, and 3 (RGB).*

Pixels were sampled randomly, covering areas close to the class edges and centres—pixels around class edges were vital in aiding the classifier's edge detection of textural features (Burnett *et al*., 2019). To attain high classification accuracy, the pixel size of the reference data corresponding to the texture classes (vegetated and shaded forest canopy gaps, and forest canopy) were kept the same, i.e. consisting of 20 × 20 pixels (Table 5.1). The same number of reference points for the three classes was collected because data imbalance reduces the accuracy and performance of classifier (Huang *et al*., 2008; Graves *et al*., 2016; Maxwell *et al*., 2018).

Table 5.1: *Reference data and their description.*

Class #	Class	Sample	Description
Class 1	Vegetated gap		Low-lying vegetation in the forest canopy
Class 2	Shaded gap		Gaps in the forest canopy that are darker because of the shadows cast by the nearby tree crowns
Class 3	Forest canopy		The topmost layer of a forest, consisting mostly of tree crowns with a few emergent trees having heights that shoot above the canopy.

Canopy gaps formed after logging of *Ocotea* trees had their dimensions recorded in the field, i.e. their dripline measurements, maximum length, and compass orientation, together with the maximum width, perpendicular to the length (Fox *et al*., 2000). Points directly below the dripline were noted and since it was impossible to all of the canopy drip-line, the boundaries were somehow generalized. Using ArcGIS, a map of canopy gaps was generated from ground survey data. The accuracy of canopy gap delineation using RS was assessed by comparing the dimensions collected from the field.

5.2.3 Feature extraction and selection

Texture analysis is an effective component of classification for higher resolution images (Kupidura, 2019). It is convenient to use because it does not need prior image segmentation (Kupidura, 2019). A crucial property of texture is the repetitive nature of the pattern(s) in an area (Suruliandi and Jenicka, 2015). The spectral information in images has been frequently used in interpreting and analysing images, however, images may have an object reflecting differently, and different objects reflecting the same (Cui *et al*., 2009). This affects the accuracy of RS image analysis. Improvement in the spatial resolution of RS images has contributed to more spatial structures and texture features, which has led to increased classification accuracy.

According to Cohen and Spies (1992), texture features drawn from images of higher spatial resolution have advantages for forestry applications. Lucieer *et al*. (2005) and

Suruliandi and Jenicka (2015) note that significantly high classification accuracies have been achieved using textural information. Among many different methods used to extract model texture from images, the texture-spectral analysis framework only evaluated the widely used GLCM texture measures and state-of-art texture descriptor Local Binary Patterns (LBP) and the multivariate local binary pattern (MLBP)—an extension of the basic LBP. The GLCMs are theoretically simple and easy to implement and they generate fewer features (Bianconi and Fernández, 2014). Even for low spatial resolution RS data, the LBP is an effective model in feature extraction, and according to Luceer *et al*. (2005), the model can be used in different RS images and applications. In order to detect canopy gaps, an overview of how these texture descriptors extracted texture from the WV-3 dataset is given.

5.2.3.1 The local binary pattern (LBP) texture-based model

The LBP model was put forth by Ojala *et al*. (2002). Ojala *et al*. (2002) defined texture T to derive the LBP operator (Equation (5.1)):

$$T = t(g_c, g_0, \dots, g_{P-1}) \qquad (5.1)$$

where g_c symbolizes the centre pixel's grey-scale value (P_c) in the neighbourhood and g_i (i = 0, …, $P - 1$) is a neighbouring pixel's grey-scale value. This study used a neighbourhood set that is symmetrically circular—a radius R with P evenly spaced pixels (Ojala *et al*., 2002) (Equation (5.2)):

$$i = -R\, sin\left(\frac{2\pi p}{P}\right); \quad j = R\, cos\left(\frac{2\pi p}{P}\right) \quad \forall\ p \in \{0, 1, \dots, P\} \qquad (5.2)$$

A pixel's circularly symmetric neighbour set consists of radius R and equally spaced pixels P. With the possibility of information loss, the texture is turned into the joint difference. This was done by subtracting the central pixel's grey value from all neighbourhood pixels. The joint difference distribution captures patterns in each pixel's neighbourhood in a P-dimensional histogram. Invariance in respect of the scaling of values of pixels or differences in illumination is attained by using the signs of the differences (Ojala *et al*., 2002) (Equation (5.3)):

$$T^* \approx t\big(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)\big) \qquad (5.3)$$

where g_P is the gray value of the p neighbour.

According to Ojala *et al*. (2002), LBP defines only the central pixel's uniformity towards its neighbours, without defining the uniformity of the neighbourhood itself. To capture uniformity in a neighbourhood set, a measure of uniformity, U was introduced. With $g_p = g_0$, U_c is defined as (Lucieer *et al*., 2005) (Equation (5.4)):

$$U_c = \sum_{i=1}^{P} | s(g_i - g_c) - s(g_{i-1} - g_c)| \tag{5.4}$$

Uniform patterns are those with $U_c \leq j$ and Ojala *et al*. (2002) indicated that the best results are achieved when j = 2. This leads to the following (Lucieer *et al*., 2005) (Equation (5.5)):

$$LBP_{c,j} = \begin{cases} \sum_{i=0}^{P-1} s(g_i - g_c) & \text{if } U_c \leq j \\ P+1 & Otherwise \end{cases} \tag{5.5}$$

On a circular radius of *R*, The $LBP_{c,j}$ operator thresholds pixels in a circular pattern at the value of the centre pixel, in the neighbourhood of *P* evenly spaced pixels. It is capable of detecting uniform patterns for any angular space quantization and spatial resolution.

The $LBP_{c,j}$ were fused with rotation invariant GLCM measures discussed in section 5.2.3.3, i.e. homogeneity, contrast, entropy, angular second moment, and correlation, which led to the following features: $LBP_{c,j}/HOM_c$, $LBP_{c,j}/CON_c$, $LBP_{c,j}/ENT_c$, $LBP_{c,j}/ASM_c$, and $LBP_{c,j}/COR_c$., for bands 3 (Green), 5 (Red), and 8 (Near Infrared 2) of the WorldView-3 (WV-3) image. The values of these features were computed and allocated to the image pixels, thus revealing textural patterns. Therefore, the histogram of the joint the $LBP_{c,j}$ and GLCM feature occurrence formed the final texture feature.

5.2.3.2 The multivariate local binary pattern (MLBP) texture model

The $LBP_{c,j}$ operator describes the texture of a single band. To improve classification accuracy, Lucieer *et al*. (2005) applied the LBP texture measure to colour images by proposing a multivariate texture model, i.e. the MLBP operator, which describes local pixel relations in three bands (Lucieer *et al*., 2005; Suruliandi and Jenicka, 2015) (Equation (5.6)):

$$MLBP_c = \sum_{i=0}^{P-1} \begin{matrix} s\left(g_i^{band\,1} - g_c^{band\,1}\right) + s\left(g_i^{band\,2} - g_c^{band\,1}\right) + s\left(g_i^{band\,3} - g_c^{band\,1}\right) \\ + s\left(g_i^{band\,1} - g_c^{band\,2}\right) + \left(g_i^{band\,2} - g_c^{band\,2}\right) + \left(g_i^{band\,3} - g_c^{band\,2}\right) + \\ s\left(g_i^{band\,1} - g_c^{band\,3}\right) + s\left(g_i^{band\,2} - g_c^{band\,3}\right) + s\left(g_i^{band\,3} - g_c^{band\,3}\right) \end{matrix} \tag{5.6}$$

Three 3 × 3 matrices describe the local texture in individual bands, while six 3 × 3 matrices compare texture among bands.

In the MLBP model, the univariate GLCM measure (e.g. HOM) was extended as multivariate homogeneity (MHOM), i.e. the individual independent homogeneities HOM3, HOM5, and HOM8 representing bands 3 (Green), 5 (Red), and 8 (Near Infrared 2), respectively, were combined into a single composite homogeneity (MHOM). An image's global texture pattern description was derived by combining the MLBP and MHOM in a 2-D histogram. In the 2-D histogram, the x ordinate denotes MLBP and the y ordinate denotes MHOM. In order to incorporate color into the MLBP model, the same procedure was repeated for the MLBP and the remaining GLCM feature composites—the respective composites of contrast (MCON), entropy (MENT), angular second moment (MASM), and correlation (MCON).

5.2.3.3 The grey level co-occurrence matrix (GLCM) model

The co-occurrence matrix (Haralick *et al*., 1973) examines the texture of a grey-scale image. Co-occurrence matrices, among the oldest image descriptors (they first appeared fifty years ago), have wide applications in almost all areas of image analysis. If $I(k,k)$ correspond to the grey-scale image of a central pixel (n_c,m_c), then the co-occurrence matrix $C_M = C_{(Dx,\ Dy)}(k,k)$ is defined as follows (Lambros *et al*., 2017) (Equation (5.7)):

$$C_M = \sum_{n=1}^{k} \sum_{m=1}^{k} \begin{cases} 1 \ if \ I(n,m) \ = k \ \ and \ I\left(n + D_x, m + D_y\right) = k \\ 0 \ otherwise \end{cases} \tag{5.7}$$

where (D_x,D_y) are defined as (Lambros *et al*., 2017) (Equation (5.8)):

$$D_x = D.\cos(\theta), \ \ D_y = D.\sin(\theta), \tag{5.8}$$

where θ defines the direction of the matrix from the central pixel (n_c,m_c) and D is the distance from the central pixel (n_c,m_c) (Figure 5.2).

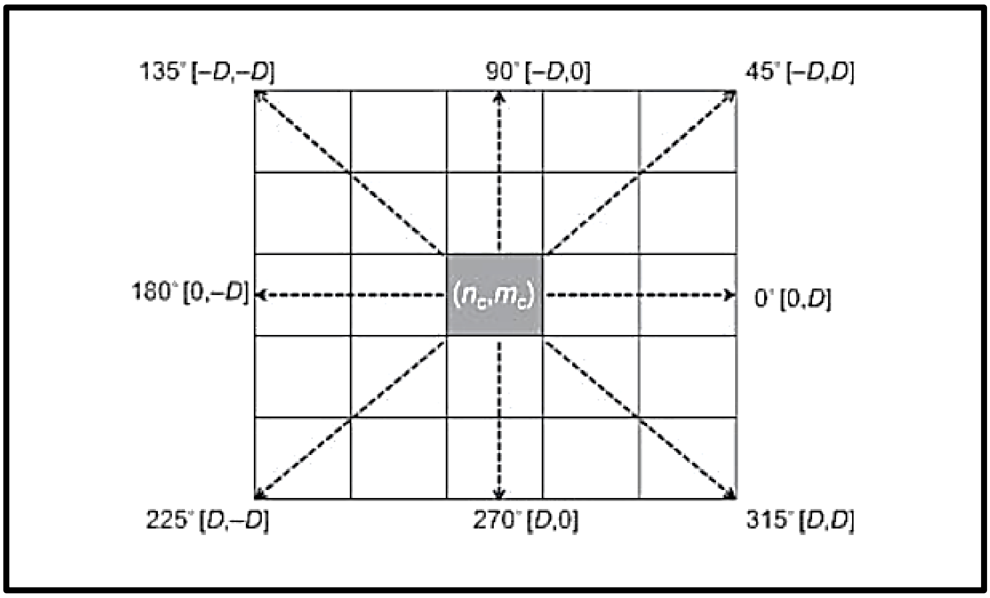


Figure 5.2. *The spatial relationships of the pixels, where D represents the distance from the central pixel (nc,mc).*

Therefore, GLCM is an excellent statistical approach used in expressing spatial relationships between each pixel and the surrounding pixels, depicting the uniqueness of the original image. GLCM can be used for forest canopy gap detection. Only five most frequently used features were applied in this study, i.e. entropy (Equation (5.9)), angular second moment (Equation (5.10)), contrast (Equation (5.11)), homogeneity (Equation (5.12)), and correlation (Equation (5.13)) (Table 5.2). Entropy measures randomness while the angular second moment is a statistical measure of the degree of pixel pair repetition. Contrast measures the amount of local dissimilarity. Homogeneity is the opposite of contrast. How a pixel in an image is correlated to its neighbour is described by correlation. The GLCM features were calculated using the following mathematical expressions (Hall-Beyer, 2017).

Table 5.2: Texture measures used in this study, their abbreviations, and equations.

Abbreviation	Name	Equation	
ENT	Entropy	$\sum_{i,j=0}^{n-1} P_{i,j}\left(-\ln P_{i,j}\right)$	(5.9)
ASM	Angular second moment	$\sum_{i,j=0}^{n-1} P_{ij}^2$	(5.10)
CON	Contrast	$\sum_{i,j=0}^{n-1} P_{i,j}\,(i-j)^2$	(5.11)
HOM	Homogeneity	$\sum_{i,j=0}^{n-1} \left(P_{i,j}/1+(i-j)^2\right)$	(5.12)
COR	Correlation	$\sum_{i,j=0}^{n-1} P_{i,j}\left[(i-\mu_i)(i-\mu_j)/\sqrt{(\sigma_i^2)(\sigma_j^2)}\right]$	(5.13)

where $P_{i,j}$ corresponds to the statistical probability values for changes between grey levels i and j at distance d and angle θ. N is the number of grey levels in the image. In the GLCM COR equation (Equation 13), μ is the mean (Equation 5.14) and σ *is* the standard deviation (Equation 5.15) of the texture image, respectively.

$$\mu_i = \sum_{i,j=0}^{n-1} i(P_{i,j})\;\mu_j = \sum_{i,j=0}^{n-1} j(P_{i,j}) \tag{5.14}$$

$$\sigma_i^2 = \sum_{i,j=0}^{n-1} P_{i,j}\,(i-\mu_i)^2;\; \sigma_j^2 \sum_{i,j=0}^{n-1} P_{i,j}\,(j-\mu_i)^2 \quad \therefore \quad \sigma_i = \sqrt{\sigma_i^2};\; \sigma_j = \sqrt{\sigma_j^2} \tag{5.15}$$

As earlier noted, GLCM is directional, therefore, sensitive to rotation (Luceer *et al*., 2005). In achieving rotation invariant descriptors, the circular symmetric neighbour set approach was used to represent the displacement, thereby offering a better description of fine textures.

5.2.4 Similarity and separability between training signatures

The G-statistic was used to evaluate the similarity between different textures by comparing bins of the histogram of a texture sample and histogram of a texture model. For the statistical global texture features the G-statistic is defined (Sokal and Rohlf, 1987) (Equation (5.16)):

$$G = 2\left(\left[\sum_{s,m}\sum_{i=1}^{tb} f_i \log f_i\right] - \left[\sum_{s,m}\left(\sum_{i=1}^{tb} f_i\right)\log\left(\sum_{i=1}^{tb} f_i\right)_i\right] - \left[\sum_{i=1}^{tb}\left(\sum_{s,m} f_i\right)\log\left(\sum_{s,m} f_i\right)\right] + \left[\left(\sum_{s,m}\sum_{i=1}^{tb} f_i\right)\log\left(\sum_{s,m}\sum_{i=1}^{tb} f_i\right)\right]\right) \quad (5.16)$$

where sample s corresponds to a histogram of the texture measure distribution while model m corresponds to a histogram of a reference area. *tb* constitutes the number of bins and *fi* represents the probability for bin *i*. For the spectral global features, the spectral distance between training signatures was evaluated to determine whether spectral classes that are almost the same could lead to reduced classification accuracy. This study implemented the Euclidean distance (Tuominen and Lipping, 2016) (Equation (5.17)):

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (5.17)$$

where x is the vector of the first spectral signature and y is the vector of the second spectral signature. *n* is the number of image bands.

5.2.5 Image classification

RF models contain bootstrapped ensembles of decision trees, they can handle independent variables in large numbers while still reporting high classification accuracy (Breiman, 2001). In this research, for the LBP model, the RF models were trained using 15 WV-3 metrics for the image blocks, as predictors to classify the samples as vegetated gap, shaded gap, or forest canopy. For the MLBP model, 5 WV-3 metrics were used to train the RF models. Before running the RF model, the number of decision trees (*ntree*) and the number of predictor variables (*mtry*) randomly selected at each node need to be defined. Each tree applies a randomized bagging approach to retrieve a training data subset and utilize it to cross-validate each tree’s result. This enables the RF models to develop an "out-of-bag" (OOB) accuracy and metrics of input in determining the significance of specific variables in the model (Breiman, 2001). The 10-fold cross-validation (CV) technique which is dependent on the OOB error was used to extract the best performing parameters to be used in the training phase of the RF models. A similar approach was used to optimize the SVM parameters, i.e. cost and gamma. Two indices of variable importance—the mean decrease accuracy (MDA) and Mean Decrease in Gini (MDG) were used in the RF models (Breiman, 2001). The MDA computes the added error rate related to an input variable's exemption from a tree while the MDG calculates the reduction in the forest-wide average in node impurities from splits on a variable (Han *et al*., 2016). The higher the MDA and MDG indices, the more influential the corresponding variables. For a robust selection of features, a combined ranking of both indices was applied in the RF models (Han *et al*., 2016).

The SVM models apply a supervised binary classifier, able to classify linearly inseparable pixels (Vapnik, 2000). The support vectors are the samples closest to the separating hyperplane. The SVM models find support vectors with an optimal margin near the separating hyperplane (Vapnik, 2000). Using kernel functions, SVM models use kernel functions to apply decision boundaries that are not linear and introduce a cost parameter (C) and gamma parameter (γ). A parameter search was carried out to choose the best combination of the C and γ. The two parameters, respectively, determine the penalty of misclassification errors and give the curvature weight of the decision boundary. The more robust radial basis function (RBF) kernel was chosen for this study

The classification of the texture features using the RF and SVM classifiers involved two phases. Firstly, the classifiers were trained using the respective known samples' global histograms and their class labels—the two should be consonant with the classifiers' corresponding pair of classes. Secondly, the unknown samples' global histograms were the input to the RF and SVM classifiers, which search for the class label of the test sample through a comparison of the respective global histograms of the test sample and the training samples. The RF and SVM classification models used 70% and 30% of the ground reference data as train and test data, respectively.

5.2.6 Morphological filtering

In the post-processing stage, it was necessary to transform the classification map so that it has only two classes, i.e. canopy gaps and forest canopy. Therefore, shaded and vegetated gap classes were merged into one class—canopy gaps. Morphological filters were implemented with ArcGIS tools whereby, thin corridors and small spaces amongst forest canopy were removed using the Shrink tool while the Expand tool was applied to enlarge the classified raster gaps. The Focal Statistics tool was used to minimize pixelation and eliminate remaining trees within gaps. The Majority filter retained the most frequently occurring value. Gap polygons with an area <100 m^2 were eliminated using the Select function.

5.2.7 Measures of model performance

Accuracy assessment is used to determine whether pixels identified in the field are classified as they should (Adam *et al*., 2014). The performance of the two machine learning classifiers was evaluated using 30% of the ground truth data. Confusion matrices were computed and averaged over 10 repetitions. The overall accuracy (OA) is computed by dividing the number of correctly classified pixels by the total pixel number—typically expressed as a percentage (Rodriguez-Galiano *et al*., 2012). The producer's accuracy (PA) is the proportion of particular classes on the ground, referred to as such by the classification map (Rodriguez-Galiano *et al*., 2012). The user's accuracy (UA) displays the likelihood that a pixel labelled as a particular feature will be placed on the classified map as such (Mutanga *et al*., 2012). The kappa coefficient (κ) is the difference between the accuracy that was observed and the expected. Therefore, the classification accuracies and kappa statistics computed from error matrices were used to evaluate how the classifiers performed.

5.3 Results

5.3.1 Similarity and separability between training signatures

The G-statistic, a nonparametric test, enabled the avoidance of erroneous assumptions regarding the distribution of features. The G-statistic score is an indication of the possibility that two sample distributions are from the same population. Therefore, a higher score means that the probability of two samples being from the same population is low, and vice versa. The Euclidean distance is 0 if signatures are alike and increase with the spectral distance of signatures. The results of spectral separability indicate that forest canopy and vegetated gaps had the lowest Euclidean Distance between them, 41. The Euclidean distance value between shaded gaps and vegetated gaps was the highest, 96, while that between the forest canopy and shaded gaps was 55. The average separability value was 64. Generally, shaded gaps, vegetated gaps, and tree crowns had good separation

5.3.2 Optimization of Random Forest and Support Vector Machine

The results of the optimization of the RF and SVM parameters for the 6 image blocks are listed in Table 5.3. The composite homogeneity (MHOM) reported the lowest OOB error of 0.097 for the RF model (*mtry* value of 3 and *ntree* value of 1500), while the SVM model achieved a CV error of 0.091 (gamma 0.01, cost 10) for image block D. The composite of homogeneity (MHOM) for image block A also performed failrly well—OOB error of 0.108 (*mtry* and *ntree* values of 3 and 1500, respectively). The SVM model attained a CV error of 0.103 (gamma and cost values were 0.1 and 1000, respectively) for the same image block. Image block F reported an OOB error of 0.106 (*mtry* value of 3 and *ntree* value of 5500) while the SVM model achieved a CV error of 0.108 ((gamma value of 1 and cost value of 100). The composite angular second moment (MASM) recorded highest OOB error of 210 for the RF model (the *mtry* and *ntree* values were 3 and 3500, respectively)—image block A. The SVM model recorded the highest CV error of 214 (the gamma and cost values were 0.1 and 100, respectively) for the composite angular second moment (MASM)—image block E.

Table 5.3: *The random forest and support vector machine model optimization parameters of the Multivariate Local Binary Pattern model.*

Image block ID	Feature	Random Forest			Support Vector Machine			Image block ID	Feature	Random Forest			Support Vector Machine		
		mtry	*ntree*	OOB error	gamma	cost	CV error			*mtry*	*ntree*	OOB error	gamma	cost	CV error
A	MHOM	3	1500	0.108	0.1	1000	0.103	D	MHOM	3	1500	0.097	0.01	10	0.091
	MCON	2	4500	0.128	1	10	0.129		MCON	2	5500	0.121	0.01	1000	0.116
	MENT	2	2500	0.204	0.1	100	0.202		MENT	2	4500	0.165	1	100	0.159
	MASM	3	3500	0.210	0.1	100	0.212		MASM	3	500	0.199	0.01	1000	0.191
	MCOR	2	2500	0.140	0.01	1000	0.132		MCOR	3	3500	0.136	0.1	10	0.127
B	MHOM	2	4500	0.109	0.01	1000	0.110	E	MHOM	2	500	0.101	1	10	0.109
	MCON	2	3500	0.129	1	100	0.122		MCON	2	1500	0.116	0.1	100	0.112
	MENT	3	1500	0.177	0.1	1000	0.174		MENT	3	5500	0.184	1	10	0.179
	MASM	2	2500	0.197	0.01	10	0.190		MASM	2	1500	0.208	0.1	100	0.214
	MCOR	2	5500	0.140	1	10	0.136		MCOR	2	6500	0.149	0.01	100	0.136
C	MHOM	3	3500	0.109	0.01	10	0.102	F	MHOM	3	5500	0.106	1	100	0.108
	MCON	3	4500	0.126	0.1	10	0.119		MCON	3	1500	0.124	0.01	1000	0.119
	MENT	3	6500	0.210	0.01	1000	0.204		MENT	3	4500	0.205	1	10	0.201
	MASM	2	3500	0.200	0.01	1000	0.197		MASM	2	500	0.199	1	10	0.194
	MCOR	2	2500	0.139	1	10	0.134		MCOR	3	1500	0.135	0.1	100	0.138

The average importance score (Figure 5.3) shows that the most important variables are the band 3's $LBP_{c,j}/HOM_c$ and $LBP_{c,j}/CON_c$, band 5's $LBP_{c,j}/HOM_c$ and $LBP_{c,j}/CON_c$, and band 8's $LBP_{c,j}/ENT_c$ and $LBP_{c,j}/ASM_c$. The least performing variables, which showed the lowest average importance scores were the band 3's $LBP_{c,j}/ENT_c$ and $LBP_{c,j}/ASM_c$, and band 5's $LBP_{c,j}/ENT_c$ and $LBP_{c,j}/ASM_c$.

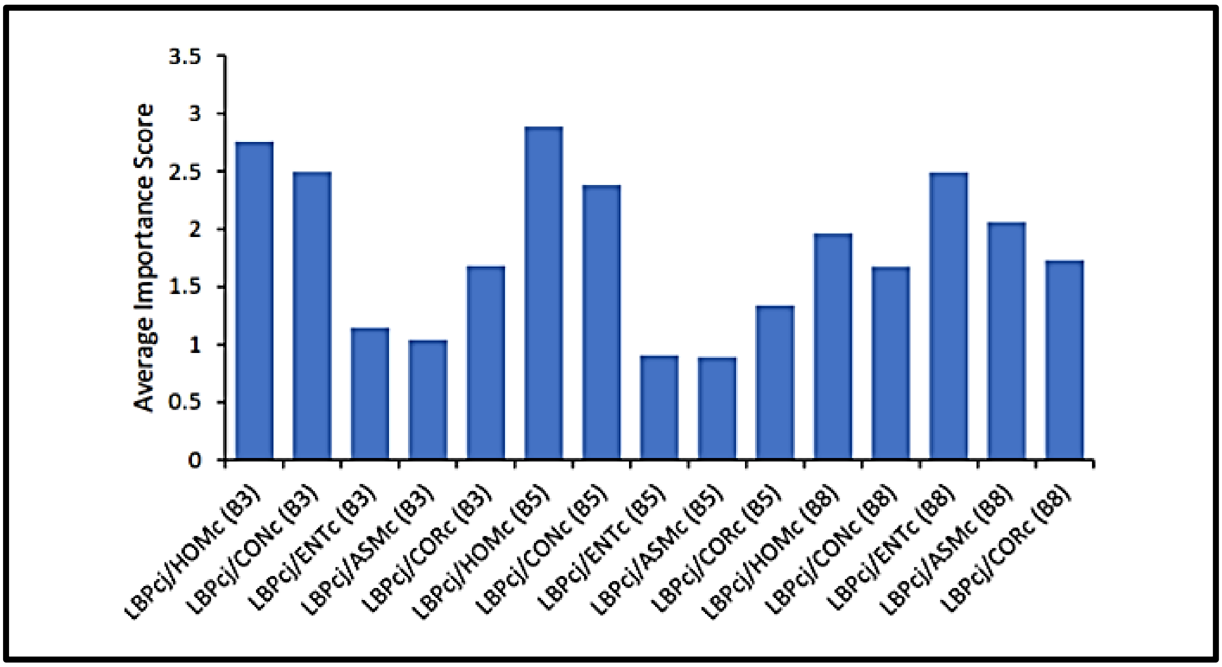


Figure 5.3. *The relative importance of the $LBP_{c,j}$ fused with grey level co-occurrence matrix (GLCM) features in discriminating between vegetated gaps, shaded gaps, and tree crowns by random forest model.*

5.3.3 Model performance

The confusion matrices in Table 5.4 show the results of the RF and SVM classifiers for the MLBP/HOM, MLBP/CON, MLBP/ENT, MLBP/ASM, and MLBP/COR models for image block D, which recorded an average overall accuracy of 86.88 ±5.1% and 89.78 ±3.7% for RF and SVM classifiers, respectively. Image block E's RF classification attained an average overall accuracy of 87.00 ±5.1% while SVM's was 87.28 ±5.8%. The SVM classification of MLBP/CON was the highest, 95.1%. The respective univariate LBP measures provided overall accuracies in the range of 67-75%

Table 5.4: *Confusion matrices for random forest classifier and support vector machine classifier for the composites of homogeneity (MLBP/HOM), contrast (MLBP/CON), entropy (MLBP/ENT), angular second moment (MLBP/ASM), and correlation (MLBP/COR) models for image block D.*

MLBP/HOM

Random Forest	FC	SG	VG	Total	UA (%)	Support Vector Machine	FC	SG	VG	Total	UA (%)
FC	28	1	3	32	88	FC	28	0	2	30	93
SG	0	29	0	29	100	SG	1	30	1	32	94
VG	2	0	27	29	93	VG	1	0	27	28	96
Total	30	30	30	90		Total	30	30	30	90	
PA (%)	93	97	90			PA (%)	93	100	90		
Overall accuracy = 93.3% Kappa = 0.90%						Overall accuracy = 94.4% Kappa = 0.92%					

MLBP/CON

Random Forest	FC	SG	VG	Total	UA (%)	Support Vector Machine	FC	SG	VG	Total	UA (%)
FC	27	1	3	31	90	FC	28	1	2	31	90
SG	1	28	1	30	91	SG	0	29	2	31	94
VG	2	1	26	29	82	VG	2	0	26	28	93
Total	30	30	30	90		Total	30	30	30	90	
PA (%)	90	93	87			PA (%)	93	97	87		
Overall accuracy = 90.0% Kappa = 0.85%						Overall accuracy = 92.2% Kappa = 0.88%					

MLBP/ENT

Random Forest	FC	SG	VG	Total	UA (%)	Support Vector Machine	FC	SG	VG	Total	UA (%)
FC	25	3	4	32	78	FC	25	3	3	31	81
SG	3	26	1	30	87	SG	3	27	1	31	87
VG	2	1	25	28	89	VG	2	0	26	28	93
Total	30	30	30	90		Total	30	30	30	90	
PA (%)	83	87	83			PA (%)	83	90	87		
Overall accuracy = 84.4%; Kappa = 0.77%						Overall accuracy = 86.7%; Kappa = 0.80%					

MLBP/ASM											
Random Forest						Support Vector Machine					
	FC	SG	VG	Total	UA (%)		FC	SG	VG	Total	UA (%)
FC	23	2	3	28	82	FC	25	3	4	32	78
SG	4	25	3	32	78	SG	3	27	1	31	87
VG	3	3	24	30	80	VG	2	0	25	27	93
Total	30	30	30	90		Total	30	30	30	90	
PA (%)	77	83	80			PA (%)	83	90	83		
Overall accuracy = 80.0%			Kappa = 0.70%			Overall accuracy = 85.6%			Kappa = 0.78%		

MLBP/COR											
Random Forest						Support Vector Machine					
	FC	SG	VG	Total	UA (%)		FC	SG	VG	Total	UA (%)
FC	26	3	4	33	85	FC	26	2	3	31	84
SG	2	27	1	30	90	SG	2	28	0	30	93
VG	2	0	25	27	78	VG	2	0	27	29	93
Total	30	30	30	90		Total	30	30	30	90	
PA (%)	87	90	83			PA (%)	87	93	90		
Overall accuracy = 86.7%			Kappa = 0.80%			Overall accuracy = 89.0%			Kappa = 0.84%		

Overall, the MLBP/HOM operator achieved the highest OA, i.e. 93.5%, 92.9%, 92.4%, 91.4%, 91.0%, and 89.8% for image blocks D, E, F, A, C, and B, respectively, for RF classifier and 94.3%, 93.1%, 92.6%, 91.8%, 91.2% and 90.8% for image blocks D, C, A, F, E, and B, respectively, for SVM classifier. Generally, the UA ranged between 78-100% for both RF and SVM classifiers. The PA was in the range of 77-100% for the RF model, and 83-100% for the SVM model. The overall results of the classification of the six image blocks are summarized in Table 5.5. For each classifier, the table reports the average classification accuracy in the form $\mu \pm \sigma$, where μ and σ are, respectively, the mean and standard deviation of the OA.

__Table 5.5:__ The overall accuracy results of the six image blocks.

	Texture descriptor	Image block A	Image block B	Image block C	Image block D	Image block E	Image block F
RF classifier	MLBP/HOM	91.4	89.8	91.0	93.3	92.9	92.4
	MLBP/CON	83.9	83.2	85.3	90.0	92.3	86.7
	MLBP/ENT	82.6	83.9	82.4	84.4	83.4	82.6
	MLBP/ASM	81.1	81.4	81.1	80.0	82.9	80.0
	MLBP/COR	84.4	84.4	84.5	86.7	83.5	85.5
	Average OA	84.7 ±4.0	84.5 ±3.2	84.9 ±3.8	86.9 ±5.1	87.0 ±5.1	85.4 ±4.7
SVM classifier	MLBP/HOM	92.6	90.8	93.1	94.4	91.2	91.8
	MLBP/CON	83.2	88.1	90.2	92.2	95.1	90.2
	MLBP/ENT	82.8	85.7	80.3	86.7	83.7	82.8
	MLBP/ASM	81.1	85.9	81.4	85.6	80.9	81.3
	MLBP/COR	87.5	85.5	86.5	90.0	85.5	84.5
	Average OA	85.4 ±4.6	87.2 ±2.3	86.3 ±5.5	89.9 ±3.7	87.3 ±5.8	86.1 ±4.6

5.3.4 Image classification

A subset of the classification results using MLBP/MHOM for image block D, whose RF and SVM model optimization parameters recorded some of the lowest OOB and CV errors, respectively, is shown in Figure 5.4. Even based on visual interpretation, the classes are mapped correctly. The Multivariate local binary pattern (MLBP) model distinguishes classes very well because it assigns distinct and precise pattern codes to show patterns. The boundaries of extracted canopy gaps are overlaid on the ground truth canopy gap areas which are shown in red.

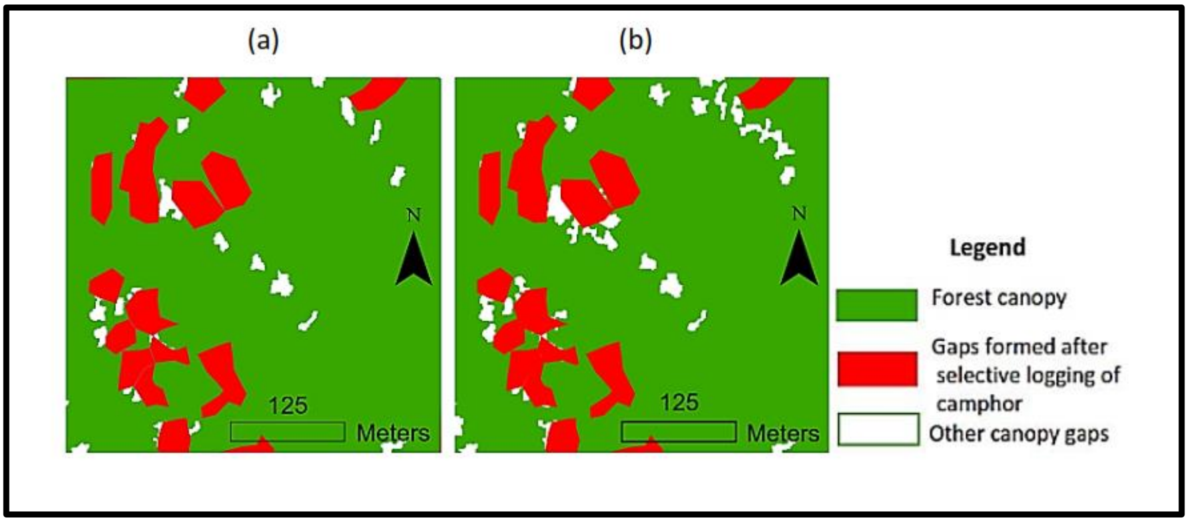


Figure 5.4. *Subset classification results of image block D based on the Multivariate texture-based (853-RGB) classification of canopy gaps from WorldView-3 image based on MLBP*$_c$ *and MHOM distribution—(a) random forest, and (b) support vector machine.*

5.4 Discussion

This study was aimed to find whether GLCM - and MLBP-based rotation-invariant feature descriptors derived from WorldView-3 imagery may be extended and used for canopy gap classification in a tropical submontane forest using remotely sensed image. The study applied an LBP model fused with a GLCM model, whereby the rotation invariant LBP operator was used to get the LBP images of subsets of images extracted from a WV-3 scene covering the study area. Then five GLCM measures of the LBP images were calculated to describe the image texture features. The LBP texture measure was applied to colour images by applying a multivariate texture model, i.e. the multivariate local binary pattern (MLBP) operator. Due to the robustness of the models, the classes were found to be separable. For the LBP model, a uniformity measure was applied to show the uniformity of the neighbourhood's pixel values. According to Ojala *et al*. (2002), in a textured image >90% of patterns are uniform.

A circular neighbourhood set comprising 8 neighbouring pixels and a radius of 1 was used, i.e. the values for P and R were 8 and 1, respectively. Large P and R values are appropriate for describing large-scale textures, and vice versa (Burnett *et al*., 2019). The circular symmetrical neighbour set approach is more robust and delivers more accurate results (Bianconi and Fernández, 2014). According to Clausi (2002), different combinations of values of P and R in neighbourhood sets might offer meaningful texture descriptions. Larger window sizes enable classifiers to extract rich textural information from a pixel, which could improve accuracy—however, a larger window size might reduce sensitivity to class edges, and eventually smooth over the image (Clausi, 2002).

The accuracy results values reported here show that good classification results were obtained. GLCM features perform better with ≤10 classes—therefore, they can outdo more powerful methods (Bianconi and Fernández, 2014). This research used just five co-occurrence descriptors, although Di Ruberto *et al*. (2015) state that a higher number of co-occurrence features could obtain excellent results. The respective univariate LBP measures provided classification accuracies in the range of 67-75%. The Multivariate LBP models gave higher classification accuracies (Table 5.5). Although the composite homogrneity (MHOM) recorded the lowest OOB and CV errors, the composite contrast (MCON) with a CV error of 0.112 (gamma and cost values of 0.1 and 100, respectively) for the SVM model outperformed all the other models to record the highest classification accuracy of 95.1%—image block E. The lowest classification accuracy (80.0%) was recorded by the RF's composite of angular second moment (MASM), for image block F. The MSAM performed poorly in all RF and SVM models. Overall, the overall accuracies (OAs) reported could have even been higher were it not for confusion between the classes.

The fusing of textural and spectral information from three WV-3 bands performed better than their basic models. The MLBP classification was applied to three bands, i.e. bands 8 (Near Infrared 2), 5 (Red), and 3 (Green) of the WV-3 image of the study area. Future research will aim to modify the model to include more than three bands, even extending it for hyperspectral data. This will explore the contribution of separate colour bands in texture analysis. It will also assist in investigating novel combinations of colour and texture for classification. Although complexity and computational demands would increase, adding more bands might not significantly increase the amount of textural information (Luceer *et al*., 2005). However, the net benefit would be increased accuracy in classification which can be worth it.

5.5 Conclusion

The accurate mapping of canopy gaps is of great importance to forest managers because it lays the groundwork for additional research and management applications, such as the comprehension of species diversity and distribution, and ecosystem dynamics. This information can be incorporated into current and future forest management activities. The spatial pattern of canopy gaps may be misrepresented if reliable methods for accurate canopy gap mapping are not used. There are many different methods that have been developed based on how texture characterization is done. Even though there are numerous effective texture analysis techniques reported in the literature, this work used a texture-spectral analysis framework to improve the GLCM model, which was fused with the MLBP model to detect SL. Even after almost 50 years, Haralick's method remains a powerful model for texture analysis.

CHAPTER SIX

An assessment of Kenya's forest policy and law on participatory forest management for sustainable forest management

This chapter is based on:

Jackson, C.M., and Adam, E. (in review). An assessment of forest policy and law on participatory forest management for sustainable forest management in Kenya. *Restoration Ecology*.

Abstract

State forest management is increasingly giving way to multi-stakeholder participation structures in many nations. When participatory forest management (PFM) is planned and put into practice in a way that makes decisions and processes transparent and local officials answerable to the public, it improves the transparency of forestry governance. By allowing the establishment of community forest associations (CFAs) to enter into forest management agreements with Kenya Forest Service (KFS), the country's statutory forest agency, Kenya’s Forest Act of 2005 solidified PFM. As a result, CFAs now have obligations and rights related to managing forests and sharing in their benefits. The goal of this study is to assess Kenya's forest policy and law on PFM and to evaluate how well institutional arrangements, legal frameworks, and policy changes aimed at promoting the shift to sustainable forest management (SFM) have worked. A literature review on MKFR’s research publications, project reports, and management plans was carried out. Previous and present forestry-related Memorandums of Understanding (MoUs), bills, policies, agreements, and legislation, were reviewed. Also information on use trends, and past and present management practices of Mt. Kenya Forest Reserve (MKFR) was gathered. The information was compared with data collected through the key interviews, the questionnaires, and what was observed in the field to find gaps and elucidate ambiguous issues. Even though laws and legal frameworks have improved, Kenyan forest governance changes have not kept up with this effort. In Mt. Kenya Forest Reserve (MKFR), there have been reports of illegal logging, among other concerns. This study found that there is a need for more inclusive and diverse regulations, even though many people support increased policing and regulation.

Keywords: Participatory Forest Management. Sustainable Forest Management. Community Forest Associations. Policy. Legal framework. Stakeholder. Transparency.

6.1 Introduction

Although forests sustain biotic and abiotic elements of the environment in various ways, sustainable management of forests is in jeopardy (Long, 2013). Forest management received more attention globally after the Food and Agricultural Organization (FAO) established its forestry department in 1945 (Galabuzi *et al*., 2015). However, it was not until 1980 that tropical deforestation was widely acknowledged as a concern when their preservation became a politically controversial issue (SIFI, 2010). Global treaties and conventions on climate change and natural resource management have promoted the advocacy for international forest policies (SIFI, 2010). Prevention of trade threatening the endangered biotic resources was given priority by the 1973 Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). In both the 1987 Brundtland Commission report and the 1992 Rio de Janeiro UN Conference on Environment and Development, the need to preserve ecosystems, and species, and to lessen the factors that contribute to deforestation were outlined as important concerns (FAO, 1999). Consequently, international forest policy has developed as a result to safeguard the planet's vulnerable forest ecosystems. The 1992 Convention on Climate Change (CCC) sought to stabilize greenhouse gas concentrations in the atmosphere, whereas the 1992 Convention on Biological Diversity (CBD) sought to protect forest, marine, and coastal area ecosystems (Galabuzi *et al*., 2015). The 1994 International Tropical Timber Agreement put a strong emphasis on sustainable forest management (SFM) (Galabuzi *et al*., 2015). The Kyoto Protocol of 1997, which extended the 1992 United Nations Framework Convention on Climate Change, commits state parties to reduce greenhouse gas emissions and makes explicit reference to land-use change and forestry (FAO, 1999). Although there have been changes to the world's forest cover, the tropical areas have been the most severely affected, resulting in significantly reduction in goods and services they provide (FAO, 2022). About 1 150 million ha of tropical rainforest had been destroyed by 1990, with about 2.3 million ha lost as a result of fragmentation, habitat loss, logging, and fires (Mayaux *et al*. 2005). Tropical forest degradation is difficult to quantify but is likely increasing (FAO, 2022). A large body of existing literature holds tropical forest loss on the effectiveness/ineffectiveness of restrictions on the use of forest resources (Robinson and Lokina, 2011).

Since Rio de Janeiro in 1992, emphasis has been placed on the value of biodiversity, particularly in protected areas (Galabuzi *et al*., 2015). Undermining efforts to achieve SFM are the high rates of forest degradation and deforestation, and the lack of forests and forestry to enhance people’s livelihoods (Mbuvi and Kung’u, 2020). Community participation in the management of state-owned forests is tailored toward SFM (Mbuvi and Kung'u, 2020). Many countries in Africa and Asia have altered their national laws and policies to involve rural communities in the management and utilisation of natural forests and woodlands—through some form of Participatory Forest Management (PFM) (Mbuvi and Kung'u, 2020). Only about 41% of the 2.2 billion ha that is expected to become more sustainably managed by 2050, according to an analysis of global progress toward SFM by MacDicken *et al*. (2015) and the 2015 Global Forest Resources Assessment (FRA) (FAO, 2015), have the prerequisites for SFM in place. In 2015, Agenda 2030 for Sustainable Development and its 17 Sustainable Development Goals (SDGs) was adopted—the SDGs build on the Millennium Development Goals (MDGs) (Katila *et al*., 2019). The SDGs incorporated a broader development policy agenda that addressed numerous aspects of economic, social, and environmental sustainability (Katila *et al*., 2019). Despite the importance of forests in achieving the SDGs, they are only mentioned in SDG 15 (Life on Land) which focuses on the protection, restoration and sustainable use of terrestrial ecosystems and halting the loss of biodiversity, and SDG 6 (Clean Water and Sanitation) which advocates for the protection and restoration of forests in one of its targets (Katila *et al*., 2019). The interconnectedness of the SDGs and targets will inevitably have an impact on forests and forest-related livelihoods—as well as the possibility of achieving forest-specific targets (Katila *et al*., 2019).

Kenyan context

Kenyan forest management transitioned from Command and Control (C-and-C) to community involvement in 1891 (Mbuvi and Kung'u, 2020). An elder council enforced a set of customary laws and ordinances that controlled the use of the forest and its resources (Mbuvi and Kung'u, 2020). This practice is observed to date in some historical and indigenous people's conservation practices, e.g. the Mijikenda people, who zoned the *Kayas* for only religious and cultural rites (Mbuvi and Kung'u, 2020). The primary objective of forestry operations between 1902 and 1908 was the preservation of the existing forests (Logie and Jones, 1968). "The Forests of East Africa," by D.E. Hutchins in 1909, guided the Forest Department's activities for many years (Mugo *et al*., 2010; Mbuvi and Kung'u, 2020). The government established the Forest Department in 1942, led by the Chief Conservator of Forests. After the Second World War (WWII), the government established a forestry subcommittee (within a development committee) that recommended the protection of forests and catchments. Until the 1957 forest policy (White Paper no. 85) was published and reviewed in 1968 sessional paper no. 1, the subcommittee's report directed the affairs of the Forest Department (Mugo *et al*., 2010; Ongugo *et al*. 2014; Mbuvi and Kung'u, 2020). However, the report failed to adequately reflect the function, obligations, and rights of communities that surround forests or those who live within them. The Sessional Paper No. 10 of 1965 (GoK, 1965b) was the first piece of policy to acknowledge the value of protecting natural resources for future generations and to express concern for environmental quality (Ongugo *et al*. 2014).

In the past two decades, Kenya's forestry sector has undergone four significant legislative and policy reforms, changing the laws, regulations, and institutions with the assistance of various capacity-building initiatives to participatory governance of resources (Mbuvi and Kung'u, 2020):

- traditional community-based forest management to Forest Act (CAP 385);
- repeal of Forest Act (CAP 385) to Forests Act, 2005;
- repeal of Forest Act, 2005 to Forest Conservation and Management (FCM) Act, 2016 (GoK, 1965a; GoK, 2016b); and
- Forest sector decentralization introduced through PFM.

The most recent update to the 1957 policy, which was developed to focus on the conservation and management of forest resources on public lands, is the 2020 draft forest policy (GoK, 2020; Mbuvi and Kungu, 2020).

To promote community-based forest management, forest tenure reforms in the Forests Act (2005) embraced PFM in gazetted forest reserves by recommending community involvement and private investment. Article 45(1) states that, "A member of a forest community may, together with other members or persons resident in the same area, register a community forest association under the Societies Act," (Forests Act, 2005). Article 45(2) stipulates that "An association registered under subsection (1) may apply to the Director for Permission to participate in the conservation and management of a state forest or local authority forest in accordance with the provisions of this Act," (Forests Act, 2005). The main objective was to increase forest biodiversity and alleviate poverty among forest-adjacent communities. The paradigm shift has resulted in a concomitant change in institutional and organizational arrangements, e.g. the formation of the Kenya Forest Service (KFS), and the Community Forest Associations (CFAs) (KFS and MENR, 2007). The 2010 constitution supports the devolution of forest resources—article 69(1)(d), "The state shall encourage public participation in the management, protection and conservation of the environment," (GoK, 2010). Therefore, it compelled the government to make sure that communities benefit from natural resources and that they get informed about the management of the resources (GoK, 2010). Decentralization in forest governance has been perceived as better resource management because it supports participation in the forest surrounding communities, accountability, and community empowerment (Monditoka, 2011). According to the 2020 draft forest policy (GoK, 2020), forest governance has experienced poor performance, therefore, more effort is required to conform to the 2010 Constitution and the national development agenda (GoK, 2010). Forest governance is the process of developing, articulating, administering, and putting into practice policies, institutions, laws, regulations, guidelines, and norms concerning ownership, access, control, obligations and rights, and SFM practices (KFS and MENR, 2007). These elements are the subject of effective forest policies. Otherwise, forest policies may result in a lack of transparency, accountability, empowerment, subsidiarity, sustainability, equity and justice, participation of the public in decision-making, and poor coordination among various governmental levels and sectors. The draft forest policy recognizes the opportunities of SFM and ensures the involvement of all relevant parties (Mbuvi and Kung'u, 2020). The National Forest Programme (GoK, 2016a) acknowledges that decentralization can provide solutions to the majority of Kenya's forest management issues. Decentralization is defined as any action whereby the central government transfers power to individuals or groups operating below it in the political, administrative, and territorial hierarchies (Adam and Eltayeb, 2016).

Poor forest governance over time has led to significant forest degradation and deforestation. At independence in 1963, Kenya’s forest cover was approximately 11%, and about 2% in the late 1990s (IEA, 2010). From 2001 to 2021, Kenya lost 368 000 ha of tree cover, equivalent to 11% decrease in tree cover since 2000, and 180 000 000 t of CO_2 emissions (GFW, 2022). The main drivers of deforestation in Kenya are illegal logging, burning of charcoal, harvesting non-timber/wood products, and degazetting and excision of forest land. Based on annual reports it is estimated that illegal charcoal burning makes up to 50%, while logging accounts for 15% and other illegal activities account for 35%. Other unlawful actions include unlawful farming, unpermitted grazing, forest fires, and unlawful hunting and honey collection (KFS and MENR, 2007). Around 80% of Kenya's total land area is made up of arid or semi-arid lands (ASALs), and only 7.4% of Kenya is forested, falling short of the United Nations' recommended minimum forest cover of 10% (NEMA, 2022). Kenya's growing population puts a strain on its natural resources, particularly the rainforest cover, which is mostly montane forests in scattered patches that are deteriorating (NEMA, 2010).

Five key factors initiated PFM in Kenya (Mbuvi and Kung’u, 2020):

- the forest conditions in Kenya were worsening at a very high rate resulting in inadequate resources;

- the unresponsive Command and Control (C-and-C) approach used by the government;
- pressure from the local communities and development partners/donors and the global community in general;
- Kenya's on-going macro-economic and institutional reforms; and
- the Kyoto protocol pushing governments to ensure communities are involved on the management of natural resources.

The shift in institutional frameworks in the governance of forests, through PFM, was intended to achieve SFM, improved livelihoods, low rates of deforestation, and equitable sharing of benefits (GoK, 2020). However, the occurrence of illicit human activities in Mt. Kenya Forest Reserve (MKFR) amid enhanced forest restoration programs may be an indication that the policy interventions may be inadequate, inappropriately designed, and implemented (CIPFA, 2012). To achieve good governance in the forest sector, there is a need to identify concrete elements within the policy and institutional framework by systematically analysing the current situation and directing efforts toward improving the systems (CIPFA, 2012). Therefore, the purpose of this study was to evaluate Kenya's forest policy and law on PFM to determine the effectiveness of changes made to institutional structures, laws, and policies to support the transition to SFM.

6.2 Methods

6.2.1 Sample design

First, respondents who were deemed to have expert knowledge on a wide range of topics—policy, socio-economic issues, and forest management—were chosen using the purposive sampling technique (Mbuvi and Kungu, 2020). They were in a better position to provide information on the effectiveness of Participatory Forest Management (PFM) in Kenya due to their interactions with the forest and/or administrative responsibilities. A total of 15 respondents from Kenya Forest Service (KFS), Kenya Wildlife Service (KWS), and the National Environment Management Authority (NEMA) were interviewed.

Second, households living adjacent to Chuka Forest (Figure 1.1) were asked to complete a structured questionnaire (Appendix 1). Ethics clearance for the questionnaires was obtained from the University of the Witwatersrand—the Human Research Ethics Committee (non-medical). For large respondent populations and when detailed information is requested, structured questionnaires are generally regarded as the best option. Equation (6.1) was used to calculate the sample size (Mbuvi and Kungu, 2020):

$$n = \frac{z^2 P(1-P)}{d^2} \tag{6.1}$$

where z represents the standard score with a 0.05 margin of error and a 95% confidence level, i.e. 1.96. P denotes the occurrence level of the phenomenon under investigation study, which is equated to 0.5 when the occurrence level is not known, and d is the study's chosen error margin at a 95% confidence level, i.e. 0.05.

Therefore,

$$n = \frac{1.96^2 \times 0.5\ (1-0.5)}{0.05^2} \qquad n = \frac{1.96^2 \times 0.5\ (0.5)}{0.05^2} \quad = 384$$

The households are those of members of the Community Forest User Groups (CFUGs) that are located up to a distance of 5 kilometres from the forest's edge, and those targeted were the heads of households or adults (Cheruiyot *et al*., 2019). The multistage sampling technique was applied to select respondents—began with selecting sub-locations within the six locations in the study area, followed by a random selection of homesteads, and finally households. By doing this, a reasonable representation was guaranteed, and a favourable impression of the state of forest conservation and community livelihood was captured (Mbuvi and Kungu, 2020). About 10% of the sample was used to conduct a pilot study to test the research instruments (Cooper and Schilder, 2007). Therefore, questionnaires were administered to 38 (Appendix 1) households while 1 KFS officer, 1 KWS officer, and 1 NEMA official (Appendix 2) and 1 KFS forest guard (Appendix 3) were selected to respond to the interview guide.

A literature review on MKFR's research publications, project reports, and management plans was carried out. Previous and present forestry-related Memorandums of Understanding (MoUs), bills, policies, agreements, and legislation, were reviewed. Information on the condition of Mt. Kenya Forest Reserve (MKFR) was gathered (Appendix 4)—also information on use trends, and past and present management practices (Galabuzi *et al*., 2015). The information was compared with data collected through the key interviews, the questionnaires, and what was observed in the field, to find gaps and elucidate ambiguous issues (Martin, 2004).

6.2.2 Data analysis

The foundation of the qualitative data analysis was a review of conversations and written texts, then responses were organized into categories (Cheruiyot *et al*., 2019). Non-parametric descriptive data were statistically analysed, presented in tables, and interpreted. The strength of the central pattern was measured using the standard deviation of the mean. Respondents gave their level of agreement/disagreement with the statements on a Likert scale of 1-5, where 1 means strongly disagree, 2 means disagree, 3 means neutral, 4 means agree, and 5 means strongly agree (Cheruiyot *et al*., 2019). Carifio and Rocco (2007) established the following weighting criteria for Likert-type data responses: Strong Disagree (SDis) 1 < SDis < 1.8; Disagree (Dis) 1.8 < Dis <2.6; Neutral (N) 2.6 < N < 3.4; Agree (A) 3.4 < A < 4.2; and Strongly Agree (SA) 4.2 < SA < 5.0.

6.3 Results

6.3.1 Characterization of respondents

Only 311 (81%) of the 384 questionnaires distributed were completed and returned. About 60% of the policymakers who were interviewed were over 40 and had more than 15 years of forestry industry experience. The respondents were predominantly male (73%) with only four female respondents (27%). All of the areas surrounding the Chuka forest were represented in the socioeconomic survey's participant pool. The 311 questionnaires returned were from respondents residing in 6 locations. Out of this, 43% of those polled were female, while 57% were male. Most of the respondents were in the 30- to 75-year age range, with the age groups 46-60 comprising the largest proportion at 39% of the respondents while respondents below 30 years comprised the smallest proportion (7%) of the respondents (Figure 6.1).

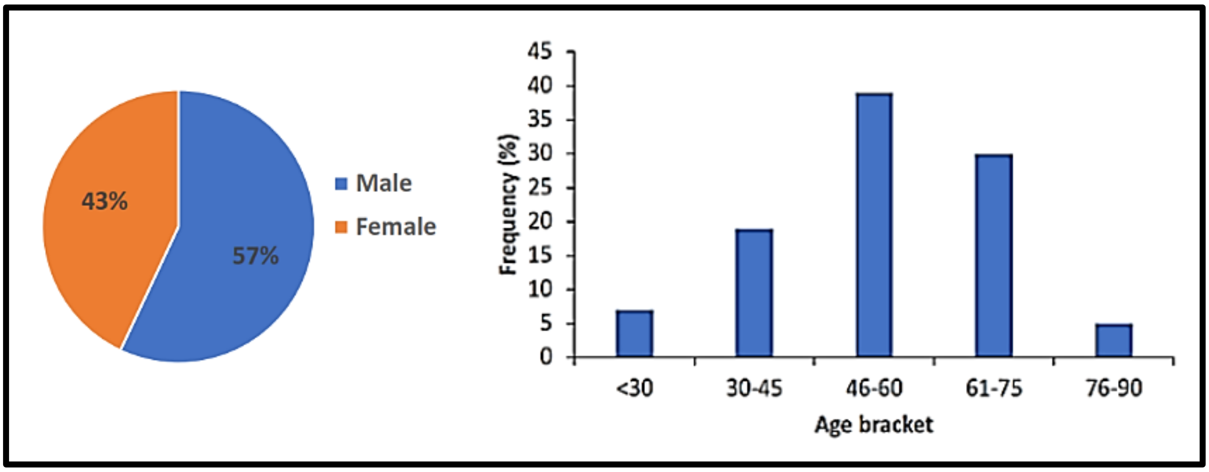


Figure 6.1. *Gender and age distribution among respondents from the questionnaire for the Community Forest Associations around Chuka Forest, Tharaka Nithi County, Kenya.*

However, participation in the survey varied from location to location with Mwonge location recording the highest number of respondents accounting for 21% followed closely by

Kiang'ondu at 20% of all respondents. Thuita, Magumoni, and Mugwe locations each recorded 15% of the respondents (Figure 6.2).

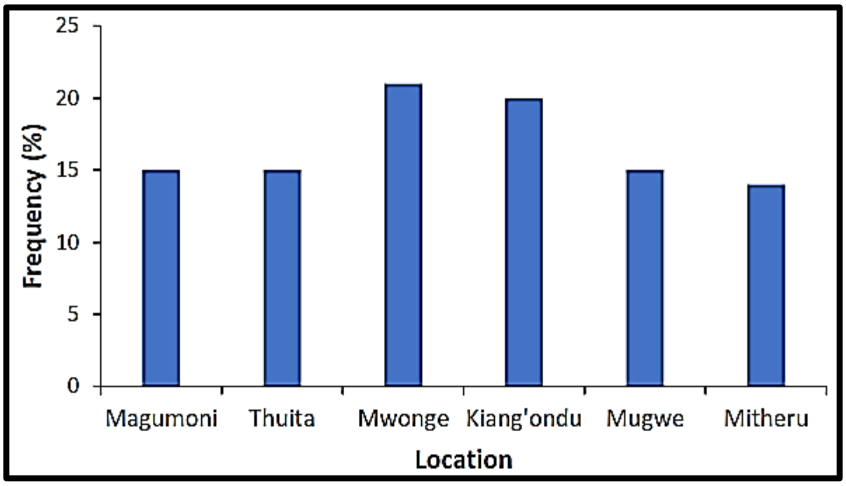


Figure 6.2. *Participation of the community forest association members around Chuka forest, Chuka, Tharaka Nithi County.*

The majority (81%) of the households were headed by men while small proportions (19%) of households were headed by women. Education levels varied among the respondents, i.e. 9% did not have any formal education, with the majority (41%) only having primary education, followed by secondary education at 32% and colleges/universities at 18%. The majority 62% of the respondents interviewed got their livelihood from agriculture, businessmen at 8%, wage labourers at 2%, and civil servants 3%. Students comprised 25% of the respondents. Therefore, the main economic activities were farming and livestock keeping. The following were identified as the current activities and benefits to the community from the forest—fodder grass, fuel wood, herbal medicines, building poles/other materials, water, forest soil, beekeeping, salt fountain, timber, water abstraction, fencing poles, and wild vegetables.

The survey showed that 77% of respondents earn between 1,000 and 5,000 KES (Kenyan Shilling) per month while 12% earn between 5,001 and 10,000 KES, and 7% earn more than 15,000 KES per month—1 USD, 121 KES. Only 1% earn less than 1000 KES (Figure 6.3). The low-income levels recorded during the survey may be due to the respondents deliberately underestimating their earnings.

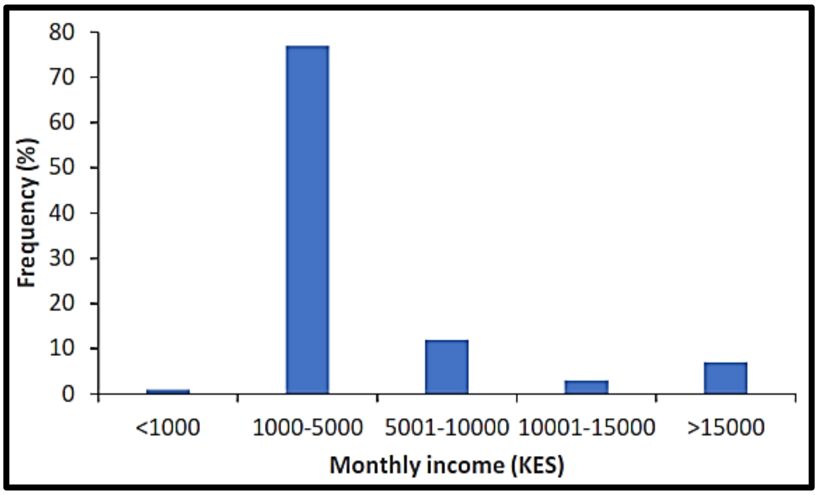


Figure 6.3. *Overall monthly income of the community forest association members around Chuka forest, Chuka, Tharaka Nithi.*

Firewood is the most common source of energy with 81% of households using it as the primary source of energy followed by charcoal at 12% and only 7% using electricity as the source of energy. Firewood is used for entirely cooking while electricity is used for lighting. The major water source type for household use among the respondents is a river which accounted for 53% of all respondents followed by wells at 16% and tap water at 15%. Other water sources recorded include springs at 13% and dam at 3%.

6.3.2 Stakeholders analysis

Mt. Kenya Forest Reserve (MKFR) and adjacent villages have a wide array of stakeholders who are involved in development activities that address environment and livelihood issues. The stakeholders consist of a wide cross-section of government agencies/public institutions (e.g. Ministry of Agriculture, Livestock, Fisheries and Irrigation, Ministry of Environment and Forestry, etc.). Others are non-governmental organizations (e.g. Farming Systems Kenya, Green Belt Movement, Nature Kenya, Rhino Ark, International Council for Research in Agroforestry, International Small Group and Tree Planting Program, World Wildlife Fund, African Wildlife Fund, etc.). Also development partners/donors (e.g., European Union, Global Environmental Facility, etc.), and Educational institutions, Faith-based organizations, and business groups (e.g., Saw millers, tea factories, hotels, mineral water companies, horticulture farmers, tour operators, etc.). Table 6.1 presents a summary of stakeholders and their roles and responsibilities in MKFR management.

Table 6.1: *Analysis of main stakeholders and their roles and responsibilities in Mt. Kenya Forest Reserve (MKFR).*

Stakeholder	Roles and responsibilities
Kenya Forest Service	Management of the MKFR ecosystem and farmlands, and introduction of plantation forests; revenue collection from the sale of forest products; forest product regulation and licensing; restoration of forests; collaboration with stakeholders on the management of the ecosystem and security; and putting out and preventing forest fires.
Kenya Wildlife Service	Wildlife preservation and protection in MKFR; managing wildlife-human conflict; security and ecosystem management issues; wetland convention and CITES steward.
Nyayo Tea Zones Dev. Corporation	Liaise with KFS in environment conservation; check forest encroachment.
Ministry of Agriculture	suggestions on food production under PELIS (Plantation Establishment and Livelihood Improvement Scheme) program; Putting conservation measures for soil and water into action; agroforestry, etc.; advice on more suitable and improved food/livestock varieties.
National Museums of Kenya	Preparation of conservation education programs and presentations; taxonomy and herbarium; education about locals' cultural customs; investigation into the historical use of cultural sites.
Forest Adjacent Communities	Security patrols; create income-generating ventures; indigenous conservation knowledge dissemination.
Tourism Stakeholders	Promoting ecotourism and educating locals about the advantages of protecting ecosystems; promote and pushing for MKFR ecosystem preservation; providing local communities with employment opportunities; establish tourism facilities e.g., hotels, campsites, etc.
Water Department	Assessment and monitoring of quality and levels of water; development, control, apportion, conservation, and regulation of water resources; empowering local communities to manage water resources.
Meteorological dept.	Provide climatic data
Education institution	Research; education
Ministry of Lands	Legal framework, policy on land use

The communities adjacent to MKFR are organized into groups (Community Forest User Groups - CFUGs) that have clearly defined responsibilities and roles in the management and conservation of MKFR. About 40 forest user groups were identified around Chuka Forest alone.

6.3.3 Participation of CFAs in conservation and management of MKFR

Forests are better managed with PFM, thus local communities play a significant role in managing forests because the community knows those who commit forest malpractice. This study sought to examine the participation of CFAs in the conservation and management of MKFR. Some of the activities of CFAs were the establishment of commercial tree nurseries, tree planting, beekeeping, creation of ecotourism opportunities, protection of MKFR through collaborative security patrols, implementing PELIS (Plantation Establishment and Livelihood Improvement Scheme), etc. Respondents were asked to express their level of agreement with various statements. The results are presented in Table 6.2.

Table 6.2: *Participation of community forest associations (CFAs) in conservation and management of MKFR; SDis = strongly disagree, Dis = disagree, N = neutral, A = agree, SA = strongly agree, SD = standard deviation, f = frequency, (%) = percentage.*

Statement	**SDis f (%)**	**Dis f (%)**	**N f (%)**	**A f (%)**	**SA f (%)**	**Mean**	**SD**
C1. Products from MKFR are essential for the subsistence of the forest surrounding communities	17 (5.5)	19 (6.1)	20 (6.4)	127 (40.8)	128 (41.2)	4.06	1.10
C2. Villagers extract products from MKFR	21 (6.8)	25 (8.0)	20 (6.4)	108 (34.7)	137 (44.1)	4.01	1.20
C3. The forest offers protection to a variety of wildlife	19 (6.1)	23 (7.4)	29 (9.3)	100 (32.2)	140 (45.0)	4.03	1.18
C4. Sustainable exploitation of the forest resources is good	19 (6.1)	20 (6.4)	21 (6.8)	108 (34.7)	143 (46.0)	4.08	1.15
C5. You report the occurrence of illegal activities in MKFR	29 (9.3)	50 (16.1)	41 (13.2)	97 (31.2)	94 (30.2)	3.60	1.32
C6. For ecological balance the forest requires protection	21 (6.8)	23 (7.3)	31 (10.0)	121 (38.9)	115 (37.0)	3.92	1.17
C7. Concerted conservation effort is required to save the forest	18 (5.8)	11 (3.5)	17 (5.5)	126 (40.5)	139 (44.7)	4.10	1.07
C8. People work relentlessly to preserve the forest	20 (6.4)	19 (6.1)	22 (7.1)	114 (36.7)	136 (43.7)	4.05	1.15
C9. Your active role in conservation is important	19 (6.1)	22 (7.1)	20 (6.4)	127 (40.8)	123 (39.6)	4.01	1.14
C10. You are willing to sacrifice more time toward conservation activities in MKFR	22 (7.1)	15 (4.8)	26 (8.4)	122 (39.2)	126 (40.5)	4.01	1.15
Overall mean and SD						3.99	1.16

Statement C1 of Table 6.2 indicates that, of the 311 respondents, 82% agreed that MKFR products are essential to their ability to survive, while 6.4% of the respondents were undecided about the statement. The remaining 11.6% of respondents disagreed. The statement's mean score of 4.06 outperformed the table's mean score of 3.99. Thus, the perception of the forest surrounding communities that MKFR products are essential for their subsistence has had a positive impact on the conservation of the forest reserve. Because the local community sees the forest as a source of livelihood, they recognize the importance of participating in forest conservation activities. In response to statement C2, ‘Villagers extract products from MKFR,’ 78.8% of respondents agreed, while 14.8% disagreed. Only 6.4 percent were unconcerned. The statement's 4.01 mean score versus a lower overall mean indicated that the local community is involved in the conservation of products such as honey, fodder grass, herbal plants, and wild vegetables. Only 13.5% of respondents disagreed with statement C3, 9.3% neither agreed nor disagree, and 77.2% agreed that the forest reserve harbor a variety of wildlife. The availability of wild animals in MKFR is due to the conservation efforts of the CFAs, which have sensitized members against wild animal poaching, as the mean score of 4.03 was higher than the mean score in the table. Statement C4, ‘Sustainable exploitation of the forest resources is good,’ received 6.8% of neutral responses, 80.7% agreed, and approximately 12.5% disagreed. The 4.08 mean score was higher than the overall mean score, indicating that sustainable exploitation of MKFR resources complements the CFA members' conservation and preservation efforts. Only 61.4%t of respondents report illegal activities occurring in MKFR (statement C5). Examples of these illegal activities include illegal logging, marijuana cultivation, unpermitted grazing, charcoal burning, and the illegal production and collection of honey, which has occasionally resulted in forest fires, animal poaching, etc. The statement's mean score of 3.60 was lower than the table's mean, indicating that failure by CFA members to report illegal activities in MKFR encourages more illegal activities, undermining forest conservation efforts. This, in turn, has an impact on the goods and services flowing to forest-adjacent communities and the country as a whole. ‘For ecological balance the forest requires protection,' says Statement C6. The statement was supported by 75.9% of respondents. 14.1% disagreed, while 10.0% were neutral. The average score was 3.92, which was a little lower than the average on the table. This implies that CFA members' failure to adequately participate in MKFR conservation and management will have an impact on its optimal ecological balance. As a result, demand for goods and services will outstrip supply. Statements C7, C8, C9, and C10 all recorded means that were higher than the table mean. This suggests that a concerted effort by CFAs surrounding MKFR to protect the forest reserve has resulted in a net of a balanced ecosystem with increased biodiversity. The benefits obtained from using harvested forest products may influence the CFA members' acceptance of the significance of their active role in the conservation of MKFR and their willingness to put in more effort to do so. The result is a well-conserved and managed forest ecosystem.

6.3.4 Participation of CFAs in the planning, implementation, evaluation, and monitoring of forest conservation in MKFR

The four stages of forest conservation and management are planning, implementation, evaluation, and monitoring. Planning enables the designing of what ought to be done, i.e. aiming at the crucial outcomes. Monitoring and evaluation assess progress toward expected results, highlight implementation challenges, and identify unintended consequences of forest conservation. Bottlenecks may be caused by a lack of stakeholder ownership and participation. The study sought to evaluate how CFA members participate in forest conservation planning, implementation, evaluation, and monitoring in MKFR. Table 6.3 summarizes the findings.

Table 6.3: *Participation of community forest association (CFA) members based on the four stages of forest conservation and management; SDis = strongly disagree, Dis = disagree, N = neutral, A = agree, SA = strongly agree, SD = standard deviation, f = frequency, (%) = percentage.*

Stage/statement	**SDis f (%)**	**Dis f (%)**	**N f (%)**	**A f (%)**	**SA f (%)**	**Mean**	**SD**
Planning							
P1. You are involved in making key decisions in MKFR conservation activities	41 (13.2)	35 (11.2)	40 (12.9)	125 (40.2)	70 (22.5)	3.48	1.31
P2. Meetings are held regularly to plan MKFR conservation activities	25 (8.0)	21 (6.8)	28 (9.0)	100 (32.2)	137 (44.0)	3.97	1.24
P3. You are involved in making MKFR management plans	22 (7.1)	26 (8.4)	29 (9.3)	112 (36.0)	122 (39.2)	3.92	1.21
P4. Forest management plans are revised and updated frequently	42 (13.5)	37 (11.9)	43 (13.8)	122 (39.2)	67 (21.6)	3.43	1.31
P5. Election of forest executive committee members are fair	39 (12.5)	40 (12.9)	42 (13.5)	117 (37.6)	73 (23.5)	3.53	1.34
Overall mean and SD						3.67	1.28
Implementation							
I6.You are regularly informed about forest conservation issues and activities	22 (7.1)	23 (7.4)	33 (10.6)	116 (37.3)	117 (37.6)	3.91	1.19
I7. You have access to tools/equipment used in MKFR's conservation activities	44 (14.2)	49 (15.7)	99 (31.8)	50 (16.1)	69 (22.2)	3.16	1.32
I8. You use extension services for forest management and conservation	42 (13.5)	37 (11.9)	44 (14.2)	113 (36.3)	75 (24.1)	3.46	1.34
I9. You are aware of rights/duties regarding law and regulations in forest management	28 (9.0)	31 (10.0)	35 (11.2)	93 (29.9)	124 (39.9)	3.82	1.30
I10. Financial resources are at your disposal to implement forest conservation activities	41 (13.2)	45 (14.5)	76 (24.4)	82 (26.4)	67 (21.5)	3.29	1.31
Overall mean and SD						3.53	1.29

Evaluation							
E11. Good collaboration exists among CFAs in MKFR conservation activities	27 (8.7)	31 (10.0)	33 (10.6)	101 (32.4)	119 (38.3)	3.82	1.28
E12. There is improved biodiversity in MKFR due to forest conservation activities	17 (5.5)	23 (7.4)	26 (8.4)	110 (35.4)	135 (43.4)	4.04	1.14
E13. Increased forest products have improved people's livelihood	24 (7.7)	23 (7.4)	22 (7.1)	122 (39.2)	120 (38.6)	3.94	1.20
E14. MKFR's conservation activities have led to more development in the community	25 (8.0)	22 (7.1)	27 (8.7)	106 (34.1)	131 (42.1)	3.95	1.23
E15. You are willing to sacrifice more time to participate in conservation activities	30 (9.6)	33 (10.6)	29 (9.3)	107 (34.4)	112 (36.0)	3.76	1.30
Overall mean and SD						3.90	1.23
Monitoring							
M16. In MKFR, you keep an eye on tree planting and maintenance	21 (6.7)	18 (5.8)	23 (7.4)	112 (36.0)	137 (44.1)	4.05	1.16
M17. You monitor illegal activities in MKFR	26 (8.4)	32 (10.3)	23 (7.4)	111 (35.7)	119 (38.3)	3.85	1.27
M18. You monitor the protection of water sources in MKFR	23 (7.4)	23 (7.4)	27 (8.7)	105 (33.8)	133 (42.8)	3.97	1.22
M19. Relevant technology is used in forest conservation and management in MKFR	98 (31.5)	113 (36.3)	37 (11.9)	41 (13.2)	22 (7.1)	2.29	1.24
M20. Monitoring reports are made available to the public regularly	38 (12.2)	43 (13.8)	51 (16.4)	99 (31.8)	80 (25.7)	3.45	1.33
Overall mean and SD						3.52	1.24

According to Table 6.3, when it came to the statement P1 that stated, ‘You are involved in making key decisions in MKFR conservation activities,’ 24.5% of respondents disagreed, 62.7% agreed, and 12.9% were neutral. The statement's mean score of 3.48 was lower than the stage's overall mean score of 3.67. This implies that CFA members' failure to participate in key decision-making does not support conservation and management activities in MKFR. As a result, residents lack a sense of ownership over the processes and are hesitant to participate actively. On statement P2, 76.3% agreed that meetings are held regularly to plan MKFR conservation activities, while 9.0% neither agreed nor disagreed. Only 14.7% of those polled were in disagreement. The mean score for the statement was 3.97, which was higher than the overall mean score for the table. This meant that when MKFR conservation meetings were held regularly, they made decisions that favoured forest conservation. Concerning statement P3, 75.2% of respondents agreed that they are involved in the development of MKFR management plans. Only 15.5% agreed, with 9.3% neutral. The mean score of 3.92 was higher than the overall mean score of the table. This demonstrated that CFA members' involvement in developing forest management plans benefits conservation efforts in MKFR. Essentially, by participating, CFA members feel like they are a part of the forest conservation process. Only 60.7% of respondents agreed with Statement P4's assertion that ‘Forest management plans are revised and updated frequently,’ while 25.4% disagreed and 13.8% were unsure. The mean score of 3.43 was lower than the overall mean score, indicating that failure to revise and update the forest management plan frequently did not support MKFR conservation because delays in incorporating issues of major concern that may require immediate attention may work against forest conservation efforts. Finally, 25.4% disagreed with statement P5 about the fairness of the election of forest executive committee members. Approximately 61.1% agreed, while 13.5% were neutral. The mean score for the statement was 3.53, which was lower than the overall mean score. This showed that unfair elections for the forest executive committee could result in ongoing conflicts that would harm conservation efforts in the MKFR. There is a lack of transparency among officials, a dictatorial tendency among some of the leaders, and mistrust and suspicion among members.

Table 6.3 shows that statement I6, You are frequently informed about issues and actions relating to forest conservation,' was made by CFAs who participated in the implementation of conservation activities in MKFR. 74.9% agreed, 14.5% disagreed, and 10.6% were neutral. The mean score of 3.91 was higher than the overall mean score of 3.53 for the 'Implementation' stage. As a result, by regularly updating CFA members on forest conservation issues and activities, they are given the pertinent information they need to conserve MKFR. On the accessibility of tools and equipment used in MKFR's conservation activities, i.e., statement I7, approximately 29.9% of respondents disagreed, while 31.8% were neutral. The statement's mean score of 3.16 was lower than the overall mean, implying that the lack of tools and equipment hurts MKFR's conservation efforts. Statement I8 found that 25.4% of respondents disagreed that they have access to extension services related to forest conservation and management in MKFR, while only 60.4% agreed. The mean score for the statement was 3.46, which was lower than the overall mean score of 3.53. This meant that the lack of extension services for CFA members would have an impact on the conservation of MKFR because they won't be able to receive guidance on issues that call for specialized involvement. In the statement I9 question about being aware of rights and duties regarding law and regulations in forest management, 69.8% agreed, 11.3% were neutral, and 19.0% disagreed. The average score of 3.82 was higher than the overall average of 3.53. This meant that CFA members are aware of their rights and responsibilities while carrying out conservation and management activities in MKFR, and they will not do anything that will harm the forest. In response to statement I10, 34.1% disagreed and 33.5% agreed that they have the financial means to carry out conservation efforts in the MKFR. Approximately 24.4% were neutral. The mean score of 3.29 for this statement is lower than the overall mean score. As a result, the lack of funds to carry out conservation activities in MKFR lowers the morale of CFA members. Members' registration and monthly/annual subscription fees are the primary sources of funding for CFAs. There are numerous major challenges, such as insufficient funding, misappropriation of funds, some members who do not contribute funds, which causes benefit sharing to become an issue, and so on. This leads to an overreliance on outside assistance, particularly from non-governmental organizations (NGOs). It becomes increasingly difficult to sustain members' enthusiasm, commitment, and efforts over time.

Statement E11 of the evaluation of conservation activities in the MKFR reveals that, while 70.8% of respondents agreed and 10.6% disagreed that CFAs work well together, 18.7% of respondents did not agree with this statement. The statement's mean score of 3.82 was lower than the overall mean score of 3.90. This means that a lack of good collaboration among the CFAs has resulted in some of them being disorganized and sometimes disagreeing, which has harmed conservation efforts in MKFR. Statement E12, ‘There is improved biodiversity in MKFR due to forest conservation activities,’ received 78.8% agreement and 8.4% neutral response. 12.9% of those polled said they disagreed. The overall score was higher than the mean score of 4.04. This demonstrated how improved biodiversity in MKFR was the result of CFAs' concerted efforts. This has also resulted in an increase in forest products and, as a result, improved livelihoods through community participation in PELIS, IGAs—Income Generating Activities, and long-term access to timber and non-timber forest products. Statement E13 has a mean score of 3.94, which is higher than the overall mean for the evaluation stage. The livelihood of the people has improved, according to 77.8% of the 311 respondents. Only 7.1% agreed, while 15.1% disagreed. In response to statement E14, 76.2 percent of respondents agreed that MKFR's conservation efforts have resulted in more development in forest-adjacent communities. Neutral respondents made up 8.7% of the sample, while 15.1% of respondents disagreed. The statement's mean score of 3.95, which was higher than the overall mean for the 'Evaluation' stage, implies that as a result of conservation efforts in MKFR, more people are now using electricity as a source of energy, progress in road network connectivity, establishment and improvement of infrastructure for schools and health facilities, and so on. The majority (70.4%) of respondents agreed with Statement E15's ‘You are willing to sacrifice more time to participate in conservation activities,’ although the statement's mean score of 3.76 is lower than the overall stage's. This meant that CFA members' failure to devote more time to conservation activities in MKFR could harm forest products and, eventually, their livelihood. Those who disagreed thought their involvement in conservation activities was sufficient, among other reasons as listed under I10, including the fact that they had other family-related errands to run. Only 9.3% of the 311 respondents agreed, while 20.3% disagreed.

In response to statement M16, 76.1% agreed that they monitor the planting and care of trees in MKFR, while 12.6% disagreed and 7.4% were indifferent. The mean score of 4.05 was higher than the mean score of 3.52 for the 'monitoring' stage. This indicates that tree monitoring and care contributed to the conservation of MKFR. Statement M17, ‘You monitor illegal activities in MKFR,’ received 74.0% agreement, 18.7% disagreement, and 7.4% neutral. The fact that the mean score of 3.85 was higher than the stage's average mean score indicated that CFA members' surveillance of unauthorized activity aided in the conservation of MKFR. 76.6% of respondents strongly agreed on CFA members monitor water sources in MKFR, i.e. statement M18. Only 8.7% of them said they were unsure. The average score of 3.97 was higher than the overall average score. It implies that CFA members supported MKFR conservation by monitoring water sources. Water quantity and quality have both improved. ‘Relevant technology is used in forest conservation and management in MKFR,’ according to M19. 67.8% of respondents opined that technology was not being used in forest conservation and management activities. While 11.9% of respondents were neutral, 20.3% of CFA participants agreed. The monitoring section's mean score of 2.29 was lower than the overall mean score. As a result, the failure to use technology such as GPS devices and communication equipment such as UHF radios in MKFR has had an impact on conservation. On the question of whether monitoring reports are made available to the public regularly (Statement M20), 26.0% disagreed, while 57.5% agreed. 16.4% of them were undecided. The monitoring section's mean score of 3.45 was lower than the overall mean. Conservation activities in MKFR are hampered by a lack of coordination and information sharing.

6.3.5 Policy and institutional framework and conservation of MKFR

Policy and institutional frameworks at both national and subnational levels create a favorable environment for the sustainable management of forests. The study sought to determine the gauge the impact of institutional, legal, and policy framework on the PFM and MKFR conservation. As a result, respondents were asked to rate their levels of agreement/disagreement with the statements that follow in Table 6.4.

Table 6.4: *Policy and institutional framework and conservation of Mt. Kenya Forest Reserve (MKFR); SDis = strongly disagree, Dis = disagree, N = neutral, A = agree, SA = strongly agree, SD = standard deviation, f = frequency, (%) = percentage.*

Statement	**SDis f (%)**	**Dis f (%)**	**N f (%)**	**A f (%)**	**SA f (%)**	**Mean**	**SD**
P/I1. There are clearly stated and assigned user rights for the forest resources	26 (8.4)	19 (6.1)	40 (12.9)	129 (41.4)	97 (31.2)	3.81	1.19
P/I2. Forest benefits are shared equitably	33 (10.6)	35 (11.2)	36 (11.6)	88 (28.3)	119 (38.3)	3.72	1.35
P/I3. Collaboration with external institutions is good for the conservation of MKFR	27 (8.7)	23 (7.4)	28 (9.0)	117 (37.6)	116 (37.3)	3.88	1.24
P/I4. Stakeholders are aware of the activities allowed and not allowed in MKFR	31 (10.0)	33 (10.6)	31 (10.0)	116 (37.3)	100 (32.1)	3.71	1.29
P/I5. MKFR conservation activities are allocated sufficient funds	99 (31.8)	123 (39.6)	21 (6.8)	39 (12.5)	29 (9.3)	2.28	1.29
P/I6. KFS has enough trained staff who assist in forest conservation activities in MKFR	92 (29.6)	115 (37.0)	19 (6.1)	58 (18.6)	27 (8.7)	2.40	1.32
P/I7. Training is carried out frequently in support of MKFR’s conservation activities	108 (34.7)	91 (29.3)	27 (8.7)	43 (13.8)	42 (13.5)	2.42	1.43
P/I8. The current forest administration has minimized illegal activities in MKFR	35 (11.2)	26 (8.4)	42 (13.5)	89 (28.6)	119 (38.3)	3.74	1.34
P/I9. Existing forest laws adequately deal with illegal activities in MKFR	91 (29.3)	112 (36.0)	24 (7.7)	41 (13.2)	43 (13.8)	2.46	1.39
P/I10. Political interventions have solved the plight of the local community concerning the extraction of forest products	87 (28.0)	93 (29.9)	61 (19.6)	32 (10.3)	38 (12.2)	2.49	1.32
Overall mean and SD						3.09	1.32

Only 72.7% agreed with the statement P/I1 that ‘There are clearly stated and assigned user rights for the forest resources,’ while 12.9% were neutral and 14.5% disagreed (Table 6.4). The statement received a mean score of 3.81, which was higher than the mean score of 3.09 in the table. This

indicates that the clarification of user rights over forest resources has had a positive impact on the conservation of the forest reserve because CFA members extract what the user rights stipulate from the forest. In response to statement P/I2, 66.6% of respondents agreed that forest resources are distributed equitably. As many as 11.6% were neutral. The mean score of 3.72 was higher than 3.09, indicating that equitably sharing forest resources reduced unnecessary conflicts among CFA members, and thus their attention was focused on forest conservation. More respondents (74.9%) agreed that collaboration with external institutions is beneficial to MKFR conservation (statement P/I3). The mean score of 3.88 was higher than the table's mean. This indicates that collaboration with institutions such as those listed in Section 6.3.3 enabled CFAs to access external sources of funding to supplement the little they collect through member registration fees. Only 10.0% of respondents were neutral toward statement P/I4, while 20.6% disagreed. 69.5% of respondents were in favor of the idea that stakeholders are aware of the activities that are and are not permitted in MKFR. The mean score of 3.71 indicates that knowledge about activities permitted and prohibited in MKFR will reduce illegal activities, resulting in increased forest products and improved livelihood for forest-adjacent communities. The majority of respondents (71.4%) disagreed with statement P/I5, i.e. MKFR conservation activities are adequately funded. The mean score of 2.28 indicates that due to a lack of funds, MKFR conservation cannot be carried out as planned. This has an impact on the quality and quantity of forest products, eventually impoverishing the livelihoods of the forest's surrounding communities. Unless significant additional funding is requested from donors, including the global conservation beneficiaries, conservation efforts will often come at a local cost, and CFAs won't be able to meet their goals for conservation-related activities. In response to statement P/I6, 66.6% of respondents agreed that the Kenya Forest Service (KFS) does not have enough trained staff to assist in MKFR forest conservation activities. The statement's mean score of 2.40 was lower than the table's overall mean, implying that a lack of trained personnel would delay conservation in MKFR because CFA members would not receive assistance in extension services or technical support on forest conservation and management. 64.0 percent of respondents disagree with the statement ‘Training is conducted on a regular basis to support MKFR's conservation efforts.’ The 2.42 mean score is significantly lower than the overall mean; thus, a lack of conducting conservation training as frequently as possible results in CFA members not being effectively equipped in forest conservation matters. The net result will be unmet conservation targets. According to statement P/I8, 66.9% of respondents agreed that the current forest administration has reduced illegal activities in MKFR. The average score of 3.74 indicated that the current KFS administration in Tharaka-Nithi County was doing an admirable job of controlling illegal activities in MKFR. This is the complete opposite of what occasionally occurs when those responsible for protecting forest resources get involved in illegal activity. Despite the accomplishment, the current administration faces a slew of challenges in its day-to-day operations, including a staff shortage, insufficient funding, insufficient tools and equipment, and so on. On statement P/I9, 65.3% of respondents believed that existing forest laws do not adequately address illegal activities in MKFR. Only 27.0% agreed, while 7.7% were neutral. The mean score of 2.46 versus the table's mean score of 3.09 indicates that leakage in current forest laws has resulted in the penetration of illegal activities in MKFR. This is in line with what the KFS officials stated, which is that low fines for offenders promote more illegal activity in MKFR and that even when arrests are made, convictions may not be realized due to flaws in the judicial system. As a result, forest crime has evolved into a low-risk, high-profit enterprise. All possible offenses should be identified and defined in the policy and legal framework governing the forest sector, and appropriate penalties should be applied. Repeat offenders should face harsher penalties. Finally, the majority of respondents, 57.9%, disagreed with statement P/I10, i.e. political interventions have solved the plight of the local community regarding the extraction of forest products. The statement's average score of 2.49 suggests that political meddling has a detrimental impact on MKFR conservation. Politicians frequently interfere with forest management affairs, for example, in their desperation to earn votes, they encourage local communities to invade forests, telling them that they are the true owners of the forests and have free access to forest resources.

6.4 Discussion

Local communities and other stakeholders taking part in forest management and conservation can improve environmental sustainability, increase forest productivity, alleviate poverty, and make forest access rules more enforceable. Government commitment will determine whether PFM is implemented because it takes time and resources to (World Bank, 1995; Mbuvi and Kung'u, 2020):

- build consensus among stakeholders;
- create new institutional arrangements;
- decentralize finance and administration;
- establish suitable rules/incentives for local involvement, and
- strengthen local organizational capability.

Participatory forest management (PFM) approaches raise questions about their long-term viability because co-management processes and institutional arrangements put in place may be easily tampered with by the more powerful members of society, even though PFM can bring about significant benefits (Schreckenberg *et al*., 2006). Still, it is difficult to assess the implications of such action because there is no solid impact evidence related to PFM, partly due to the inability to measure the range of benefits and costs for different groups of people (Menzies, 2002).

According to the policymakers interviewed, some of the outcomes of PFM in Kenya are enhanced natural resource management, and improved access to forest resources. Others are increased public participation in forest conservation and management, and decreased costs incurred by the central government in forest conservation. Thus, PFM and related issues have enhanced forest governance in MKFR and throughout Kenya. Positive outcomes in PFM, according to consensus, are dependent on local governments being accountable to resource users. About 69.9% of respondents supported the idea of improved relations between the state and the communities.

6.4.1 Consequences of illegal activities in Mt. Kenya Forest Reserve

The policymakers noted that beyond the economic consequences, forest crimes have far-reaching social and environmental impacts. The livelihoods of hundreds of thousands of poor people depend on MKFR. Any illegal activities affect them directly, adding to their vulnerability. Such crimes also impair delicately balanced forest ecosystems, because they disturb the conservation of forest resources and biodiversity (KFS and MENR, 2007; KFS, 2010). Many forest laws advocate for the adoption of a sustainable management approach, thus they include social welfare, e.g. recreation, livelihood support, etc., and conservation of forest ecosystem services, e.g. carbon sequestration, watershed conservation, biodiversity protection, etc., which do not have a market price (FAO/ITTO, 2005). By absorbing greenhouse gases, MKFR significantly contributes to Kenya's water catchment protection and climate control. This has resulted in environmental damage because illegal activities are a threat to biodiversity, and destabilize soils and groundwater, thus affecting agricultural activity around the forest. Therefore, when illegal acts take place in forests, complex ecosystems are disturbed or may disappear. In Kenya, forests serve as the primary source of domestic fuel wood and are a crucial source of raw materials for both wood-based industries and non-wood forest products (KFS, 2010). Therefore, illegal activities may interfere with market forces for forest products, e.g. timber products, thus hindering sustainable management, which is subject to higher costs linked to good husbandry, proper tax declaration, etc. (FAO/ITTO, 2005; Chebii, 2015). Therefore, illegal activities in MKFR have lowered the market value of forest products. The comparative profitability of sustainably produced forest products has diminished, hence giving a competitive advantage to the illegal operators (Chebii, 2015). In general, investments in forestry and related sectors are minimal since donors and investors shy away from investing where investment risks are very high and there are leakages in forest laws, or generally, the rule of law is weak (Chebii, 2015). This is characteristic of many countries within the tropics, Kenya inclusive. Therefore, such a dilemma has hampered conservation activities in MKFR. Also, illegal activities in MKFR have resulted in a reduced supply of forest resources, which to some extent has affected the livelihoods of some members of the forest adjacent communities. Forest mismanagement creates a scenario whereby demand for forest resources becomes more than the supply of the supply, and as noted previously, this may create conflict among CFA members. Illegal activities in MKFR have deprived the government of revenue due to uncollected tax, stumpage fees, and other related costs. The relationship between illegal activities and inequality is a complicated one, especially when the rule of law is weak; in MKFR, although the poor may illegally extract forest resources, it is only the most powerful community members who eventually dominate the utilization of forest resources. Therefore, corruption will affect the livelihoods of the poor more than the others. The benefits of illegal activities accruing to the poor are therefore transitory (FAO/ITTO, 2005; Chebii, 2015). Moreover, illegal activities in MKFR may also weaken the government's capacity to offer services, e.g. technical assistance and proper delimitation and consolidation of property rights, to the forest-dependent poor.

6.4.2 Stumbling blocks in forest governance reform affecting conservation of MKFR

The policymakers acknowledged the increase in population and poverty as a crucial factor dragging the government's efforts in attaining sustainable forest management (SFM) through PFM. Due to the lack of resources on private and public lands, the population must turn to nearby protected areas for food and other necessities, which creates the possibility of encroachment and other illegal activities. In fact, due to their low economic status, 81% of the respondents depend on firewood obtainable from the forest. Because of poverty and a high reliance on natural forests for livelihoods, especially during droughts, people frequently violate forest laws and regulations

(KFS, 2010). Those who believed they have a claim to ownership as well as access to the goods and services provided by the forest constituted 19%, while 11.3% neither agree nor disagree that they are aware of their rights/duties regarding law and regulations in forest management. The problem is compounded by the failure to have frequent training among the CFA members on their role, rights, and responsibilities in the conservation of forests, and the need to protect them for the common good. The respondents who do not report illegal activities happening in MKFR form 25.4%, while 13.2% do not know whether to report or not. Disagreements have been witnessed among the forest user groups concerning the election of forest executive committee members. Equitable sharing of forest resources has led to disagreements. This degrades the integrity of those tasked with the governance of MKFR. Another stumbling block in forest governance reform affecting the conservation of MKFR is the condition that communities through CFAs must come up with management plans for them to be involved in managing adjacent forests. Most CFA members lack the technical know-how to produce and/or implement the plans. Worse still, KFS does not have adequately trained staff who can assist. Alternatively, the CFAs hire professionals who can design such plans. In some cases, it would take time to obtain authority approval.

The ineffective legal and policy frameworks for enforcing forest laws and managing them are also noted. For instance, the overlap of numerous Acts that deal with forestry, such as the Agriculture Act, Water Act, the Wildlife Act, etc., which are not harmonized, makes it unclear who is responsible for what and where (KFS and MENR, 2007; Chebii, 2015). The disagreements over the use of forest resources by various users and agencies are indicative of the conflicting laws governing the forestry sector. For example, despite providing for joint law enforcement, the memorandum of understanding between KWS and KFS lacks coherent implementation (Chebii, 2015). The conflict can be acerbated if the two players together with the Ministry of Water and Sanitation impose fees for environmental services. The same confusion is created due to the dual gazettement of parts of MKFR, thus it becomes difficult when it comes to activities such as patrolling, firefighting, and the overall accountability in law enforcement (KFS and MENR, 2007). Another challenge is the fact that the forest policy mainly focuses on government programs and pays little attention to the provision of financial and technical resources to forestry programs (Chebii, 2015). Also sometimes forest directives, e.g. presidential ban on logging may not be followed up with implementation guidelines (World Bank, 2001). This leaves room for manipulation by those in charge. Further, in most cases, the government passes laws but the issuance of the corresponding operational regulations is delayed. Uncertainty looms about how the law should be applied in the meantime (Contreras-Hermosilla, 2011). For example, it took 10 years before the 2005 policy was implemented and three more years before the Forests Act came into being (GOK, 2007; Walubengo, 2007). The delay demonstrates the issue of vested interest in the forestry sector. In other cases, operational regulations may be ambiguous, thus prone to subjective interpretations. Therefore, the key to SFM is the lack of interference in forest policy formulation by political pressures and the government‘s commitment to its policies (Chebii, 2015).

Laws may be technically inapplicable if they advocate activities, procedures, and institutional arrangements which do not go hand in hand with the required financial and human resources (Contreras-Hermosilla, 2011; Smyle *et al*., 2016). Laws can be socially unacceptable if they are perceived to be unfair, e.g. when they ignore or even punish local practices and norms (Smyle *et al*., 2016; Cheruiyot *et al*., 2019). Coupled with inadequate/lack of public participation in their design and decision-making processes, this can result in long-term irreversible socio-economic and environmental impacts, and increased levels of illegal forest acts (Kairu *et al*., 2018; Chebii, 2015; Mbuvi and Kung'u, 2020). Most forest-surrounding communities will refuse to comply with forest laws if they believe it will jeopardize their livelihoods. In some circumstances, it may be challenging to adhere to forestry-related laws and regulations, particularly for small farmers and economically disadvantaged rural communities who may not be able to comply with expensive and complex legal requirements (FAO/ITTO, 2005; Chebii, 2015). According to a senior KFS official, '…payment for some services can be high for the community to access/afford them.' These are some of the reasons why even logging bans put in place to check illegal activities in MKFR, have sometimes proven counter-productive. The forest land tenure is also often unclear or discriminatory, creating major barriers to effective forest governance in MKFR (KFS and MENR, 2007). Traditional rights may not be legally noted even though the forest adjacent communities may have been the original owners (Contreras-Hermosilla, 2011; Chebii, 2015). Therefore, the lack of secure and legal forest land tenure results in irresolvable conflict between communities and the government (Chebii, 2015). This remains a barrier to the effective implementation of PFM.

Forest laws may be under-utilized or not utilized at all due to a lack of political will, institutional failure, corruption, or outright criminal mentality (KFS and MENR, 2007; Contreras-Hermosilla, 2011). This leads to selective application and enforcement of forest laws and regulations, creating opportunities for neglect and abuse, resulting in poor forest governance

(World Bank 2001). Corruption and overall disregard for the rule of law, together with abuse of office take different forms, for example:

- harvesting more forest products than one has paid for;
- irregular issuance of licenses, e.g. KFS single-handedly issues the General Forest License (GFL) for harvesting forest products;
- bribing to avoid arrests/prosecution;
- bribing to get harvesting rights;
- irregular allocation of forest land to repay political debts; and
- favouritism in employment/promotions, etc.

Such is the threat to forest law enforcement and governance (KFS and MENR, 2007). The policymakers agreed that corruption impedes sustainable forest management (SFM). About 21% of the respondents from the local community felt that corruption exists in the management of MKFR. The way the legislation is designed also influences law enforcement. Leakage in law enforcement and low penalties for offenders make many people perceive the law as unfair (Contreras-Hermosilla, 2011; Kairu *et al*., 2018). The lack the adequate human and financial resources to enact forest law compliance leads to poor forest law enforcement and governance (Contreras-Hermosilla, 2011; Smyle *et al*., 2016; Kairu *et al*., 2018). In some places, for policiticians to win votes, they encourage forest adjacent communities to invade forests and extract forest products. This made planning and application of forest management practices ineffective because it weakens the forest inspection and protection function. Weak government institutions lead to illegal activities in MKFR since the chance of being punished is low (Kairu *et al*., 2018). Forests are subject to intense and detailed regulation compared to other land uses—overregulation is often linked with more corruption (FAO/ITTO, 2005). The most powerful in society know how to wade through the complex network of forest laws and regulations to get authorization to harvest forest resources. The poor in society will also definitely design their schemes on how to illegally access the forest to get forest products (Kairu *et al*., 2018). Therefore, overregulation is deemed to be among the major causes of environmental degradation (FAO/ITTO, 2005). The challenge with policy and institutional reforms and programs is that they are delicate and time-consuming (may take years) (KFS and MENR, 2007). They may involve changes in structure and relationship with other institutions, and other workers' careers. KFS has taken a long time to be where it is, and reforms are still ongoing. Therefore the future is bleak, i.e. it may not be clear what governance and law enforcement reform measures will or not. Thus, the government's commitment and political will are crucial. Another challenge of forest governance reform affecting the conservation of MKFR is a weak judicial system (Chebii, 2015). A reasonably functioning judiciary is essential for the effective implementation of forestry laws. But, Kenyan courts are mostly understaffed and underfinanced, and overburdened with caseloads. As well, the training of judges who deal with typical forest-related matters may be weak and will not necessarily involve any training in forestry laws (FAO, 2012; Chebii, 2015).

Finally, the dimensions of forest governance are closely interlinked, therefore, while solving one or a few problems seems to be a relief, it may not be sufficient enough to attain optimal forest governance. To date, Kenya can claim better management of forest resources, through improved frameworks for governance. This has been achieved through significant institutional, legal, and policy reforms toward SFM through PFM. This study can however already identify areas that require special attention to achieve optimal management of forest resources.

6.4.3 Policy failure and recommendations

In most cases, numerous forest laws result in conflicting regulations (FAO/ITTO, 2005; Contreras-Hermosilla, 2011; Chebii, 2015). Such regulations need to be harmonized so that they are consistent with each other and be able to define who is responsible for what (Chebii, 2015). According to FAO/ITTO (2005), a correlation exists between the abundance of regulations, the size of corruption, and the illegal market. Developing countries mostly have excessive regulations (World Bank, 2001). This is mostly a sign of a weak policy framework that cannot solve the long-term objectives of SFM through PFM (FAO/ITTO, 2005; Kairu *et al*., 2018). This increases the danger of legal discrepancies, ambiguities, and conflicting jurisdictions, and further confuses the understanding of the forest regulatory framework (Chebii, 2015). This has left room for subjective interpretation of forest law, increasing the chances of corruption thriving. Thus, flaws in forest policy and regulatory frameworks may lead to profound undesirable effects (Contreras-Hermosilla, 2011). The following guidelines may limit excessive regulations, overlaps, and discrepancies (FAO/ITTO, 2005):

- before law-making, an effective forest policy should be in place to guide legislation formulation;
- an extensive legal analysis should be done to identify overlaps and inconsistencies to abridge the legal framework;
- following the identified overlaps and inconsistencies, reduce the number of rules;

- discard any official decision or bureaucratic action called for in the law or regulations not directly linked to the attainment of a policy objective;
- adapt the rest of the rules but they must be consistent with other existing laws, including those of other departments and sectors; and
- immediately after the new laws have been ratified translate new legislative acts into working norms and regulations.

Also, forest governance failures can be due to policies and laws in other sectors of the economy, e.g. policies and regulations on land tenure, agricultural expansion, and industrial and transportation infrastructure development (Schmithüsen, 2002; Chebii, 2015).

Development of subsidiary legislation is required for SFM, especially targeting fiscal policies and particularly tax regimes. For example, royalty should be levied on the raw materials used in charcoal production; encourage usage of environmentally desirable products like Liquid Petroleum Gas (LPG) and kerosene by lowering their tariffs and taxes, and reducing tariffs on imported wood/wood products. A significant challenge is a long period it may take to design these subsidiary legislations through consultative and participatory processes, and eventually the gazettement process (KFS and MENR, 2007). Also, the entire process of developing these subsidiary legislations and guidelines and their implementation require both financial and human resources (Smyle *et al*., 2016). Therefore, the focus should be on recognizing urgent subsidiary legislations and concentrate on their development and implementation.

6.4.4 Institutional challenges and recommendations

The current poorly equipped forest rangers are few, they cover large tracts of forest land, with wild animals especially elephants and the terrain is difficult. Thus, they feel demoralized because most of the time they walk long distances in the thick forest with neither GPS nor radio communication. The capacity for law enforcement is low (Chebii, 2015; Kairu *et al*., 2018). Most people opine that illegal activities in MKFR are because of low morale of staff, corruption, and the low number of forest rangers. Thus, to achieve effective policing, more funds (a challenge) are required to employ extra well-trained and equipped forest rangers, purchase relevant equipment, training of forest officers as prosecutors, among other needs. This is the opposite of expectations that the number of forest rangers should go down as CFAs become more involved in forest conservation and management activities in MKFR. Another problem related to law enforcement manifests itself in form of contradictory and/or impossible to comply with requirements governing the forest sector, therefore, they must be eliminated (Chebii, 2015). A proper procedure must be followed to do so, as explained in the previous section. But the guiding principle should be how the regulations affect actors in the forest sector, i.e. concerning their financial, technical, and managerial capacity to follow the legal requirements and their needs (Contreras-Hermosilla, 2011). In circumstances where regulations reduce profitability thereby preventing realistic compliance, the government may offer support, e.g. financial incentives and/or compensation to motivate forest actors to willingly observe the law (FAO/ITTO, 2005).

KFS does not get sufficient funding from the state and the little they collect from charging for forest products harvested by CFAs is insignificant (Chebii, 2015). Since 2018 there is a moratorium on logging in place to help streamline management and governance in the forestry sector in Kenya, following an outcry from the public because of illegal logging that led to a reduction in water levels in major rivers. Therefore KFS should strive to innovate ways of generating finances to achieve SFM in MKFR. Examples are introducing ecotourism levies, charging for environmental services, e.g. imposing tariffs on the water as well as electricity-generating companies for using water, and competitive long-term leases of forest plantations in areas designated as such. To reduce overdependence on indigenous forests, KFS should champion the development of private enterprises to exploit commercial tree farming on private and trust lands for forest products, e.g. gums and resins, timber, and biofuels (*Jatropha curcas*). Eventually, this will contribute significantly to increased forest cover. But, a challenge manifests itself when the regulation's main objective turns out to be the government aiming to take control of forest lands, raise revenue, etc. (KFS and MENR, 2007).

Failure by stakeholders to participate in forest conservation and management activities, e.g. failure by some CFA members to involve themselves in making key decisions in MKFR conservation activities, coupled with failure to avail monitoring reports regularly, and poor quality of information whenever they are made public, is a major concern. A participatory approach to forest law-making ensures transparency, controls corruption, supports greater equity, reduces the gratuitous influence of the 'powerful' in society, and promotes the rule of law. This will make stakeholders air their grievances and safeguard their rights from absolute unilateral decisions by governments. A forest policy accepted by all portrays a sense of joint ownership i.e. involvement of stakeholders from all sectors legitimizes the policy in society.

Research contributes to the protection of forests from pests and diseases, and it is a crucial aspect in making sure that forest products are not undervalued (KFS and MENR, 2007). Forest research is also key in making informed decisions concerning policy on, e.g. tree and seed improvement, advantages and disadvantages of the 'shamba' system, etc. The 'shamba' system is a way to afforest government land where exotic trees have been cut, then the community is allowed to plant crops until canopy closure (Chebii, 2015). It was banned in 2003 because it was full of controversy and widespread malpractice, leading to environmental degradation, but the ban was lifted in 2007 (Chebii, 2015). Thus, both local and global institutions specialized in forestry research should collaborate and adequately address major issues working against achieving SFM through PFM.

Finally, to have an effective judicial system to deal with cases involving offenders involved in illegal activities in forests, the forestry governance institutions among other stakeholders should lobby for broad reforms in the judicial system. The weak knowledge of forestry law matters among the judiciary should be alleviated through professional training programs in forestry and related laws for those involved in natural resource management together with their advocates, judges, lawyers, administrators, and NGOs (FAO/ITTO, 2005).

6.4.5 Other challenges

6.4.5.1 Insufficient information

Transparency is key to the reduction of corruption and stopping government officials from taking part in illegal activities (Chebii, 2015). Lack of transparency interferes with decisions on procurement contracts, awarding and monitoring of concession contracts, fine calculation and collection for illegal activities, and staff responsibilities (FAO/ITTO, 2005; Contreras-Hermosilla, 2011; Chebii, 2015). Policy reforms on the granting and monitoring of concessions and subsidies, as well as accountability issues in forest administration, all contribute to greater transparency. Full disclosure of documents motivates people to be directly involved in forest law enforcement thereby equipping them with relevant knowledge on forestry issues. According to FAO/ITTO (2005), the following information should be available in the public domain, on the Internet, and in printed form through forest administrations:

- all laws and related rules, regulations, decrees, and decisions in the forestry sector;
- forest inventory showing how much forest there is, the type, location, and a value assessment;
- a map indicating how forests are used;
- agreements on concession and investment;

- details regarding the ownership of concession;
- the forest department structure, roles of forest department personnel and other stakeholders, and contact details;
- government officials' information regarding family ties to the logging industry;
- information about banned companies;
- conditions and rules for the award of concessions;
- independent inspection reports; and
- prosecution of offenders.

6.4.5.2 Participatory Forest Management

As noted by Cheeseman *et al*. (2016), the PFM decentralization model ensures equitable distribution of resources because more resources are brought to the grassroots level. PFM may not be able to comprehend community norms, roles, and obligations in the same way as those of KFS (KFS and MENR, 2007). In Kenya, draft PFM guidelines are already in place, but there is a need to validate and develop relevant tools for training and capacity building in local communities (Chebii, 2015; Kairu *et al*., 2018). As well, the degree of participation is subject to application to the Director of KFS. According to one of the KFS officials in Tharaka-Nithi County, the benefits received from MKFR will motivate local communities to protect the forests. This translates to better management and lower costs for KFS. However, questions remain about whether the incentives are sufficient to have a meaningful impact on local communities, how long it will take for the impact to be realized, and how long it will last. PFM is prone to political interference because forest resources are considered to be a free good (Chebii, 2015). Clear rules with clearly defined responsibilities and contributions are essential for PFM's success because it has a cost to the stakeholders (Chebii, 2015; Kairu *et al*., 2018).

The Ministry of Environment and Natural Resources has already started a review of the National Forest Policy to amend the current Forest Conservation and Management Act 2016. Concerns include:

- the disparity between the objectives in the National Forest Policy 2020 document and the actual policy statements—its objectives are too numerous and some are repetitive;
- a lack of adequate information on public participation; and
- the phase-out of the PELIS—but retaliation and forest destruction may follow.

6. 5 Conclusion

Previously, the government tried other community-based forest conservation approaches but the PFM decentralization model has proven to be effective. PFM involves communities surrounding forests and other stakeholders in managing forests, in a manner that contributes to increased conservation awareness and livelihoods of the communities. Through PFM, a sense of ownership has been realized by forest-adjacent communities who actively engage in forest conservation activities in MKFR. Although there is a lot of good news to write about PFM, forests are still faced with threats ranging from poaching of wild animals to illegal logging, despite heavy investment in the creation of awareness against them, championing for alternative livelihoods, and better forest governance. This chapter has noted some underlying hurdles in the current policy and legal frameworks which have affected adequate forest law enforcement and governance (FLEG):

- numerous forest laws that are conflicting—a sign of a weak policy framework;
- low capacity for law enforcement because the forest guards are few and poorly equipped;
- the KFS does not get sufficient funding from the government—this has interfered with KFS's objective of enhancing forest conservation and management;
- failure by some CFA members to involve themselves in making key forest conservation and management decisions;
- failure to present monitoring reports regularly, and poor quality of information whenever they are made public; and
- inadequate understanding and appreciation of forestry law among the judiciary make the forest law itself amorphous.

Forest policies and practices have changed significantly over time as a result of shifting societal demands; they must continue to change in order to stay relevant, helpful, and responsive to shifting societal demands. To shape a broader vision for Kenya's forestry sector in the years to come, national policies should aim to incorporate future needs and trends in them.

CHAPTER SEVEN

The remote sensing of forest canopy gaps in a selectively logged sub montane tropical forest reserve: A synthesis

7.1 Introduction

Unsustainable selective logging (SL) has been identified as a significant factor contributing to the degradation of tropical forests, but the rate and amount of the degradation is not clear (Asner *et al*., 2009). Because of this uncertainty, the effects of forest degradation on biodiversity in the tropics have been misunderstood and misinterpreted (Borrego and Skutsch, 2014). The spatial pattern of canopy gaps may be misrepresented in the absence of reliable methods for accurate mapping of canopy gaps. The accurate delineation of gaps serves as the foundation for future forestry research and management applications.

The size of canopy gaps has been estimated by ground-based measurements, but these methods are prone to error, time-consuming, and strenuous in rough terrain, among other factors. Using remote sensing (RS) methods, attempts were made to reduce the effort involved in collecting canopy gap data. However, because of its partial disturbance of the forest canopy and small scale, SL can be difficult to detect and quantify (Dalagnol *et al*., 2019). Very high resolution (VHR) i.e. <1 m per pixel satellite imagery and airborne data are suitable for the accurate delineation of small forest canopy gaps (Dalagnol *et al*., 2019). Before the introduction of high-resolution optical data, traditional aerial photography was successfully used to map canopy gaps (Malahlela *et al*., 2014). Nonetheless, data acquisition is expensive, but technological advances have revitalized its use via unmanned aerial vehicles (UAV). Unmanned aerial vehicles (UAVs) and LiDAR data have been used to detect gaps (Asner *et al*., 2013). Calibrated VHR SAR data have been used to detect canopy gaps caused by individual tree removal (Baldauf, 2009). However, each of the data types has its challenges.

It is difficult to locate a gap and define its boundaries using any method (Vepakomma *et al*., 2008). According to Nyamgeroh (2015), the studies on canopy gap delineation are few, and methods to delineate canopy gaps are even fewer. As a result, dependable remote sensing (RS) techniques that can detect and map canopy gaps are required.

In this thesis, the objectives were:

1. To review the status, trends, potentials, and challenges and recommended future directions in remote sensing (RS) techniques used to map selective logging (SL) in the tropical forests from 1992 to 2019;
2. To use species-level tree classification to map threatened trees species in a selectively logged sub-montane heterogeneous tropical forest—MKFR;

3. To explore the potential of spectral features—visible-near-infrared bands and vegetation indices—to model canopy gaps caused by logging of Ocotea trees;
4. To apply a texture-spectral analysis framework—using GLCM - and MLBP-based rotation-invariant feature descriptors—to exploit the potential of the very high resolution WorldView-3 multispectral imagery to model canopy gaps; and
5. To assess how inadequacies in Kenya's forest policy and law on participatory forest management led to illegal activities in forests in Kenya.

7.2 Remote sensing of selective logging in tropical forests

The application of moderate resolution RS imagery such as Landsat TM or SPOT 4 to detect canopy gaps from SL may present spectral confusion of canopy gaps due to natural disturbances such as a windfall. Due to the sub-pixel scale of SL gaps, the broad spectral range of Landsat wavelengths cannot detect spectrally subtle changes. Fusion of panchromatic and multispectral images and upscaling of spatial resolutions are some of the commonly applied techniques to enhance lower resolution imagery. In intensive SL analysis, high spatial resolution (5-10 m pixel size) images enable the detection of logging roads, log landings, and tree fall gaps. But, for SL where only individual trees are targeted, the application of satellites or UAVs (or manned aircraft) VHR data is viable in detecting and mapping disappearing tree crowns. Canopy height models (CHM) acquired through repeat surveys of LiDAR are applied in detecting the removal of individual trees. LiDAR data can also estimate aboveground biomass (AGB) loss due to logging. LiDAR is the only RS method that can map SL with high accuracy. Pinage *et al*. (2019) used LiDAR data to map canopy gaps in Eastern Amazon—did not involve explicit ground-based validation for canopy gaps because it was not necessary. The major challenge of using LiDAR to detect SL is that LiDAR data is costly, and thus limits its application. Among optical, RADAR, and LiDAR sensors, LiDAR is the richest source of data because it provides full cross-sectional information on forests. Data availability, cost, and the complexity of the analysis are the major factors that determine the type of data to use rather than which system would provide the most comprehensive information about a forest. Thus, VHR satellite data such as WorldView-3 (WV-3) come in handy.

A variety of RS techniques which differ in their mapping goals, approach, geographic extent, and map accuracies, have been applied in detecting and estimating the intensity of SL in tropical forests. These techniques have a myriad of challenges. Most RS methods used in monitoring SL in tropical forests use Landsat imagery, which can detect selectively logging at moderately high intensities, i.e. >20 m^3 ha^{-1}; 3-7 trees ha^{-1}). But the methods are unable to quantify the extent and duration of logging damage in areas experiencing lower logging intensities, i.e. <20 m^3 ha^{-1}) (Hethcoat *et al*., 2018). The detection of small canopy gaps has been done using VHR RS data. The application of visual interpretation in identifying areas degraded by SL activities is time-consuming because it requires more workload. The accuracy of visual interpretation is highly dependent on the skills of the interpreter. In the application of the Spectral Mixture Analysis (SMA) technique, the differences in spectral reflectance due to topography and atmospheric conditions determine the accuracy of detecting SL, thus extensive field studies are always necessary. This limits the application of SMA over large areas. Post-classification analysis has also been used in mapping SL in the tropics, but this technique can only be applied in bi-temporal data analysis. Since Landsat-based monitoring is not adequate to detect low-intensity SL, other studies have used object-based change detection techniques. Unlike pixel-based analyses, object-based analyses minimize the high local spectral variability of the canopy and reduce the mismatch of tree crown geo-misallocation (Dalagnol *et al*., 2019). The classification of RS data could be effective and efficient thanks to machine learning (ML) which can handle high dimensionality data and map extremely complex classes (Maxwell *et al*., 2018).

In validating SL maps, *in situ* data is the most reliable, although the major challenge is the collection of adequate reference samples in a dense forest. If *in situ* data is not available, RS data, data from model simulations, or prior knowledge of the study area by experts can be used. According to Banskota *et al*. (2014), a validation technique must have an appropriate sampling and responsive design for the selection and labelling of reference samples. The results of accuracy assessment provide the possibility of comparing maps, therefore, providing RS researchers with the information they need to assess the reliability of new methods and modelling techniques (Morales-Barquero *et al*., 2019).

7.3 Classification of endangered tree species using an imbalanced dataset

This study was aimed to use WorldView-2 (WV-2) data to map endangered tree species in MKFR by making use of an imbalanced dataset and random forest (RF) and support vector machine (SVM) classifiers. The imbalanced dataset was a result of the inability to collect adequate samples of some tree species (e.g. *Anthocleista grandiflora*, *Prunus Africana*, *Ocotea usambarensis*, and *Albizzia gummifera*) in the study area. Only eight samples were collected for *Ocotea usambarensis*, therefore, this tree species was combined with other tree species with fewer samples to form the 'other woody vegetation' class. The study found that there were spectral overlaps within the tree species classes. The overlaps could have been due to, among others:

- the complex forest structure in the study area;
- some stands were multi-storied; and
- some tree crowns that were relatively small leading to mixed pixels.

This ultimately resulted in misclassifications between species, which was in agreement with their inter-specific separability as calculated using the Jeffries-Matusita distance. Only *Syzygium guineense* could be separated from the other six tree species. Generally, values greater than 1.9 indicate that the classes are separable, while 1.7 - 1.9, shows that the separation is fairly good, but below 1.7, the separation is poor. Examples with very low separability are *Newtonia buchananii* and *Albizzia gummifera* (0.59), and *Newtonia buchananii* and *Anthocleista grandiflora* (0.61). Also, the tree species with the lowest J-M values reported significantly low F1 scores for both classifiers and vice versa.

The variable importance measurement in the RF algorithm identified band 5 (Red), band 6 (Red Edge), band 7 (Near Infrared 1), and band 8 (Near Infrared 2) as the most crucial variables in the classification process; they reported the highest mean decrease in accuracy (MDA) values. Therefore, the visible and the infrared portions of the electromagnetic spectrum were relevant in mapping endangered tree species in the study area. The MDA was also used to show bands that were crucial in detecting particular tree species.

The difference in sample size across classes in the original dataset resulted in a rise in the rate of classification errors; tree species with more sample sizes had commission errors which were higher compared to omission errors, and vice versa. Also, tree species with small samples reported very high F1-score variability across iterations. When using an imbalanced dataset, the final classification under-predicts the classes with fewer samples. Therefore, test samples of smaller classes are most likely misclassified than those of the dominant classes. The degree of bias in the model was reduced when training data size was standardized across classes, thus higher classification accuracies were attained.

7.4 Mapping canopy gaps using narrow-band vegetation indices

The study aimed to explore the potential of vegetation indices extracted from WorldView-3 (WV-3) imagery to detect canopy tree loss associated with SL of *Ocotea usambarensis* in MKFR. The

RF's variable importance measurement identified the means of vegetation indices such as the CARI, the NPCI, and the REPI, that of the shadow index (SDI), the standard deviation (SD) of the ARI, and the brightness feature, together with individual means of the visible-near-infrared (VNIR) bands as being critical in detecting canopy gaps. The variable importance measurement also identified the means of the CARI, the CRI-2, the MCARI, the PSRI, the REPI, and those of bands 1 (Coastal), 3 (Green), 5 (Red), and 6 (Red Edge) to help identify vegetated gaps, while the means of bands 4 (Yellow) and 5 (Red) were crucial in identifying shaded gaps. The brightness feature, and the means of bands 2 (Blue) and 7 (Near Infrared 1), the SDI and the NPCI, and SD of ARI majorly aided in the identification of forest canopy class. The vegetation indices were mostly derived from bands 5 (Red) and 6 (Red Edge) of the multispectral WV-3 imagery.

A correlation matrix was computed from the reference samples to determine whether or not two features share the same characteristics. Only correlation values of more than 0.70 are significant. Therefore, pairs of variables with such values are correlated, providing redundant information. Examples are the means the NPCI, the CARI, the MCARI, the CRI-2, and the PSRI, and those of the visible bands. The highest negative correlations were reported between the mean of the REPI and those of the CARI, the CRI-2, the MCARI, the NPCI, and the PSRI, i.e. –1.000, –0.718, –0.943, – 0.877, and –0.749, respectively. A negative correlation means that a variable is the opposite of the other. Therefore, depending on the threshold value, the two variables might be of importance during classification. The relationship between two variables is opposite when there is a perfect negative correlation, e.g. the relationship between means of the REPI and the CARI. High correlation among pairs of variables makes such variables less important in canopy gap modelling; one merit of ML classifiers is that they can efficiently handle this collinearity.

Both the transformed divergence (TD) index and Jeffries–Matusita (J-M) distance separability measures were used to determine the distance between any two signatures. Within one standard deviation (Figure 4.5), the variables depicted varying degrees of overlap between shaded and vegetated gaps, and forest canopy. For example, the means of the CRI-2, the NPCI, and the PSRI, those of the visible bands, and SD of the ARI. The means of the CARI, the MCARI, and the REPI, and that of band 6 (Red Edge), and the brightness feature, did not show any spectral overlap between shaded gaps and the other two classes which showed major spectral overlaps. Only the means of bands 7 (Near Infrared 1) and 8 (Near Infrared 2), and the SDI were able to fully discriminate between the three classes. According to the TD index and J-M distance values from

Table 4.3, the comparison between the forest canopy and vegetated gaps recorded the lowest value of separation.

All classes reported very high user's accuracy (UA) and producer's accuracy (PA), resulting in high overall accuracy (OA) for RF and SVM models. This was attributed to the features extracted from the pansharpened WV-3 multispectral data. Also, the large sample size for the classes provided more accurate mean values. Manual delineation of the forest landscape classes may have also contributed to higher classification accuracies because only relevant pixels were targeted. Overall, the shaded gap class reported the highest UA and PA, since the other two classes represent vegetation, and their spectral properties are different from the non-vegetation class. The results also showed that the SDI assisted in accurately mapping the shaded gaps. The forest canopy had lower UA and PA—probably due to the wide range of reflectance characteristics by tree crowns in the WV-3 imagery.

7.5 Enhancing the detection and mapping of canopy gaps

This study also aimed to enhance the detection of canopy gaps by fusing grey level co-occurrence matrix (GLCM) and multivariate local binary pattern (MLBP) based rotation-invariant descriptors derived from the WV-3 multispectral dataset. Initially, the rotation invariant local binary pattern (LBP) operator was used to get the LBP images of a subset of images extracted from a WV-3 scene covering the study area. Then the GLCMs of the LBP images were calculated using the homogeneity, contrast, entropy, angular second moment, and correlation to describe the image texture features. The LBP texture measure was applied to colour images by applying a multivariate texture model, i.e. the multivariate local binary pattern (MLBP) operator. To describe the uniformity of the values of pixels, a uniformity measure was applied to the LBP measure. The circular symmetrical neighbour set approach was applied in this study because it provides beneficial effects in terms of accuracy and robustness. Although larger window sizes enable the drawing of more texture information from a pixel, thus could increase accuracy, enlarging the scope of the GLCM window can reduce sensitivity to class edges, and effectively smooth over the image. This negatively impacts the classification accuracy.

The accuracy report shows that better classification results were obtained—GLCM features perform better with few classes i.e. 10 or less. The Multivariate LBP models recorded higher classification accuracies as compared to the univariate LBP measures. The SVM classification of

MLBP/CON reported the highest classification accuracy of 95.1%. Therefore, the fusing of textural and spectral information from three WV-3 bands performed better than their basic models.

7.6 Forest policy and institutional framework on participatory forest management

An assessment of Kenya's forest policy and law on participatory forest management (PFM) was done to determine to what extent have the alterations to institutional structures, laws, and policies aimed at promoting the shift to SFM been successful. This was done by interviewing officials from KWS, NEMA, and KFS who are at the policy-making level. Also, data were collected from the members of community forest user groups (CFUGs) living adjacent to Chuka Forest in MKFR. The involvement of regional groups and other interested parties in the management and preservation of forests can promote environmental sustainability, increase the productivity of forests, alleviate poverty, and make rules governing access to forests more enforceable. But, among the major concerns of PFM are that the PFM methods might not be as beneficial to the poor as they are designed because co-management processes and the institutional frameworks put in place may be interfered with by the more powerful members of society. According to the policymakers interviewed, PFM and related issues have improved forest governance in Mt. Kenya Forest Reserve (MKFR), and Kenya in general. The general agreement was that positive outcomes in PFM depend on the accountability of local governments to resource users.

However, some key weaknesses were noted in the policy and institution framework governing the conservation and management of forests through PFM in Kenya. For example, there are cases of numerous forest laws that are conflicting; there exists a correlation between the abundance of regulations, and the size of corruption and the illegal market. This is a sign of a weak policy framework. There is also a low capacity for law enforcement because the forest guards are demoralized because they are few and poorly equipped. Most people think that illegal activities in MKFR are due to low morale of staff, corruption, and the low number of forest rangers. Another major challenge is that KFS does not get sufficient funding from the government. This has severely dented KFS's objective of enhancing conservation and management in MKFR. Also, failure by stakeholders to participate in forest conservation and management activities, e.g. failure by some CFA members to involve themselves in making key decisions in MKFR conservation activities (the reason why a high number of respondents—than expected, chose 'neutral' in the questionnaires), together with the failure to present monitoring reports regularly, and poor quality of information whenever they are made public. A collaborative approach to forest law-making promotes transparency, checks corruption, ensures greater equity, and supports the rule of law. Again, major judicial reforms are necessary. Inadequate understanding and appreciation of forestry law among the judiciary make the forest law itself amorphous. Since transparency can eradicate corruption and stop the involvement of government representatives in criminal activity, full disclosure of documents, e.g. information about companies found operating illegally, and an explicit ban to take part in subsequent auctions or allocation procedures for concessions, the rules and conditions for the award of procurement contracts and concessions, together with a report after the bidding on which lots were awarded to whom, how much was paid for them, prosecution of offenders, inspection reports from independent bodies, among others.

In recognizing the rules, obligations, and responsibilities of the KFS staff by the communities, PFM may have limited capacity. There is still a need to certify and create pertinent tools for educating and increasing the capacity of local communities. The local communities will be motivated to conserve MKFR because of the benefits they get resulting in better management and lower costs for KFS. But, are the incentives good enough to make a real impact in the local communities? How long will it take for the impact to be felt? For how long will it last? PFM is prone to political interference because forest resources are considered to be a free good, therefore, clear guidelines with well-defined roles and responsibilities lead to its success.

7.7 Conclusions

The primary goals of this study were to assess the potential of multispectral remote sensing (RS) techniques to identify and map tree species that are at risk, as well as to detect and map individual canopy tree loss brought on by SL in a closed-canopy tropical forest. Two machine learning (ML) classification algorithms—random forest (RF) and support vector machine (SVM)—were used.

The following findings from the study objectives are the foundation for the main conclusions:

- Although several studies have been involved in monitoring SL in tropical forests, mapping the areal extent or impacts of SL using RS techniques remains a challenge. This study established that monitoring SL in tropical forests requires a combination of methods and sensors.
- Standardized sample sizes across classes contributed to higher classification accuracies, therefore, this study was able to identify and map the endangered tree species. The most

significant bands in the mapping of the endangered tree species include the bands 5 (Red), 6(Red Edge), 7 (Near Infrared 1), and 8 (Near Infrared 2).

- The application of pansharpened WV-3 data and ML algorithms detected vegetated canopy gaps resulting from illegal logging of camphor trees, and vegetation indices derived from bands 5 (Red) and 6 (Red Edge) were equal to the task. The WV-3 sensor's spectral and spatial characteristics were responsible for the relatively high classification accuracy.
- The fusion of the GLCM and MLBP operator improved the detection of vegetated canopy gaps in MKFR. The fusing of textural and spectral information from three WorldView-3 bands performed better than their basic models.
- Although the PFM decentralization model is being lauded as the best, forests are still faced with a myriad of malpractices such as illegal logging. This is despite a lot of investment in creating awareness of such vices. Consequently, this study has identified underlying hurdles in the current policy and institutional framework which if rectified can lead to adequate forest law enforcement and governance (FLEG).

This research has shown how gaps in forest canopies in tropical areas can be identified and quantified using the VHR WV-3 multispectral dataset, and how this could be used to identify illegal logging. This has a significant contribution toward carbon accounting of forests, as well as the mapping of proliferation of invasive species. The international community is becoming increasingly interested in tropical forests and their critical role in the global carbon cycle. None of the algorithms used in carbon estimation of forests account for canopy gaps. Selective logging (SL) which is a major cause of forest degradation in tropical forests, causes long-term carbon emissions into the atmosphere. The accuracy of carbon estimates can be improved, provided canopy gaps are accurately identified and correctly estimate trees in per unit area of forest. The integration of both optical images and LiDAR data may boost canopy gap classification—further contributing to accurate estimation of carbon densities of forests.

Canopy gaps can stimulate the regrowth of secondary forests dominated by fast-growing species, significantly reducing the presence of shade-tolerant species. This interferes with the diverse range of biodiversity in tropical forests, thus interferes with forest health and the availability of essential ecosystem functions and services. Forest managers and policy makers can use this information to make better-informed decisions when identifying priority areas or species for conservation. Therefore, forest management is critical to the global carbon budget.

7.8 The future

This research provided alternative methods of detecting and mapping vegetated canopy gaps in a submontane tropical forest using multispectral satellite data. The study established that a variety of methods and sensors are especially important for monitoring SL in tropical forests. Although a significant increase in studies involving mapping SL has been reported, several challenges still abound which require the attention of researchers.

The following are recommendations for solving some of these challenges:

- While metrics extracted from the multispectral satellite data used in this study are useful in mapping canopy gaps, there is a need to try other data types (e.g., hyperspectral data) and methods in combination with ML or statistical classifiers.
- More sophisticated data fusion techniques by researchers are expected soon that can be more reliable and effective in mapping SL.
- Commitment by space agencies for the systematic and coordinated acquisition of RS data and with an open data policy—the success of data integration will rely on the compatibility of multi-source RS data, specifically the consistency in spatial, spectral, temporal, and radiometric resolutions.
- Technological advancement is required concerning UAVs to collect and share UAV data at broader scales and at minimal costs; the building and deploying a fleet of high-altitude imaging drones should also be explored.
- Innovation of automated methods for processing LiDAR data to lower data processing costs—will enable data acquisition of extensive areas while providing repeatable and consistent estimates of vital forest attributes.
- Further research is needed on RS techniques used in monitoring SL that effectively apply the increasingly diverse and complex remotely sensed data acquired or projected to be acquired from space and airborne sensors.

REFERENCES

Achard, F., Eva, H.D., Stibig, H-J., Mayaux, P., Gallego, J., Richards, T., and Malingreau, J-P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, 297, 999-1002.

Adam, E., Mutanga, O., Odindi, J., and Abdel-Rahman, E.M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35, 3440-3458.

Adam, Y. O., and Eltayeb, A. M. (2016). Forestry decentralization and poverty alleviation: A review. *Forest Policy and Economics*, 73, 300-307.

Andersen, H.E., Reutebuch, S.E., McGaughey, R.J., D'Oliveira, M.V. and Keller, M. (2014). Monitoring selective logging in western Amazonia with repeat lidar flights. *Remote Sensing of Environment*, 151, 157-165.

Antropov, O., Rauste, Y., Seifert, F.M. and Hame, T. (2015). Selective logging of tropical forests observed using L- and C-band SAR satellite data. In: Proceedings of the "IEEE International Geoscience and Remote Sensing Symposium (IGARSS)". Milan (Italy) 26-31 July 2015. IEEE Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 3870-3873.

Anwar, S., and Stein, A. (2012). Detection and spatial analysis of selective logging with geometrically corrected Landsat images. *International Journal of Remote Sensing*, 33, 7820-7843.

Arivazhagan, S., and Ganesan, L. (2003). Texture classification using wavelet transform. *Pattern Recognition Letters*, 24, 1513-1521.

Arivazhagan S, Ganesan L, and Priyal S (2006) Texture classification using Gabor wavelets based rotation-invariant features. *Pattern Recognition Letters*, 27, 1976-1982.

Arroyo-Mora, P., Kalacska, M., and Chazdon, R. (2009). Spectral unmixing of forest canopy recovery in selectively logged units in a tropical lowland forest, Costa Rica. In: Anais "XIV Simposio Brasileiro de Sensoriamento Remoto" [Annals XIV Brazilian Symposium on Remote Sensing]. Natal (Brasil) 25-30 Apr 2009. INPE, Sao Jose dos Campos, Sao Paulo, Brazil, pp. 2539-2546.

Asner, G.P., Keller, M., Pereira, R., and Zweede, J.C. (2002). Remote sensing of selective logging in Amazonia: assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. *Remote Sensing of Environment*, 80, 483-496.

Asner, G.P., Keller, M., Pereira, R., Zweede, J.C., and Silva, J. (2004). Canopy damage and recovery after selective logging in Amazonia: field and satellite studies. *Ecological Applications*, 14, 280-298.

Asner, G.P., Knapp, D.E., Broadbent, E.N., Oliveira, P.J.C., Keller, M., and Silva, J.N. (2005). Selective logging in the Brazilian Amazon. *Science*, 310, 480-482.

Asner, G.P., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Knapp, D.E., and Silva, J.N.M. (2006). Condition and fate of logged forests in the Brazilian Amazon. *Proceedings of the National Academy of Sciences USA*, 103, 12947-12950.

Asner, G.P., Hughes, R.F., Vitousek, P.M., Knapp, D.E., Kennedy-Bowdoin, T., and Boardman, J. (2008). Invasive plants transform the three-dimensional structure of rainforests. *Proceedings of the National Academy of Sciences USA*, 105, 4519-4523.

Asner, G.P., Keller, M., Lentini, M., Merry, F., Souza, Jr., C. (2009). Selective logging and its relation to deforestation. In: Keller, M., Bustamente, M., Gash, J., Dias, P.S., eds. Amazonia and global change. Geophysical Monograph Series. Vol. 186 Washington, DC: American Geophysical Union. P. 25-42.

Asner, G.P., Kellner, J.R., Kennedy-Bowdoin, T., Knapp, D.E., Anderson, C., and Martin, R.E. (2013). Forest canopy gap distributions in the southern Peruvian Amazon. *PLoS ONE,* 8, e60875.

Athanasiou, L.S., Fotiadis, D.I., and Michalis, L.K. (2017). Plaque Characterization Methods Using Intravascular Ultrasound Imaging, Editor(s): Athanasiou, L.S., Fotiadis, D.I., and Michalis, L.K., Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging, Academic Press, Pages 71-94.

Baldauf T, and Köhl M (2009). Use of TerraSAR-X for forest degradation mapping in the context of REDD: In proceedings of the World Forestry Congress in Buenos Aires on 23 October 2009, Buenos Aires, Argentina.

Baldauf, T. (2013). Monitoring reduced emissions from deforestation and forest degradation (REDD+): capabilities of high-resolution active remote sensing. Ph.D. thesis, Biology Department, University of Hamburg, Hamburg, Germany, pp. 94.

Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Kellner, J.R., and Wright, S.J. (2015). Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. *PLoS ONE*, 10, e0118403.

Banskota, A., Kayastha, N., Falkowski, M.J., Wulder, M.A., Froese, R., and White, J.C. (2014). Forest monitoring using Landsat time-series data - A review. *Canadian Journal of Remote Sensing*, 40, 362-384.

Battles, J.J., Dushoff, J.G., and Fahey, T.J. (1996). Line intersect sampling of forest canopy gaps. *Journal of Forest Science*, 42, 131-138.

Betts, H.D., Len J. Brown, L.J., and Stewart, G.H. (2005). Forest canopy gap detection and characterisation by the use of high-resolution Digital Elevation Models. *New Zealand Journal of Ecology*, 29, 95-103.

Bianconi. F., and Fernández, A. (2014). Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform. *Pattern Recognition Letters*, 48:34-41.

Blackburn, G.A. (1998). Quantifying chlorophylls and carotenoids from leaf to canopy scale: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66, 273-285.

Bonnet, S., Gaulton, R., Lehaire, F., and Lejeune, P. (2015). Canopy gap mapping from airborne laser scanning: An assessment of the positional and geometrical accuracy. *Remote Sensing*, 7, 11267-11294.

Borg, I., and Groenen, P.J.F. (2005). Modern multidimensional scaling: Theory and applications, 2nd ed.; Springer Science + Business Media: Berlin, Germany.

Borrego A, Skutsch M (2014) Estimating the opportunity costs of activities that cause degradation in tropical dry forest: Implications for REDD+. *Ecological Economics*, 101, 1-9.

Boyd, D. S., Hill, R. A., Hopkinson, C., and Baker, T. R. (2013). Landscape-scale forest disturbance regimes in southern Peruvian Amazonia. *Ecological Applications*, 23, 1588-1602.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

Broadbent, E.N., Zarin, D.J., Asner, G.P., Pena-Claros, M., Cooper, A., and Littell, R. (2006). Recovery of forest structure and spectral properties after selective logging in lowland Bolivia. *Ecological Applications*, 16, 1148-1163.

Broadbent, E.N., Asner, G.P., Keller, M., Knapp, D.E., Oliveira, P.J.C., and Silva, J.N. (2008). Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biological Conservation*, 141, 1745-1757.

Brokaw, N.V.L. (1982). The definition of treefall gap and its effect on measures of forest dynamics. *Biotropica*, 14, 158-160.

Brokaw, N.V.L., and Grear, J.S. (1991) Forest structure before and after Hurricane Hugo at three elevations in the Luquillo Mountains, Puerto Rico. *Biotropica*, 23, 386-392.

Buja, A., Swayne, D.F., Littman, M.L., Dean, N., Hofmann, H., and Chen, L. (2007). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, 17, 444-472.

Burivalova, Z., Bauert, M.R., Hassold, S., Fatroandrianjafinonjasolomiovazo, N.T., and Koh, L.P. (2015). Relevance of global forest change data set to local conservation: a case study of forest degradation in Masoala National Park, Madagascar. *Biotropica*, 47, 267-274.

Burnett, M.W., White, T.D., McCauley, D.J., De Leo, G.A., and Micheli, F. (2019). Quantifying coconut palm extent on Pacific islands using spectral and textural analysis of very high-resolution imagery. *International Journal of Remote Sensing*, 40, 1-27.

Bussmann, R.W., and Beck, E. (1995). Regeneration- and cyclic processes in the Ocotea-Forests (*Ocotea usambarensis* Engl.) of Mount Kenya. *Verhandlungen der Gesellschaft für Ökologie*, 24, 35-38.

Bussmann, R.W. (1996). Destruction and management of Mount Kenya's forests. *Ambio*, 25, 314-317.

Carifio, J., and Rocco, J.P. (2007). Ten common misunderstandings, misconceptions, persistent myths, and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3, 106-116.

Chavez, P.S. Jr. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24, 459-479.

Chawla, N.V., Japkowicz, N., and Kolcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6, 1-6.

Chebii, J.K. (2015). Forest management and conservation in Kenya: a study of the role of law in the conservation of forest resources. Doctoral thesis. University of South Africa.

Cheeseman, N., Lynch, G., and Willis, J. (2016). Decentralisation in Kenya: the governance of governors. *Journal of Modern African Studies*, 54, 1–35.

Chen, C., A. Liaw, and and Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. Berkeley, CA: University of California Berkeley. https://statistics.berkeley.edu/sites/default/files/techreports/666.pdf

Chen, X., Gerlach, B., and Casasent, D. (2005). Pruning support vectors for imbalanced data classification. In Proceedings of International Joint Conference on Neural Networks, Montreal, QC, Canada, 31 July - 4 August 2005; pp. 1883–1888.

Cheruiyot, J., Otieno, L., Omutoko, and Rambo, C. (2019). Institutionalization of forest conservation program for enhancement of participatory forest management, Bomet County, Kenya. *International Journal of Academic Research and Reflection*, 7, 28-38.

Chia, E.L., Hubert, D., Enongene, K., and Tegegne, Y.T. (2020). An AHP assessment of barriers to adopting sustainable forest management practices in the context of carbon emission reductions in Cameroon. *Journal of Sustainable Forestry*, 39, 379-391.

Chicco, D., and Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, 6.

CIPFA (2012). Delivering Good Governance in Local Government. London: The Chartered Institute of Public Finance and Accountancy (CIPFA).

Clark, M.L., Clark, D.B., and Roberts, D.A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, 91, 68-89.

Clark, M.L., Roberts, D.A., and Clark, D.B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96, 375-398.

Clark, M.L., and Roberts, D.A. (2012). Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing*, 4, 1820-1855.

Clausi, D.A. (2002). An Analysis of Co-Occurrence Texture Statistics as a Function of Grey Level Quantization. *Canadian Journal of Remote Sensing*, 28, 45-62.

Cohen, W.B., and Spies, T.A. (1992). Estimating structural attributes of Douglas fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, 41, 1-17.

Conde, T.M., Higuchi, N., and LIMA, A.J.N. (2019). Illegal selective logging and forest fires in the northern Brazilian Amazon. *Forests*, 10, 61.

Contreras-Hermosilla, A. (2011). People, governance and forests-The stumbling blocks in forest governance reform in Latin America. *Forests*, 2, 168-199.

Cooper, D., and Schindler, P. (2007). Business Research Methods. NewYork, McGraw-Hill Higher Education

Costa, O.B., Matricardi, E.A.T., Pedlowski, M.A., Miguel, E.P., and Gaspar, R.O. (2019). Selective logging detection in the Brazilian Amazon. *Floresta e Ambiente*, 26, e20170634.

Cross, G.R., and Jain A. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5, 25-39.

Cui, H., Qian, H., Qian, L., and Li, Y. (2009). Remote sensing experts classification system applying in the land use classification in Guangzhou City: In proceedings of the 2nd International Congress on Image and Signal Processing, CISP'09, 17-19 October 2009, Tianjin, China.

Dalagnol, R., Phillips, O.L., Gloor, E., Galvao, L.S., Wagner, F.H., Locks, C.J., Luiz, E.O.C., and Aragao, L.E. (2019). Quantifying canopy tree loss and gap recovery in tropical forests under low-intensity logging using VHR satellite imagery and airborne lidar. *Remote Sensing*, 11, 817.

Da Ponte, E., Fleckenstein, M., Leinenkugel, P., Parker, A., Oppelt, N., and Kuenzer, C. (2015). Tropical forest cover dynamics for Latin America using Earth observation data: a review covering the continental, regional, and local scale. *International Journal of Remote Sensing*, 36, 3196-3242.

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., and McMurtrey, J.E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74, 229-239.

De Grandi, E.C., Mitchard, E., Woodhouse, I.H., and De Grandi, G.D. (2015). Spatial wavelet statistics of SAR backscatter for characterizing degraded forest: a case study from Cameroon. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 3572-3584.

De Wasseige, C., and Defourny, P. (2004). Remote sensing of selective logging impact on tropical forest management. *Forest Ecology and Management*, 188, 161-173.

Descals, A., Szantoi, Z., Beck, P., Brink, A., and Strobl, P. (2017). Automated detection of selective logging using Smallsat imagery. *IEEE Geoscience and Remote Sensing Letters*, pp. 1-5.

Deutscher, J., Perko, R., Gutjahr, K., Hirschmugl, M., and Schardt, M. (2013). Mapping tropical rainforest canopy disturbances in 3D by COSMO-SkyMed spotlight InSAR-Stereo data to detect areas of forest degradation. *Remote Sensing*, 5, 648-663.

DigitalGlobe (2010). The Benefits of the 8 Spectral Bands of WorldView-2. Available online: https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/35/DG-8SPECTRAL-WP_0.pdf (accessed on 2 February 2019).

DigitalGlobe (2019). WorldView-3. Above + Beyond. Available online: http://worldview3.digitalglobe.com/ (accessed on 2 February 2019).
Di Ruberto C., Fodde G., and Putzu L. (2015). Comparison of Statistical Features for Medical Colour Image Classification, LNCS of the International Conference on Computer Vision Systems, 9163, 3-13.
D'Oliveira, M.V.N., Reutebuch, S.E., McGaughey, R.J., and Andersen, H-E. (2012). Estimating forest biomass and identifying low-intensity logging areas using airborne scanning lidar in Antimary State Forest, Acre State, Western Brazilian Amazon. *Remote Sensing of Environment*, 124, 479-491.
Edwards, D.P., Tobias, J., Sheil, D., Meijaard, E., and Laurance, W.F. (2014). Maintaining ecosystem function and services in logged tropical forests. *Trends in Ecology and Evolution*, 29, 511-520.
Ellis, P., Griscom, B., Walker, W., Gonçalves, F., and Cormier, T. (2016). Mapping selective logging impacts in Borneo with GPS and airborne lidar. *Forest Ecology and Management*, 365, 184-196.
Englhart, S., Jubanski, J., and Siegert, F. (2013). Quantifying Dynamics in Tropical Peat Swamp Forest Biomass with Multi-Temporal LiDAR Datasets. *Remote Sensing*, 5, 2368-2388.
FAO (1999). Developments in international conventions and agreements. In: the state of the world’s forest. FAO cooperate document repository, http://www.fao.org
FAO and ITTO (2005). Best practices for improving law compliance in the forest sector. FAO forestry paper 145. FAO and ITTO, Rome.
FAO (2010). Global forest resources assessment. FAO, Rome.
FAO (2012). State of the World’s Forests: the first ten issues: In State of the World’s Forests, FAO, Rome.
FAO (2015). Global forest resources assessment. FAO Forestry Paper No. 1**. FAO, Rome.
FAO (2022). The State of the World’s Forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies. FAO, Rome.
Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E., and de Souza Filho, C.R. (2016). Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, 179, 66-78.
Ferreira, M.P., Wagner, F.H., Aragão, L.E.O.C., Shimabukuro, Y.E., and de Souza Filho, C.R.S. (2019). Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 119-131.
Féret, J.-B., and Asner, G.P. (2012). Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and LiDAR. *Remote Sensing*, 4, 2457-2476.
Feret, J.B., and Asner, G.P. (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. IEEE *Transactions on Geoscience and Remote Sensing*, 51, 73-84.
Foody, G.M., and Mathur, A. (2004). Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93, 107-117.
Forest Act (2005). Forest Act 2005. Government of Kenya.
Fox, T.J., Knutson, M.G., and Hines, R.K. (2000). Mapping forest canopy gaps using air-photo interpretation and ground surveys. *Wildlife Society Bulletin*, 28, 882-889.
Franke, J., Navratil, P., Keuck, V., Peterson, K., and Siegert, F. (2012). Monitoring fire and selective logging activities in tropical peat swamp forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5, 1811-1820.

Furusawa, T., Pahari, K., Umezaki, M., and Ohtsuka, R. (2004). Impacts of selective logging on New Georgia Island, Solomon Islands evaluated using very-high-resolution satellite (IKONOS) data. *Environmental Conservation*, 31, 349-355.

Galabuzi, C., Eilu, G., Nabanoga, G. N., Turyahabwe, N., Mulugo, L., Kakudidi, E., and Sibelet, N. (2015). Has the evolution process of forestry policies in Uganda promoted deforestation? *The International Forestry Review*, 17, 298-310.

Gamon, J.A., Serrano, L., and Surfus, J.S. (1997). The photochemical reflectance index: An optical indicator of photosynthetic radiation-use efficiency across species, functional types, and nutrient levels. *Oecologia*, 112, 492-501.

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

Gatti, L.V., Gloor, M., Miller, J.B., Doughty, C.E., Malhi, Y., Domingues, L., Basso, L., Martinewski, A., S. C. Correia, C., Borges, V., Freitas, S., Braz, R., Anderson, L., Rocha, H., Grace, J., Phillips, O., and Lloyd, J. (2014). Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. *Nature*. 506, 76-80.

Gaulton, R., and Malthus, T.J. (2010). Lidar mapping of canopy gaps in continuous cover forests; a comparison of canopy height model and point cloud-based techniques. *International Journal of Remote Sensing*, 31, 1193-1211.

Getzin, S., Nuske, R.S., and Wiegand, K. (2014). Using unmanned aerial vehicles (UAV) to quantify spatial gap patterns in forests. *Remote Sensing*, 6, 6988-7004.

GFW (2022). Global Forest Watch. Available online: https://www.globalforestwatch.org (accessed on 10 October 2022).

Ghasemi, N., Sahebi, M.R., and Mohammadzadeh, A. (2013). Biomass estimation of a temperate deciduous forest using wavelet analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 765-776.

Ghate, R. (2004). Un-commons in the commons: Community-initiated forest resource management. New Delhi, India, Concept Publishing Company.

Gibson, L., Lee, T.M., Koh, L.P., Brooks, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F., Lovejoy, T.E., and Sodhi, N.S. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, 478, 378-381.

Gitelson, A.A., Kaufman, Y.J., and Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58, 289-298.

Gitelson, A.A., Kaufman, Y.J., Stark, R., and Rundquist, D. (2002a). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, 80, 76-87.

Gitelson, A.A., Merzlyak, M.N., and Chivkunova, O.B. (2001). Optical properties and non-destructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74, 38-45.

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

GoK (1965a). African socialism and its application to planning in Kenya. Sessional Paper No. 10 of 1965. Nairobi, Kenya: Government Printer.

GoK (1965b). The Forest Act (Cap 385) of the laws of Kenya. Revised Edition 1982 (1962) Government Printer, Nairobi. Kenya.

GoK (2007). Sessional Paper No. 1 of 2007 on Forest Policy. Ministry of Environment and Natural Resources, Nairobi.

GoK (2010). The Constitution of Kenya 2010. Nairobi, Kenya: Government Printer

GoK (2015). Fifth national report to the conference of parties to the convention on biological diversity. Available online: https://www.cbd.int/doc/world/ke/ke-nr-05-en.pdf (accessed on 16 June 2019).

GoK (2016a). National forest programme (2016-2030). Ministry of Environment and Natural Resources. Nairobi, Kenya: Government Printer.

GoK (2016b). The Forest Conservation and Management Act, 2016. Kenya Gazette Supplement No. 155 (Acts No. 34). Nairobi, Kenya: Government Printer.

GoK (2020). Forest Policy. Ministry of Environment and Forestry. Government Printer, Nairobi. Kenya.

Graça, P.M.L.A., Maldonado, F.D., Santos, J.R., and Keizer, E.W.H. (2015). Multi-temporal analysis of radiometric changes in satellite images of forest edges to infer selective-logging areas in the Amazon forest. *Acta Amazonica*, 45, 35-44.

Graves, S.J., Asner, G.P., Martin, R.E., Anderson, C.B., Colgan, M.S., Kalantari, L., and Bohlman, S.A. (2016). Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. *Remote Sensing*, 8, 161.

Grecchi, R.C., Beuchle, R., Shimabukuro, Y.E., Aragão, L.E.O.C., Arai, E., Simonetti, D., and Achard, F. (2017). An integrated remote sensing and GIS approach for monitoring areas affected by selective logging: A case study in northern Mato Grosso, Brazilian Amazon. *International Journal of Applied Earth Observation and Geoinformation*, 61, 70-80.

Guitet, S., Pithon, S., Brunaux, O., Jubelin, G., and Gond, V. (2012). Impacts of logging on the canopy and the consequences for forest management in French Guiana. *Forest Ecology and Management*, 277, 124-131.

Hall-Beyer, M. (2017). GLCM Texture: A Tutorial v. 3.0 March 2017. University of Calgary PRISM: University of Calgary's Digital Repository.

Hamunyela, E., Verbesselt, J., and Herold, M. (2016). Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sensing of Environment*, 172, 126-138.

Han, H., X. Guo, and H. Yu. (2016). Variable selection using mean decrease accuracy and mean decrease Gini based on random forest. In 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 219-224. Beijing, China.

Haralick, R. M., Shanmuga, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems Man and Cybernetics*, 3, 610-621.

He, D.C. and Wang, L. (1990). Texture unit, texture spectrum, and texture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 509-513.

He, H., and Garcia, E.A. (2009). Learning from imbalanced data. *IEEE Transactions* on *Knowledge* and *Data Engineering*, 21, 1263-1284.

Healey, J.R., Price, C. and Tay, J. (2000). The cost of carbon retention by reduced impact logging. *Forest Ecology and Management*, 139, 237-255.

Heiskanen, J., Korhonen, L., Hietanen, J., and Pellikka, P.K.E. (2015). Use of airborne lidar for estimating canopy gap fraction and leaf area index of tropical montane forests. *International Journal of Remote Sensing*, 36 (10), 2569-2583.

Hernández-Gómez, I.U., Cerdan-Cabrera, C.R., Navarro-Martínez, A., Vazquez-Luna, D., Armenta-Montero, S., and Ellis, E.A. (2019). Assessment of the CLASlite forest monitoring system in detecting disturbance from selective logging in the Selva Maya, Mexico. *Silva Fennica*, 53, 1.

Hethcoat, M.G., Edwards, D.P., Carreiras, J.M.B., Bryant, R.G., França, F.M., and Quegan, S. (2019). A machine learning approach to map tropical selective logging. Remote Sensing of Environment, 221, 569-582.

Hethcoat, M.G., Carreiras, J.M.B., Edwards, D.P., Bryant, R.G., Peres, C.A., and Quegan, S. (2020). Mapping pervasive selective logging in the south-west Brazilian Amazon 2000–2019. *Environmental Research Letters*, 15, 094057.

Hethcoat, M.G., Carreiras, J.M.B., Edwards, D.P., Bryant, R.G., and Quegan, S. (2021). Detecting tropical selective logging with C-band SAR data may require a time series approach. *Remote Sensing of Environment*, 259, 112411.

Hirschmugl, M.G., Gallaun, H., Dees, M., Datta, P., Deutscher, J., Koutsias, N., and Schardt, M. (2017). Methods for mapping forest disturbance and degradation from optical earth observation data: a Review. *Current Forestry Reports*, 3, 32-45.

Hoekman, D. (2016). Automated wide-area forest change monitoring with Sentinel-1 and TerraSAR-X. In: Proceedings of the GFOI R&D and GOFC-GOLD land cover Science meeting, October 31 - November 4, 2016, The Hague, The Netherlands.

Horler, D.N.H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4, 273-288.

Huang, C., Song, K., Kim, S., Townshend, J.R.G., Davis, P., Masek, J.G., and Goward, S.N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112, 970-985.

Hur, J.-H., Ihm, S.-Y., and Park, Y.-H. (2017). A variable impacts measurement in random forest for mobile cloud computing. *Wireless Communications and Mobile Computing*, 1-13.

IEA (2010). Devolution in Kenya: Prospects, challenge and the future. IEA Research Paper Series No. 24, Nairobi, Kenya.

Immitzer, M., Atzberger, C., and Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, 4, 2661-2693.

Jackson, C.M., and Adam, E. (2020). Remote sensing of selective logging in tropical forests: Current state and future directions. *iForest - Biogeosciences and Forestry*, 13, 286-300.

Jackson, C.M., and Adam, E. (2021). Machine Learning Classification of Endangered Tree Species in a Tropical Submontane Forest Using WorldView-2 Multispectral Satellite Imagery and Imbalanced Dataset. *Remote Sensing*, 13, 1-20.

Jenicka, S. and Suruliandi, A. (2011). Comparative study of classification algorithms with modified multivariate local binary pattern texture model on remotely sensed images. In: Proceeding of the International Conference on Recent Trends in Information Technology (ICRTIT, 2011). Chennai, Tamil Nadu, pp: 848-852.

Jenicka, S., Suruliandi, A. (2014a). A textural approach for land cover classification of remotely sensed image. *Computer Science and Information Technologies*, 2, 1-9.

Jenicka, S., Suruliandi, A. (2014b). Fuzzy texture model and support vector machine hybridization for land cover classification of remotely sensed images. *Journal of Applied Remote Sensing*, 8, 1-17.

Jenicka, S., and Suruliandi, A. (2015). Comparison of soft computing approaches for texture-based land cover classification of remotely sensed image. *Research Journal of Applied Sciences, Engineering and Technology*, 10, 1216-1226.

Janssen, L.L.F., and van der Wel, F.J.M. (1994). Accuracy assessment of satellite-derived land-cover data - A review. *Photogrammetric* Engineering *and* Remote Sensing, 60, 419-426.

Jensen, J.R. (2005). Introductory digital image processing: A remote sensing perspective, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2005; pp. 210–215.

Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15, 52-60.

Kairu, A., Upton, C., Huxham, M., Kotut, K., Mbeche, R., and Kairo, J. (2018). From shiny shoes to muddy reality: understanding how meso-state actors negotiate the implementation gap in participatory forest management. *Society & Natural Resources*, 31, 74-88.

Kamarulzaman, A., Wan Mohd Jaafar, W.S., Abdul Maulud, K.N., Saad, S.N.M., Omar, H., and Mohan, M. (2022). Integrated segmentation approach with machine learning classifier in detecting and mapping post selective logging impacts using UAV imagery. *Forests*, 13, 48.

Kankeu, R.S., Sonwa, D.J., Atyi, R., and Nkal, N. (2016). Quantifying post logging biomass loss using satellite images and ground measurements in Southeast Cameroon. *Journal of Forestry Research*, 27.

Karlson, M., and Ostwald, M. (2015). Remote sensing of vegetation in the Sudano-Sahelian zone: A literature review from 1975 to 2014. *Journal of Arid Environments*, 124, 257-269.

Katila, P., Pierce Colfer, C., De Jong, W., Galloway, G., Pacheco, P., and Winkel, G. (Eds.). (2019). Sustainable Development Goals: Their Impacts on Forests and People. Cambridge: Cambridge University Press.

Kavzoglu, T., and Mather, P.M. (2000). The use of feature selection techniques in the context of Artificial Neural Networks. In Proceedings of the 26th Annual Conference of the Remote Sensing Society (CD-ROM), Leicester, UK, 12-14 September 2000.

Kent, R., Lindsell, J., Laurin, G., Valentini, R., and Coomes, D. (2015). Airborne LiDAR detects selectively logged tropical forest even in an advanced stage of recovery. *Remote Sensing*, 7, 8348-8367.

KFS (2010). Mt. Kenya Forest Reserve Management Plan 2010-2019. Available online: http://www. kenyaforestservice.org/documents/MtKenya.pdf (accessed on 3 January 2019).

KFS and MENR (2007). Forest Law Enforcement and Governance in Kenya. A paper prepared for the East African Community-led Regional process in the framework of the Ministerial declaration, Yaounde, Cameroon, October 16 2003 on the Africa Forest Law Enforcement and Governance (AFLEG). Government of Finland, World Bank and World Conservation Union (IUCN).

Kigomo, B.N. (1987). The growth of camphor (*Ocotea usambarensis Engl*.) in plantation in the eastern Aberdare range, Kenya. *East African Agricultural and Forestry Journal*, 52, 141-147.

Kim, M.S. (1994). The use of narrow spectral bands for improving remote sensing estimation of fractionally absorbed photosynthetically active radiation (fAPAR). M.Sc. Thesis, College Park, MD, Department of Geography, University of Maryland. Available online: https://drum.lib.umd.edu/handle/1903/24796 (accessed on 23 May 2020).

Kim, K.I., Jung, K., Park, S.H., and Kim, H.J. (2002). Support vector machines for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 1542-1550.

Kim, S., Schreuder, G.F., Briggs, D., and Hinckley, T. (2009). Individual tree species identification using lidar-derived crown structures and intensity data. Doctoral thesis, 1-137. The University of Washington, College of Forest Resources.

Kleinschroth, F., Healey, JR, Sist, P., Mortier, F., and Gourlet-Fleury, S. (2016). How persistent are the impacts of logging roads on Central African forest vegetation? *Journal of Applied Ecology*, 53, 1-17.

Koltunov, A., Ustin, S.L., Asner, G.P., and Fung, I. (2009). Selective logging changes forest phenology in the Brazilian Amazon: evidence from MODIS images time series analysis. *Remote Sensing of Environment*, 113, 2431-2440.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, 5, 221-232.

Kuemmerle, T., Erb, K-H., Meyfroidt, P., Müller, D., Verburg, P., Estel, S., Haberl, H., Hostert, P., Jepsen, M., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, H., Zanden, E., and Reenberg, A. (2013). Challenges and opportunities in mapping land use intensity globally. *Current opinion in environmental sustainability*, 5, 484-493.

Kuntz, S. (2012). X-band SAR for forestry-TerraSAR-X. In: Proceedings of the GEOFCT 3rd Science and Data Summit (SDS3), 6-9 February 2012, Arusha, Tanzania.

Kupidura, P. (2019). The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. *Remote Sensing*, 11, 1-20.

Kuter, S. (2021). Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. *Remote Sensing of Environment*, 255, 112294.

KWS (2010). Mt Kenya Ecosystem Management Plan 2010–2020. Available online: http://www.kws.go.ke/file/1470/download?token=1lO6G3zI (accessed on 16 February 2019).

Lange, S., Bussmann, R.W., and Beck, E. (1997). Stand structure and regeneration of the subalpine *Hagenia abyssinica* of Mt. Kenya. *Botanica Acta*, 110, 473-480.

Langner, A., Miettinen, J., Kukkonen, M., Vancutsem, C., Simonetti, D., Vieilledent, G., Verhegghen, A., Gallego, J., and Stibig, H-J. (2018). Towards operational monitoring of forest canopy disturbance in evergreen rain forests: a test case in continental Southeast Asia. *Remote Sensing*, 10, 544.

Laporte, N.T., and Lin, T.S. (2003). Monitoring logging in the tropical forest of the Republic of Congo with Landsat imagery. *IEEE International Geoscience and Remote Sensing Symposium*, 4, 2565-2567.

Laurance, W.F. et al. (2014). Agricultural expansion and its impacts on tropical nature. *Trends* in *Ecology* and *Evolution*, 29, 107-116.

Lei, Y., Treuhaft, R., Keller, M., Dos-Santos, M., Gonçalves, F., and Neumann, M. (2018). Quantification of selective logging in tropical forest with spaceborne SAR Interferometry. *Remote Sensing of Environment*, 211, 167-183.

Li, J., and Wang G.A. (2013). Criminal Identity Resolution Using Personal and Social Identity Attributes: A Collective Resolution Approach. In Book: Intelligent Systems for Security Informatics, Editor(s): Yang, C., Mao, W., Zheng, X., Wang, H. Academic Press, pp. 107-124.

Li, H., Jing, L. and Tang, Y. (2017). Assessment of pansharpening methods applied to worldview-2 imagery fusion. *Sensors*, 17, 89.

Liao, S. and Chung, A.C.S. (2007). Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 15–20 April, Honolulu, HI, 1221-1224.

Liao, S., Law, W.K., Chung, C.S. (2009). Dominant local binary patterns for texture classification. *IEEE Transactions on Image Processing*, 18, 1107-1118.

Lima, T.A., Beuchle, R., Langner, A., Grecchi, R., Griess, V.C., and Achard, F. (2019). Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian Amazon. *Remote Sensing*, 11, 961.

Liu, D., Xia, F. (2010). Assessing object-based classification: Advantages and limitations. *Remote Sensing Letters*, 1, 187-194.

Logie, J.P.W., and Jones, G.A. (1968). Land use planning for forestry in Kenya. *East African Agricultural and Forestry Journal*, 33 (sup1), 59-62.

Long, A. (2013). REDD+, adaptation, and sustainable forest management: Toward effective polycentric global forest governance. *Tropical Conservation Science*, 6 (3), 384-408.

Lucieer, A., Stein, A., and Fisher, P. (2005). Multivariate texture-based segmentation of remotely sensed imagery for extraction of objects and their uncertainty. *International Journal of Remote Sensing*, 26, 2917-2936.

MacDicken, K.G., Sola, P., Hall, J.E., Sabogal, C., Tadoum, M. and, de Wasseige, C. (2015). Global progress toward sustainable forest management. *Forest Ecology and Management*, 352, 47–56.

Malahlela, O., Cho, M.A., and Mutanga, O. (2014). Mapping canopy gaps in an indigenous subtropical coastal forest using high-resolution worldview-2 data. *International Journal of Remote Sensing*, 35, 6397-6417.

Malczewski, J. (2016). Multicriteria Analysis. In: B. Huang and T.J. Cova (Editors) *Comprehensive Geographic Information Systems*. pp. 197-217. Oxford: Elsevier.

Marissiaux, Q. (2018). Characterizing tropical forest dynamics by remote-sensing using very high resolution and Sentinel-2 images. M.Sc. Thesis, Faculty of Bioengineers, Catholic University of Louvain, Belgium.

Martin, G.J. (2004). Ethnobotany; a Methods Manual. People and plants initiative, Earthscan London.

Masiliūnas, D. (2017). Evaluating the potential of Sentinel-2 and Landsat image time series for detecting selective logging in the Amazon. M.Sc. Thesis Report GIRS-2017-27, Wageningen University and Research. Available online: https://edepot.wur.nl/426144 (Accessed on 18 August 2020).

Matricardi, E.A.T., Skole, D.L., Cochrane, M.A., Qi, J., and Chomentowski, W. (2005). Monitoring selective logging in tropical evergreen forests using Landsat: multitemporal regional analyses in Mato Grosso, Brazil. *Earth Interactions*, 9, 1-24.

Matricardi, E.A.T., Skole, D.L., Cochrane, M.A., Pedlowski, M., and Chomentowski, W. (2007). Multi-temporal assessment of selective logging in the Brazilian Amazon using Landsat data. *International Journal of Remote Sensing*, 28, 63-82.

Matricardi, E.A.T., Skole, D.L., Pedlowski, M.A., Chomentowski, W., and Fernandes, L.C. (2010). Assessment of tropical forest degradation by selective logging and fire using Landsat imagery. *Remote Sensing of Environment*, 114, 1117-1129.

Matricardi, E.A.T., Skole, D.L., Pedlowski, M.A., and Chomentowski, W. (2013). Assessment of forest disturbances by selective logging and forest fires in the Brazilian Amazon using Landsat data. *International Journal of Remote Sensing*, 34, 1057-1086.

Maxwell, A.E., Warner, T.A., and Fang F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39, 2784-2817.

Mayaux, P., Holmgren, P., Achard, F., Eva, H., Stibig, H., and Branthomme, A. (2005). Tropical forest cover change in the 1990s and options for future monitoring. *International Journal of Biological Science*, 360, 373-384.

Mbuvi, M.T.E., and Kung'u, J.B. (2020). Policy makers' perspective on impacts of decentralizing forest management in Kenya on forestry conservation and community livelihoods. *Environmental Management and Sustainable Development*, 9, 47-67.

Melendy, L., Hagen, S.C., Sullivan, F.B., Pearson, T.R.H., Walker, S.M., Ellis, P., Kustiyo, Sambodo, A.K., Roswintiarti, O., Hanson, M.A., Klassen, A.W., Palace, M.W., Braswell, B.H., and Delgado, G.M. (2018). Automated method for measuring the extent of selective logging damage with airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 139, 228-240.

MENR (2016). National Forest Programme of Kenya, MENR, Nairobi, Kenya.

Menzies, N. (2002). Global gleanings: lessons from six studies of community-based forest management around the World. A report prepared for the Ford Foundation's Environment and Development Affinity Group (EDAG).

Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., and Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment*, 159, 307-317.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., and Rakitin, V.Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 105, 135-141.

Miettinen, J., Stibig, H.J., and Achard, F. (2014). Remote sensing of forest degradation in Southeast Asia-Aiming for a regional view through 5-30 m satellite data. *Global Ecology and Conservation*, 2, 24-36.

Mitchard, E.T.A. (2016). A Review of Earth Observation Methods for Detecting and Measuring Forest Change in the Tropics. *Ecometrica*, Edinburgh, UK, pp. 2-4.

Mitchell, A.L., Rosenqvist, A., and Mora, B. (2017). Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. *Carbon Balance Management*, 12, 9.

Miettinen, J., Stibig, H.J., and Achard, F. (2014). Remote sensing of forest degradation in Southeast Asia: aiming for a regional view through 5-30 m satellite data. *Global Ecology and Conservation*, 2, 24-36.

Monditoka, A.K. (2011). Decentralized forest governance - A policy perspective. Centre for Economic and Social Studies. Working Paper No. 93. RULNR Working Paper No. 8. Begumpet, Hyderabad. India.

Monteiro, A., Souza, C.M., and Barreto, P. (2003). Detection of logging in Amazonian transition forests using spectral mixture models. *International Journal of Remote Sensing*, 24, 151-159.

Morales-Barquero, L., Lyons, M.B., Phinn, S.R., and Roelfsema, C.M. (2019). Trends in remote sensing accuracy assessment approaches in the context of natural resources. *Remote Sensing*, 11, 2305.

Mugenda, O.M., and Mugenda, A.G. (2003). Research methods: Quantitative and qualitative approaches. African Centre for Technology Studies, Nairobi.

Mugo, E., Nyandiga, C., and Gachanja, M. (Eds.). (2010). Development of forestry in Kenya (1900-2007): Challenges and lessons learnt. Kenya Forest Working Group, Nairobi Kenya.

Mutanga, O., and Skidmore, A.K. (2004). Narrow Band Vegetation Indices Overcome the Saturation Problem in Biomass Estimation. *International Journal of Remote Sensing*, 25, 3999-4014.

Mutanga, O., Adam, E., and Cho, M.A. (2012). High-density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18, 399-406.

Nagendra, H., and Rocchini, D. (2008). High-resolution satellite imagery for tropical biodiversity studies: The devil is in the detail. *Biodiversity Conservation*, 17, 3431-3442.

Nakashizuka, T., Katsuki, T., and Tanaka, H. (1995). Forest canopy structure analyzed by using aerial photographs. *Ecological Research*, 10, 13-18.

Nature Kenya (2019). Ecosystem service assessment for the restoration of Mount Kenya Forest. Nature Kenya. Nairobi.

NCSS (2019). Chapter 435. Multidimensional scaling. Available online: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Multidimensional_Scaling.pdf (accessed on 18 November 2021).

Neba, G.S., Kanninen, M., Atyi, R., and Sonwa, D. (2014). Assessment and prediction of above-ground biomass in selectively logged forest concessions using field measurements and remote sensing data: A case study in South East Cameroon. *Forest Ecology and Management*, 329, 177-185.

Negrón-Juárez, R.I., Chambers, J.Q., Marra, D.M., Ribeiro, G.H.P.M., Rifai, S.W., Higuchi, N., and Roberts, D. (2011). Detection of subpixel treefall gaps with Landsat imagery in Central Amazon forests. *Remote Sensing of Environment*, 115, 3322-3328.

Nellemann, C. (2012). Green Carbon, Black Trade. Illegal Logging, tax fraud and laundering in the World's tropical forests. A Rapid Response Assessment. INTERPOL Environmental Crime Programme, Arendal: United Nations Environment Programme and GRID-Arendal, Birkeland Trykkeri AS, Spelefjellveien 1, 4760 Birkeland, Norway, pp.1

NEMA (2010). Kenya State of the Environment and Outlook 2010; Supporting the Delivery of Vision 2030. Available online: http://www.enviropulse.org/documents/Kenya_SOE.pdf (accessed on 3 January 2019).

NEMA. (2022). National Environment Management Authority (NEMA)—Towards 10 percent forest cover. https://www.nema.go.ke/index.php?option=com_content&view=article&id=306:towards-10-percent-tree-cover&catid=10&Itemid=462

Ng'eno, J.K. (1996). Kenya. In: Proceedings of Country Report to the FAO International Technical Conference on Plant Genetic Resources, Leipzig, Germany, 17-23 June *1996*.

Nyongesa, K.W., and Vacik, H. (2019). Evaluating Management Strategies for Mount Kenya Forest Reserve and National Park to reduce fire danger and address interests of various stakeholders. *Forests*, 10, 426.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2001) A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. *Advances in Pattern Recognition*, 399-408.

Ojala T., Pietikäinen M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 971-987.

Olofsson, P., Giles, M.F., Herold, M., Stehman, S., Curtis, E.W., and Wulder, M. (2014). Good Practices for Assessing Accuracy and Estimating Area of Land Change. *Remote Sensing of Environment*, 148, 42-57.

Omar, H. (2010). Commercial timber tree species identification using multispectral Worldview-2 Data. In D. Globe (Ed.): Geoinformation Program, Environment and Forestry Division Forest Research Institute Malaysia (FRIM).

Ongugo, P.O., Langat, D., Oeba, V.O., Kimondo, J.M., Owuor, B., Njuguna, J., Okwaro, G., and Russell, A.J.M. (2014). A review of Kenya's national policies relevant to climate change adaptation and mitigation: In from Mount Elgon. Working Paper 155. Bogor, Indonesia: CIFOR.

Osazuwa-Peters, O.L., Chapman, C.A., and Zanne, A.E. (2015). Selective logging: does the imprint remain on tree structure and composition after 45 years? *Conservation Physiology*, 3, cov012.

Ota, T., Ahmed, O.S., Minn, S.T., Khai, T.C., Mizoue, N., and Yoshida, S. (2019). Estimating selective logging impacts on aboveground biomass in tropical forests using digital aerial photography obtained before and after a logging event from an unmanned aerial vehicle. *Forest Ecology and Management*, 433, 162-169.

Otto, J-C., Prasicek, G., Blöthe, J., Schrott, L. (2018). GIS Applications in Geomorphology. *Comprehensive Geographic Information Systems*, 2018, 81-111.

Papeş, M., Tupayachi, R., Martínez, P., Peterson, A.T., Asner, G.P., and Powell, G.V.N. (2013). Seasonal variation in spectral signatures of five genera of rainforest trees. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 339-350.

Pearson, T. R. H., Bernal, B., Hagen, S. C., Walker, S. M., Melendy, L. K., and Delgado, G. (2018). Remote assessment of extracted volumes and greenhouse gases from tropical timber harvest. *Environmental Research Letters*, 13, 065010.

Peñuelas, J., Baret, F., and Filella, I. (1995). Semi-empirical indexes to assess carotenoids chlorophyll-a ratio from leaf spectral reflectance. *Photosynthetica*, 31, 221-230.

Pietikainen, M., Maenpaa, T., and Viertola, J., (2002), Color texture classification with color histograms and local binary patterns, Proceeding of International Workshop on Texture Analysis and synthesis, pp. 109-112. - Copenhagen, Denmark
Duration: 1-2 Jun 2002.

Pinagé, E.R., Matricardi, E.A.T., Leal, F.A., and Pedlowski, M.A. (2016). Estimates of selective logging impacts in tropical forest canopy cover using RapidEye imagery and field data. *iForest*, 9, 461-468.

Pinagé, E.R., Keller, M., Duffy, P., Longo, M., Dos-Santos, M.N., and Morton, D.C. (2019). Long-Term impacts of selective logging on Amazon forest dynamics from multi-temporal airborne lidar. *Remote Sensing*, 11, 709.

Pithon, S., Jubelin, G., Guitet, S., and Gond, V. (2013). A statistical method for detecting logging-related canopy gaps using high-resolution optical remote sensing. *International Journal of Remote Sensing*, 34, 700-711.

Poudyal, B.H., Maraseni, T., and Cockfield, G. (2018). Evolutionary dynamics of selective logging in the tropics: A systematic review of impact studies and their effectiveness in sustainable forest management. *Forest Ecology and Management*, 430, 166-175.

Pringle, M.J., Schmidt, M., and Muir, J.S. (2009). Geostatistical interpolation of SLC-off Landsat ETM+ images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 654-664.

Pu, R., and Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment*, 124, 516-533.

Putz, F.E., Dykstra, D.P. and Heinrich, R. (2000). Why poor logging practices persist in the tropics. *Conservation Biology*, 14, 951-956.

Qi, J., Wang, C., Matricardi, E., and Skole, D. (2002). Improved selective logging detection with Landsat images in tropical regions. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4, 2078-2080.

Qu, Y., Shaker, A., Silva, A.C., Klauberg, C., and Pinagé, E.R. (2018). Remote sensing of leaf area index from lidar height percentile metrics and comparison with MODIS product in a selectively logged tropical forest area in eastern Amazonia. *Remote Sensing*, 10, 970.

Rahm, M., Van Wolvelaer, J., Vrieling, A., and Mertens, B. (2013). Detecting forest degradation in the Congo Basin by optical remote sensing. In: Proceedings of the "ESA Living Planet Symposium." Edinburgh (UK), 9-13 September 2013. ESA, Noordwijk, The Netherlands, PP. 19.

Rakwatin, P., Longépé, N., Isoguchi, O., Shimada, M., and Uryu, Y. (2010). Mapping tropical forest using ALOS PALSAR 50m resolution data with multiscale GLCM analysis, *IEEE*

International Geoscience and Remote Sensing Symposium, 25-30 July 2010, Honolulu, Hawaii, USA. pp. 1234-1237.
Rauste, R., Antropov, O., Häme, T., Ramminger, G., Gomez, S., and Seifert, F.M. (2013). Mapping selective logging in tropical forest with space-borne SAR data. In: Proceedings of the "ESA Living Planet Symposium." Edinburgh (UK), 9-13 September 2013. ESA, Noordwijk, The Netherlands, pp. 168-173.
Read, J.M., Clark, D.B., Venticinque, E.M., and Moreira, M.P. (2003). Application of merged 1-m and 4-m resolution satellite data to research and management in tropical forests. *Journal of Applied Ecology*, 40, 592-600.
Rex, F., Silva, C., Paula, A., Corte, A., Klauberg, C., Mohan, M., Cardil, A., da Silva, V.S., de Almeida, D.R.A., Garcia, M., Broadbent E.N., Ruben Valbuena, R., Stoddart, J., Merrick, T., and Thomas Hudak, A.T. (2020). Comparison of statistical modeling approaches for estimating tropical forest aboveground biomass stock and reporting their changes in low-intensity logging areas using multi-temporal LiDAR data. *Remote Sensing*, 12, 1498.
Richards, J.A., and Jia, X. (1999). Remote sensing digital image analysis: An introduction, 3rd ed.; Springer-Verlag, Berlin/Heidelberg, Germany.
Robinson, E.J.Z., and Lokina, R.B. (2011). A spatial-temporal analysis of the impact of access restrictions on forest landscapes and household welfare in Tanzania. *Forest Policy and Economics*, 3, 79-85.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for landcover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
Runkle, J.R. (1981). Gap regeneration in some old-growth forests in the eastern United States. *Ecology*, 62, 1041-1051.
Runkle, J.R. (1992). Guidelines and sample protocol for sampling forest gaps. General technical report, PNW-GTR-283, USDA Forest Service. Pacific Northwest Research Station, Portland, Oregon, U.S.A. 44 pp.
Schreckenberg, K. Luttrell, C. and Moss, C. (2006). Participatory forest management: an overview. Projects: ARPIP: Participatory forest management and poverty and People orientated forestry.
Schmithüsen, F. (2002). Cross-sector linkages related to sustainable forest management, policy and legal aspects. Technical meeting on cross sector policy impacts between forestry and other sectors. FAO, Rome, Italy, 18 September 2002.
Senf, C., Wulder, M., Campbell, E., and Hostert, P. (2016). Using Landsat to assess the relationship between spatiotemporal patterns of western spruce budworm outbreaks and regional-scale weather variability. *Canadian Journal of Remote Sensing*, 42, 706-718.
Shahi, K., Shafri, H.Z.M., and Taherzadeh, E. (2014). A novel spectral index for automatic shadow detection in urban mapping based on WorldView-2 satellite imagery. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 8, 1769-1772.
Shao, Z., Zhou, W., Zhang, L., and Hou, J. (2014). Improved color texture descriptors for remote sensing image retrieval. *Journal of Applied Remote Sensing*, 8, 083584.
Shimabukuro, Y.E., Beuchle, R., Grecchi, R.C., and Achard, F. (2014). Assessment of forest degradation in Brazilian Amazon due to selective logging and fires using time series of fraction images derived from Landsat ETM+ images. *Remote Sensing Letters*, 5, 773-782.
Shimizu, K., Ahmed, O.S., Ponce-Hernandez, R., Ota, T., Win, Z.C., Mizoue, N., and Yoshida, S. (2017). Attribution of disturbance agents to forest change using a Landsat time series in tropical seasonal forests in the Bago Mountains, Myanmar. *Forests*, 8, 218.

SIFI (2010). Forest on the global political arena, In: (F. Inge-marson Ed.) international forest policy; an overview. Ake Barklund, KSLA.

Silva, C. A., Hudak, A. T., Vierling, L. A., Klauberg, C., Garcia, M., Ferraz, A., Keller, M., Eitel, J., and Saatchi, S. (2017). Impacts of airborne lidar pulse density on estimating biomass stocks and changes in a selectively logged tropical forest. *Remote Sensing*, 9, 1068.

Singh, M., Evans, D., Tan, B.S., and Nin, C.S. (2015). Mapping and characterizing selected canopy tree species at the Angkor World Heritage site in Cambodia using aerial data. *PLoS ONE*, 10, e0121558.

Slik, J.W.F., Arroyo-Rodrguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bel-lingham, P.J., and van den Berg, E., et al. (2015). An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences*, 112, 7472-7477.

Smyle, J., Collins, S., and Biason, C. (2016). Rethinking forest regulations: Overcoming the challenges of regulatory reform. Project: Rights and Resources Initiative.

Sofan, P., Vetrita, Y., Yulianto, F., and Khomarudin, M.R. (2016). Multi-temporal remote sensing data and spectral indices analysis for detection tropical rainforest degradation: a case study in Kapuas Hulu and Sintang districts, West Kalimantan, Indonesia. *Natural Hazards*, 80, 1279-1301.

Sokal, R.R. and Rohlf, J. (1987). Introduction to biostatistics. 2nd edition, Freeman and Company, New York, 363.

Solberg, R., Malnes, E., Amlien, J., Danks, F., Haarpaintner, J., Høgda, K-A., Johansen, B.E., Karlsen, S.R., and Koren, H. (2008). State of the art for tropical forest monitoring by remote sensing. In: A review carried out for the Ministry for the Environment of Norway and the Norwegian Space Centre. Norwegian Computing Centre, Oslo, Norway, pp. 11.

Somers, B., and Asner, G.P. (2012). Hyperspectral time series analysis of native and invasive species in Hawaiian rainforests. *Remote Sensing*, 4, 2510-2529.

Souza, C.M., and Barreto, P. (2000). An alternative approach for detecting and monitoring selectively logged forests in the Amazon. *International Journal of Remote Sensing*, 21, 173-179.

Souza, C.M., Firestone, L., Silva, L.M., and Roberts, D. (2003). Mapping forest degradation in the eastern Amazon from SPOT 4 through spectral mixture models. *Remote Sensing of Environment*, 87, 494-506.

Souza, C.M., Roberts, D., and Cochrane, M.A. (2005). Combining spectral and spatial information to map canopy damage from selective logging and forest fires. *Remote Sensing of Environment*, 98, 329-343.

Souza, C.M., Siqueira, J.V., Sales, M.H., Fonseca, A.V., Ribeiro, J.G., Numata, I., Cochrane, M.A., Barber, C.P., Roberts, D.A., and Barlow, J. (2013). Ten-year Landsat classification of deforestation and forest degradation in the Brazilian Amazon. *Remote Sensing*, 5, 5493-5513.

Spaias, L., Suomlainen, J., and Tanago, J.G.D. (2016). Radiometric detection of selective logging in tropical forest using UAV-borne hyperspectral data and simulation of satellite imagery. The 2016 European Space Agency Living Planet Symposium, Prague, the Czech Republic from 9-13 May 2016.

Stone, T.A., and Lefebvre, P. (1998). Using multi-temporal satellite data to evaluate selective logging in Pará, Brazil. *International Journal of Remote Sensing*, 19, 2517-2526.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9, 307.

Suruliandi A (2009). A study on classification of remotely sensed multispectral mages-a textural approach. Ph.D. dissertation work submitted to Manonmaniam Sundaranar University, Tamil Nadu, India.
Suruliandi, A., and Jenicka, S. (2015). Texture-based classification of remotely sensed images. *International Journal of Signal and Imaging Systems Engineering*, 8, 260-272.
Suruliandi, A., and Ramar, K. (2008). Local texture patterns - a univariate texture model for classification of images. In: Proceedings of the 16th International Conference on Advanced Computing and Communications, IEEE, held on 14-17 December 2008, Chennai, India. pp 32-39.
Tangki, H., and Chappell, N.A. (2008). Biomass variation across selectively logged forest within a 225-km^2 region of Borneo and its prediction by Landsat TM. *Forest Ecology and Management*, 256, 1960-1970.
Trichon, V., and Julien, M-P. (2006). Tree species identification on large-scale aerial photographs in a tropical rain forest, French Guiana-application for management and conservation. *Forest Ecology and Management*, 225, 51-61.
Tuominen, J., and Lipping, T. (2016). Spectral Characteristics of Common Reed Beds: Studies on Spatial and Temporal Variability. *Remote Sensing*, *8*, 181.
Unser, M. (1995). Texture classification and segmentation using wavelet frames. *IEEE transaction on Image Processing*, 4, 1549-1560.
UW Madison Libraries (2019). Top 10 databases. University of Wisconsin-Madison Libraries, WI, USA, website. [online] URL: http://www.library.wisc.edu/find/top-10-databases/
Vapnik, V. (2000). The nature of statistical learning theory, 2nd ed.; Springer-Verlag, New York, NY, USA.
Vepakomma, U., St-Onge, B., and Kneeshaw, D. (2008). Spatially explicit characterization of boreal forest gap dynamics using multi-temporal LiDAR data. *Remote Sensing of Environment*, 112, 2326-2340.
Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114, 106-115.
Verhegghen, A., Eva, H., and Achard, F. (2015). Assessing forest degradation from selective logging using time series of fine spatial resolution imagery in the Republic of Congo. In: Proceedings of the "IEEE International Geoscience and Remote Sensing Symposium (IGARSS)". Milan (Italy) 26-31 July 2015. IEEE Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 2044-2047.
Wagner, F.H., Ferreira, M.P., Sanchez, A., Hirye, M.C.M., Zortea, M., Gloor, E., Phillips, O.L., Filho, C.R.S., Shimabukuro, Y.E., and Aragão, L.E.O.C. (2018). Individual tree crown delineation in a highly diverse tropical forest using very high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 362-377.
Walubengo, D. (2007). Participatory forest management and the new forest legislation: Is it a promise or a reality? Proceedings of the 1st national participatory forest management conference, 6–8 June 2007, Kenya Forestry Research Institute (KEFRI) Headquarters, Nairobi, Kenya.
Wang, C., Qi, J., and Cochrane, M. (2005). Assessment of tropical forest degradation with canopy fractional cover from Landsat ETM+ and IKONOS Imagery. *Earth Interactions*, 9, 1-18.
Wang, T., Zhang, H., Lin, H., and Fang, C. (2016). Textural-spectral feature-based species classification of mangroves in Mai Po Nature Reserve from Worldview-3 imagery. *Remote Sensing*, 8, 24.
Wang, Y., Ziv, G., Adami, M., Mitchard, E., Batterman, S.A., Buermann, W., Marimon, B.S., Marimon-Junior, B.H., Reis, S., Rodrigues, D., and Galbraith, D. (2019). Mapping tropical

disturbed forests using multi-decadal 30 m optical satellite imagery. *Remote Sensing of Environment*, 221, 474-488.

Wass, P., (Ed.) (1995). Kenya's Indigenous Forests.' Status, Management and Conservation. IUCN, Gland, Switzerland, and Cambridge, UK. xii + 205pp.

Watrin, O.S., and Rocha, A.M.A. (1992). Levantamento da vegetacao natural e do uso da terra no municipio de Paragominas (PA) utilizando imagens TM/Landsat [Survey of natural vegetation and land use in the municipality of Paragominas (PA) using TM/Landsat]. *Boletim de Pesquisa* 124, EMBRAPA/CPATU, Bele´m, PA, Brazil, pp. 40.

Wedeux, B.M.M., and Coomes, D.A. (2015). Landscape-scale changes in forest canopy structure across a partially logged tropical peat swamp. *Biogeosciences*, 12, 6707-6719.

Whitehead, K., and Hugenholtz, C. H. (2014). Remote sensing of the environment with small unmanned aircraft systems (UASs), Part 1: A review of progress and challenges. *Journal of Unmanned Vehicle Systems*, 02, 69-85.

Whittle, M., Quegan, S., Uryu, Y., Stüewe, M., and Yulianto, K. (2012). Detection of tropical deforestation using ALOS-PALSAR: A Sumatran case study. *Remote Sensing of Environment*, 124, 83-98.

Win, R.N., Suzuki, R., and Takeda, S. (2012). Remote sensing analysis of forest damage by selection logging in the Kabaung Reserved Forest, Bago Mountains, Myanmar. *Journal of Forest Research*, 17, 121-128.

Wooller, M.J., Swain, D.L., Ficken, K.J., Agnew, A.D.Q., Street-Perrott, F.A., and Eglinton, G. (2003). Late Quaternary vegetation changes around Lake Rutundu, Mount Kenya, East Africa: Evidence from grass cuticles, pollen and stable carbon isotopes. *Journal of Quaternary Science*, 18, 3-15.

World Bank (1995). Participation in Forest and Conservation Management. Social Development Notes; No. 10. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/11668

World Bank (2001). A revised forest strategy for the World Bank Group. Draft 30 July 2001, World Bank, Washington DC.

Wulder, M.A., White, J.C., Coops, N.C., and Butson, C.R. (2008). Multi-temporal analysis of high spatial resolution imagery for disturbance monitoring. *Remote Sensing of Environment*, 112, 2729-2740.

Yuan, J., Wang, D., and Li, R. (2014). Remote sensing image segmentation by combining spectral and texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 16-24.

Zalazar, L.R.M. (2013). Use of lidar system for monitoring and management of forest resources. M.Sc. Thesis report, Environmental Science, Institute Gulich (CONAE/UNC)-FaMAF (UNC).

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., and Castro-Esau, K. (2006). Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. *Remote Sensing of Environment*, 105, 129-141.

Zhou, J., Yan Guo, R., Sun, M., Di, T.T., Wang, S., Zhai, J., and Zhao, Z. (2017). The effects of GLCM parameters on LAI estimation using texture values from Quickbird satellite imagery. *Scientific Reports*, *7*, 7366.

Appendix I: Questionnaire for Community Forest Associations

The questionnaire has been prepared for carrying out a survey. Please your cooperation in filling it up will go a long way in my academic pursuit and conservation of the forest. The information collected will be used for academic purposes only. Do not indicate your name or any form of identification. Information shared will be held in utmost confidentiality and anonymity.

Respondent's Demographics and General Information

Date:	Time:
Location:	Village:

1. Household composition

Status	Description	Age	Sex M/F)	Educational level	Livelihood
Head of household					
Spouse					
Member 1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Description: (1) Husband, (2) Wife, (3) Child, (4) Relative, (5) Orphan, (6) Visiting worker, (7) Dependent, and (8) Female head

Educational level: (1) no formal education, (2) Primary, (3) Secondary, (4) College/University

Livelihood: (1) No work, (2) Agriculture, (3) Student, (4) Business, (5) Wage laborer, (6) Salaried employee, (7) Infant, and (8) Other – specify

2. For how long have you been in this location?

Less than 1 year ( ) 1-5years ( ) 5-10 years ( ) 10 years or more ( )

3. Land ownership: Communal ( ) Private ( )

4. What do you use your land for?

Type of land	Size	% of the total land

Land type- (1) Cash crop plantation (2) Natural forest/woodland, (3) Woodlot, (4) Arable, (5) Wetland, (6) Grassland pasture

5. What is your position in the Community Forest Association (CFA)?

6. For how long have you been a CFA member?

7. Income per month in KES

1	<1000	
2	1000-5000	
3	5001-10 000	
4	10 001-15 000	
5	>15 000	

8. Water source ……………………………

9. Primary source of energy ………………………….

10. Distance to forest area from the village …………………... km

Historical Review

11. Compare the following factors as far back as you can remember? (Mark one alternative)

Factor		**Much decreased**	**Decreased**	**The same**	**Increased**	**Much increased**	**Don't know**
Climate	Temperature						
	Rainfall						
Vegetation	Grass						
	Bush						
	Trees						
	Bare Ground						
Water	Distance to source						
	Amount						
	Quality						
Soil Erosion							
Wildlife	Species						
	Numbers						
Population							
Livestock	Numbers						
	Rotation possibility						
Agriculture	Food production						
	Cultivated area						

Conservation of Mt. Kenya Forest Reserve

12. Do you agree/disagree with the following statements about the involvement of CFAs in the conservation of Mt. Kenya Forest Reserve?

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
C1. Products from MKFR are essential for the subsistence of the forest surrounding communities					
C2. Villagers extract products from MKFR					
C3. The forest offers protection to a variety of wildlife					
C4. Sustainable exploitation of the forest resources is good					
C5. You report the occurrence of illegal activities in MKFR					
C6. For ecological balance the forest requires protection					
C7. A concerted conservation effort is required to save the forest					
C8. People work relentlessly to preserve the forest					
C9. Your active role in conservation is important					
C10. You are willing to sacrifice more time toward conservation activities in MKFR					

Participation of CFA members based on the four stages of forest conservation and management

13. Do you agree/disagree with the following statements about the involvement of the CFAs in the planning, implementation, evaluation, and monitoring of Mt. Kenya Forest Reserve?

Stage/statement	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Planning					
P1. You are involved in making key decisions in MKFR conservation activities					
P2. Meetings are held regularly to plan MKFR conservation activities					
P3. You are involved in making MKFR management plans					
P4. Forest management plans are revised and updated frequently					
P5. The election of forest executive committee members is fair					
Implementation					
I6.You are regularly informed about forest conservation issues and activities					
I7. You have access to tools/equipment used in MKFR's conservation activities					
I8. You use extension services for forest management and conservation.					
I9. You are aware of rights/duties regarding law and regulations in forest management					
I10. Financial resources are at your disposal to implement forest conservation activities					

<u>Evaluation</u>					
E11. Good collaboration exists among CFAs in MKFR conservation activities					
E12. There is improved biodiversity in MKFR due to forest conservation activities					
E13. Increased forest products have improved people's livelihood					
E14. MKFR's conservation activities have led to more development in the community					
E15. You are willing to sacrifice more time to participate in conservation activities					
<u>Monitoring</u>					
M16. You monitor tree planting					
M17. You monitor illegal activities in MKFR					
M18. You monitor the protection of water sources in MKFR					
M19. Relevant technology is used in forest conservation and management in MKFR					
M20. Monitoring reports are made available to the public regularly					

Policy and institutional framework and conservation of MKFR

14. Do you agree/disagree with the following statements on policy and institutional framework and conservation of Mt. Kenya Forest Reserve?

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
P/I1. There are clearly stated and assigned user rights					
P/I2. Forest benefits are shared equitably					
P/I3. Collaboration with external institutions is good for the conservation of MKFR					
P/I4. Stakeholders are aware of the activities allowed and not allowed in MKFR					
P/I5. MKFR conservation activities are allocated sufficient funds					
P/I6. KFS has enough trained staff who assist in forest conservation activities in MKFR					
P/I7. Training is carried out frequently in support of MKFR's conservation activities					
P/I8. The current forest administration has minimized illegal activities in MKFR					
P/I9. Existing forest laws adequately deal with illegal activities in MKFR					
P/I10. Political interventions have solved the plight of the local community concerning the extraction of forest products					

Appendix II: An interview schedule for KFS, KWS, and NEMA Officials

1. How long have you worked at your workstation?

………………………………………………………………………………………………

2. In your view, do you think integrating the concept of community participation through Participatory Forest Management (PFM) approaches is a key move in championing forest conservation and management practices at the local level?

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

3. Which measures have you introduced to enhance participation by the local community in the conservation and management of forest activities?

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

4. Is the adjacent community accorded equal chance as other parties in the sharing of forest resources?

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

5. What concerns (different types of resources and the intensity of the problems encountered) does the community come across in accessing forest resources?

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
6. Has the idea of integrating adjacent communities in forest management activities enhanced the livelihood (economic, social, and cultural benefits) of the people in the area?
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
7. Compare the exploitation of forest resources by the community in the past and now?
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
8. In your opinion, how would you like forest resources to be exploited in the future?
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

9. Do you experience any managerial challenges because of integrating local communities in forest management and conservation practices?

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

10. In your view, what measures can the government put in place (at the local level) to enhance community awareness and enhance the efficiency of community participation in forestry conservation and management efforts?

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

11. What steps can the government take to enhance equitable, sustainable, and effective exploitation of resources in Mt. Kenya Forest Reserve?

………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

12. What lessons and experiences have you learned from involving adjacent communities in forest management and conservation?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Appendix III: An interview schedule for KFS Forest Guards

1. How long have you worked at your workstation?

……………………………………………………………………………………………

2. What illegal activities take place in Mt. Kenya Forest Reserve?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

3. How can you describe the relationship between the management of the forest and the adjacent community?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

4. Do you think the community is adequately involved in forest conservation and management activities by the government?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

5. Do you think the integration of adjacent communities has promoted forest conservation and management effort?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

6. What challenges do you encounter in your line of duty?

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

7. Which areas do you think the government can improve to enhance community participation and improve forest management?

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Appendix IV: Observation Checklist

State of user rights

User right	Present	Absent	Stage of implementation	Challenges observed
Grazing livestock on pastures of forest				
Grass harvesting				
Growing crops (nonresident cultivation)				
Collect fuel and dry wood				
Collect medicinal & aromatic herbs/plants				
Collect fruits (specify)				
Contracts for silvicultural activities				
Beekeeping on forest land				
Community-based wood and non-wood industries				
Other use (specify)				