8-2021

PREDICTING TROPICAL RAINFOREST DEFORESTATION USING MACHINE LEARNING, REMOTE SENSING & GIS: CASE STUDY OF THE CROSS RIVER NATIONAL PARK, NIGERIA

Ossai Alu

Predicting Tropical Rainforest Deforestation using Machine Learning

Remote Sensing & GIS: Case study of the Cross River

National Park, Nigeria

by

Ossai D. Alu

A thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Geographic Information Science

August, 2021

Thesis Committee:
Kirk Stueve (Chairperson)
Mikhail Blinnikov
William Cook

Abstract

Population growth, urban sprawl, agricultural expansion, and illegal logging has led to losses in forested land in most parts of the world, especially in a highly populated country like Nigeria. The Cross River National Park (CRNP) in southeastern Nigeria with an area just above 4000km^2 is designated a biodiverse hotspot and one of the oldest rainforests in Africa. As with all other tropical forests spread across the globe the CRNP is not immune to these factors that threaten its existence. The focus of this study is to analyze the change of forest cover at the Oban division of the Cross River National Park using multi-temporal remotely sensed data to predict and model the future probability of deforestation within the area of interest. This study made use of the Landsat West Africa Land Use/Land Cover Time Series dataset for the years 1975, 2000 and 2013 and Landsat 8 operational land imager (OLI) imagery for the year 2020 in a post classification change detection model to determine the extent of change in forest cover classes. Random forest decision tree machine learning algorithm was used to predict the future risk of forest cover loss using the datasets produced from the post classification change detection. The model related deforestation probabilities with several physical and anthropogenic factors such as elevation, slope angle, solar radiation, aspect, topographic roughness, soil type, distance from roads, distance from towns, distance from rivers, distance from plantations and population density. The results from the change detection analysis showed that from 1975 to 2020 the forest cover declined by 1909km^2 a rate of 42km^2 per year. The random forest regression analysis predicted areas of the forest with modest to high deforestation probabilities and indicated that socio-economic factors are major drivers of deforestation in the region rather than physical factors.

Table of Contents

Page

List of Tables 6

List of Figures..7

Chapter

1. Introduction 9

Deforestation ……………………………...10

Research Question...15

Study Objectives...15

2. Literature Review...17

Land Cover Change Detection using Remote Sensing and GIS...............................18

Deforestation Detection Research using Remote Sensing & GIS..............................21

Change Detection Techniques...22

Predictive Modelling of Land Use/Land Cover Changes 22

3. Methodology 27

Study Area 27

Oban Division 29

The Cross-River State Superhighway 30

Flowchart of the Methodology 31

Land cover change detection 33

Pre-processing...35

Image Processing...35

Image Classification Schemes……………………………………………………. 35

Post-Classification Comparison Change Detection …………………………..…. 37

Chapter Page

Accuracy Assessment……………………………………………………………...38

Forest Loss Risk Prediction ……………………………………………………….39

Predictor Variable Selection……………………………………...……………….41

Physical Variables...………………………....………………………………….41

Human/Anthropogenic Variables…………..…..……………………...………….46

Model Creation using Random Forest Regression Analysis ….……………………..52

Model Evaluation…………...……………………………………………………..52

4. Results 54

Image Classification and Change Detection of the CRNP 1975 - 2020 54

1975 Land Use Land Cover Analysis 54

2000 Land Use Land Cover Analysis……………………….................................55

2013 Land Use Land Cover Analysis……………………….................................57

2020 Land Use Land Cover Analysis……………………….................................58

Land Use/Land Cover Accuracy Assessment – 2020……………………..............60

Land Cover Change Detection within CRNP…………………………......................60

LULC change detection 1975 – 2000…………………..61

LULC change detection 2000 – 2013……………………..61

LULC change detection 2013 – 2020…………………..62

Deforestation probability assessment using random forest (RF) ……….……........63

Multi-collinearity test of independent variables……………………….................63

Deforestation Probability Analysis in the 1975-2000 period ………..........................64

Model Assessment ………………………...66

Deforestation Probability Analysis in the 2000-2013 period ……………..................68

Chapter Page

Model Assessment ………………………...69

Deforestation Probability Analysis in the 2013-2020 period …………….................71

Model Assessment ……………………………………….……………................73

5. Discussion …………………...76

Mapping the probability of forest loss …………………….….……................77

Model 1 (1975 to 2000) ………………...…………………….….……................77

Model 2 (2000 to 2013) ………………...…………………….….……................78

Model 3 (2000 to 2020) ………………...…………………….….……................79

Conclusion ………………...…………………….….……...................................81

Limitations and Future Studies ………………...…………………….….……......82

References 83

Appendix 89

List of Tables

Table Page

1. Major satellite remote sensing data available 19

2. General steps required to perform digital change detection 20

3. Change detection techniques ………………...22

4. Sample confusion matrix for map classification assessment 38

5. Land classes area and percentage of total area for year 1975.. 54

6. Land classes area and percentage of total area for year 2000...55

7. Land classes area and percentage of total area for year 2013...57

8. Land classes area and percentage of total area for year 2020...58

9. Significant class changes 1975-2000 ……………………………………………………61

10. Significant class changes 2000-2013 ..…………………………………………………61

11. Significant class changes 2013-2020 …………………..………………………………62

12. Multi-collinearity test data for independent variables …………………………………..63

13. Importance of independent variables – 1975 to 2000 ………………………………..…64

14. Risk categories and percentage of total area for model 1975 – 2000 …………………...67

15. Importance of independent variables – 2000 to 2013……………………..……………68

16. Risk categories and percentage of total area for model 2000 – 2013………………...…71

17. Importance of independent variables – 2013 to 2020…………………..………………72

18. Risk categories and percentage of total area for model 2013 - 2020……………………75

19. Deforestation probability statistics ……………………………………………………...75

List of Figures

Figure Page

1. Map showing global tropical rainforest regions in green ………………………………...10

2. 1966 Federal Surveys Map of Nigeria Forest Reserves…………………………………...13

3. Slash and burn farming...14

4. Illegal logging of the rain forest 14

5. The Cross River National Park showing the study area in green....................................... 28

6. Wildlife at the CRNP……………..……………………………………………………….29

7. Flowchart of methodology………..……………………………………………………….31

8. 1975 West Africa Land Use Land Cover Time Series dataset of the area of interest.........36

9. Response Variable Selection..40

10. Elevation of study area..42

11. Slope Angle raster of the Area of Interest …………………………..……….……….42

12. Slope Aspect raster ………………………………………………..…………………43

13. Solar Radiation raster ………………………………………………….…………..44

14. Soil type map for area of interest …………………………………………………….45

15. Topographic Roughness Index ………………………………..…………………….…46

16. Raster of distance from rivers ……………………………………………..…………...47

17. Raster of distance from plantations ……………………..……………………………48

18. Raster of distance from settlements ………………..…………………………………...49

19. Raster of distance from roads ………………….………………………………………50

20. Population Density raster of the area of interest ……………………………………..……51

21. Land cover classes - 1975 ……………………………………………………….……...54

22. Land classes of the area of interest and area in km2 - 1975 ……………………….……...55

Figure Page

23. Land classes of the area of interest and area in km^2 - 2000 ………………….……..…….....56

24. Land cover classes - 2000 ……………………………………………………….……...56

25. Land classes of the area of interest and area in km^2 - 2013 ……………………………57

26. Land cover classes - 2013 ……………………….…………...................................58

27. Land classes of the area of interest and area in km^2 - 2020 ………..…………….……......59

28. Land cover classes – 2020 ………………..……………………………..……….……...59

29. Land classes over a 45 - year period ……………….…….……………………...…...60

30. Change in area period 1975 – 2000 (km^2) ………………………………………..……...61

31. Change in area period 2000 – 2013 (km^2)………………………………….…….……...62

32. Change in area period 2013 – 2020 (km^2) ……………………..………..…….……...62

33. VIF Plot …………………………………..……………………………….……........64

34. Plot of importance of independent variables 1975-2000.………..……………….……...65

35. ROC curve and AUC for period 1975 - 2000 ………………..…………………….……...66

36. Deforestation probability model – 1975 to 2000 ……………….……………….…...67

37. Plot of importance of independent variables ……………………………….……...69

38. ROC curve and AUC for period 2000 - 2013 ……………………..………….……...70

39. Deforestation probability model – 2000 to 2013 ……………………..…….……..70

40. Plot of importance of independent variables ………………..…………….……........72

41. ROC curve and AUC for period 2013 - 2020 …………………..……………….……...73

42. Deforestation probability model – 2013 to 2020 ……………….………………….…...74

Chapter 1: Introduction

Biodiversity or biological diversity describes the enormous variety of life on Earth, it is also used to refer to every living thing including plants, bacteria, animals, and humans in one region or ecosystem (National Geographic Society, 2020).

Biodiversity is not uniform all over the planet, with areas having extremely high levels of biodiversity called hotspots. Some examples of hotspots include Brazil (Amazon), South Africa and Madagascar. Biodiversity plays an important role in sustenance and life cycle of organisms on the planet, it is also an essential part of the solution to climate change, economic growth, and cultural identity of nations (Brooks et al., 2002).

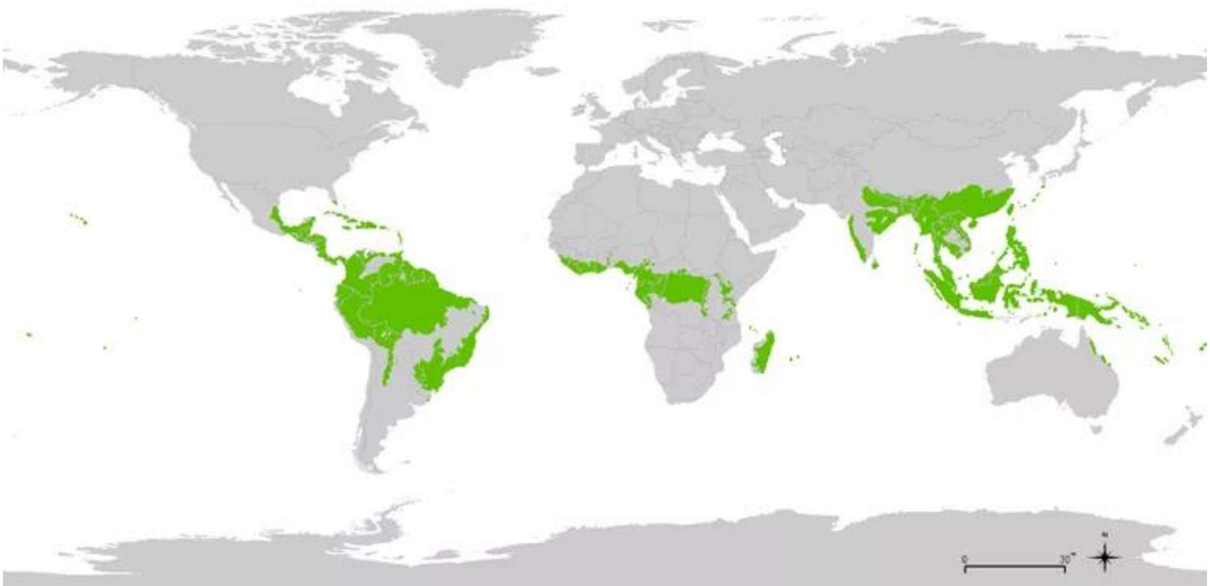
Nearly half of the world's vascular plant species and a third of terrestrial vertebrates are endemic to 25 "hotspots" of biodiversity. None of these hotspots have more than a third of their pristine habitat remaining, historically they covered 12% of the lands surface, but today their intact habitat covers only 1.4% of the land (Brooks et al., 2002). Forests are a large part of the planets biodiverse ecosystem and are vital for maintaining ecosystem balance. Although they cover less than 30% of the earth's surface, forests house an estimated two thirds of all life on the planets land mass (National Geographic Society, 2020).

Myers (1980), describes tropical forests as evergreen, or at least partly evergreen, forests, in areas receiving not less than 100 mm of precipitation in any month for two out of every three years, with mean annual temperature of 24-plus °C. These forests are exceptionally important by virtue of (a) their biodiversity, and (b) their carbon stocks with all that implies for global climate systems. They are important for many other reasons too, notably their special hardwood timbers and their watershed functions. In terms of their biodiversity, they altogether harbor more species than the rest of the world put together.

By far the greatest concentration of tropical rainforest is in the Amazon basin, this huge area of forest covers northwestern Brazil and stretches into the neighboring countries of Colombia, Ecuador, Peru, and Venezuela accounts for two thirds of the world's tropical rain forest. Asia has the next largest area mostly in Indonesia, Malaysia, and Papua New Guinea. Africa has about 15% of the total, stretching from the Guinean forest in the southern part of West Africa to Congo Basin in Central Africa.

Figure 1

Map Showing Global Tropical Rainforest Regions in Green



Deforestation and Deforestation Hotspots

The Food and Agriculture Organization (F.A.O) defines deforestation broadly as the conversion of forested areas (whether human induced or not) to non-forest land use such as arable land, urban use, logged area or wasteland. (FAO, 2007).

Research by (Myers, 1993) considered demographic, economic, institutional, and political factors that influence human demand for forest resources and developed several criteria in a model to project areas around the world that would mostly likely suffer significant deforestation due to progressive expansion of human demand. These criteria include:

- Demographic criteria: High population growth rates putting pressure on the exploitation of natural resources.
- Socio-economic criteria: Areas with prevalent or high levels of peasant poverty, landlessness, unemployment, and lack of rural development puts pressures on the exploitation of forest resources.
- Migration criteria: Refugee migrations between countries or internal migration due to land loss or land use changes is a criterion that can put pressure on forest resources.
- Infrastructure criteria: Rapid and unplanned infrastructure expansion are inimical to tropical forest conservation.
- Market-place criteria: demand driven commercial logging is one of the largest factors of deforestation in most areas
- Political criteria: areas targeted for settlement through political demarches or rulers looking to make political statements
- Physical criteria: Areas prone or susceptible to forest burning
- Atmospheric-pollution criteria: this criterion is especially pertinent in areas that may turn out to be predisposed to acid-rain injury, notably logged-over and slash-and-burn areas: a forest that is already stressed by logging or patchy burning may well prove less capable of resisting acid rain than one which remains intact.
- Climatic criteria: areas that may experience forest changes due to reduced moisture recycling
- Compounded impact criteria: for areas that feature two or more threats and may thereby become susceptible to synergized interactions whereby one factor's impact tends to reinforce that of another factor in multiplicative rather than additive fashion.

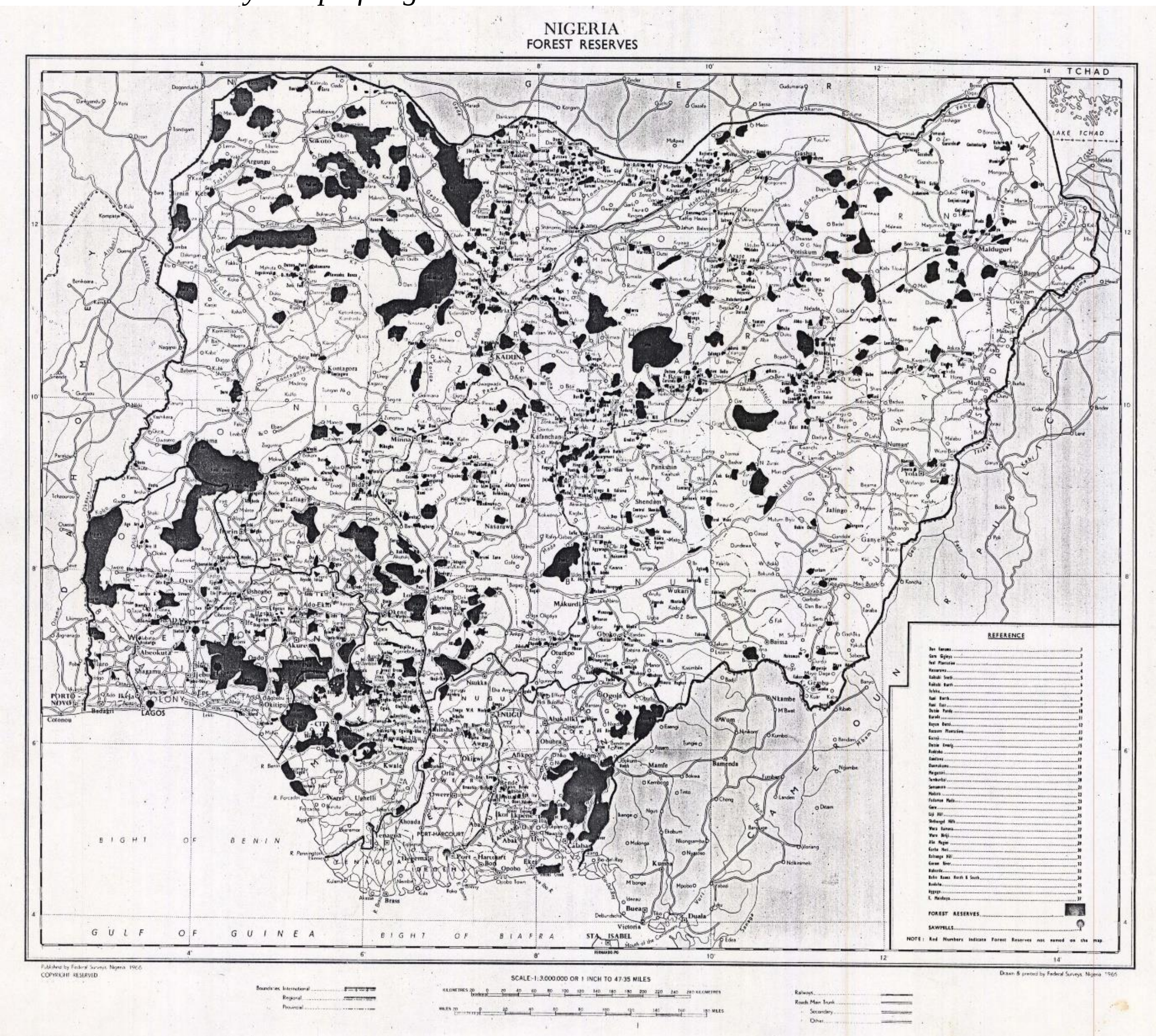
Following the identification of the ten (10) criteria for detecting deforestation hotspots on a temporal and spatial model, (Myers, 1993) declared the following countries/areas as tropical hotspots or deforestation fronts:

1. Southern Mexico
2. Central America
3. Colombian Choco
4. Western Amazonia
5. Southern and Eastern Amazonia in Brazil
6. Northern Bolivia
7. ***Eastern Nigeria and Southwestern Cameroon***
8. Madagascar
9. Eastern Myanmar (Burma)
10. Northern and Northeastern Thailand
11. Vietnam
12. Eastern Malaysia
13. Northern Sumatra and East and South Kalimantan (Borneo) in Indonesia
14. Philippines

Nigeria is home to several national forests and parks with an estimated total area of forest reserves put at 10 million hectares, which is about 10% of the total land area of the country. According to the World Resources Institute, Nigeria is home to 4,715 different types of plant species, and over 550 species of breeding birds and mammals, making it one of the most ecologically vibrant places on the planet., however, with this immense biodiverse wealth the country has had a poor record managing forest resources, the F.A.O in 2010 lists Nigeria as having one of the highest deforestation rates in the world.

Figure 2

1966 Federal Surveys Map of Nigeria Forest Reserves



In its feature report dated September 2, 2010, titled “Battling to preserve Nigeria’s rainforest” by Matthew Knight for the Cable News Network (CNN) a forest activist Odigha Odigha was quoted as saying that "The rate of deforestation in Nigeria is quite frightening, what is left is less than ten percent of the whole country and more than 50 percent of that is in Cross River State."

Nigeria is estimated to have lost/cleared about 410,000 ha of tropical forest land per annum between 1990 – 2005 which is approximately 4% of the country’s tropical forest disappearing every year (Juergen Blaser, 2006), activities such as illegal and legal logging,

slash and burn farm clearing for large scale plantations as we all as small holder farming has put a strain on this resource. Rather than these activities abating, they seem to be growing larger as the years roll by, leading to more forest land disappearing in the southern part of Nigeria.

This apparent and uncontrolled grab for natural resources in most African countries, especially Nigeria has led to wanton destruction of the natural ecosystem leaving the country very vulnerable to various natural disasters such as erosion, desertification, and land degradation. As such it is important that studies must be carried to ascertain the extent of deforestation and its likely impact.

Figure 3

Slash and Burn Farming - Courtesy Mongobay.com

Figure 4

Illegal Logging of The Rain Forest - Courtesy Guardian Nigeria 09/23/2020

Drawing on present and historical remote sensing imagery as well as GIS mapping tools for modelling and analysis, data sets will be processed, analyzed, and classified to produce temporal maps and models to assess the spatial extent of the Cross River National Park (CRNP) over time. The methodology adopted for this research includes a time series analysis of remotely sensed images (mostly originating from Landsat platforms) with adequate radiometric and geometric corrections carried out.

Datasets such as elevation, soil type, topographic roughness, slope angle, aspect, solar radiation, and human impact assessments of the AOI will be employed in the research. The use of R statistical programming language as well as Python to provide predictive models would be employed. Image and GIS analysis would be carried out using ArcGIS and Erdas Imagine. Ultimately with this research, I hope to provide a study to be used by local government agencies, planners, and conservationists in their work, protecting this natural treasure called the Cross River National Park.

Research Question:

The overarching question for this study is: *What physical and human factors influence deforestation in Cross River National Park, Nigeria?* In terms of the physical factors, it relates to the topographic, geological, pedologic and limnologic aspects of the landscape while the human/anthropogenic effects refer to such factors as population density, presence of roads, rivers, large plantations, and urban expansion.

Study Objectives

To answer the overarching question a few objectives or goals need to be met, these objectives cover several parts of the study and as such are vital in achieving the results of this study, they include:

1. Determination of land use change of the area of interest showing the areas most affected, spatial diffusion of the change and the rate of change.
2. Determination of significant factors influencing forest cover change.
3. Determine the future probability of forest change using machine learning modelling techniques.

Chapter 2: Literature Review

Biophysical materials and human-made features on the surface of the earth are subject to change over time, though not all features change with time but those that do change tend to experience either steady or dynamic changes over time. It is important that these changes be monitored accurately so that the physical and human processes at work can be fully understood, and negative effects mitigated (Jensen, 2015).

Humans in their quest to satisfy their wants and needs since existence have always modified the land, and as time has gone by so has the need for food and other essentials gone up due to increased population as such LULCC has become more apparent and unprecedented. The rapid changes have led to such global concerns as climate change, biodiversity loss, soil, air, water pollution and natural disasters such as flooding, erosion, and desertification globally.

Land use and land cover change (LULCC) is a general term for the human modification of Earth's terrestrial surface (Ellis, 2007). Also, Sealey et al., (2018) describe land cover change as the loss of natural areas, particularly loss of forests to urban or exurban development, or the loss of agricultural areas to urban or exurban development.

Land cover refers to the physical and biological cover over the surface of land, including water, vegetation, bare soil, and/or artificial structures while the definition of land use may depend on which branch of science you tend towards, from natural science land use is defined in terms of syndromes of human activities such as agriculture, forestry and building construction that alter land surface processes including biogeochemistry, hydrology, and biodiversity. In the field of social science and land management land use is define more broadly to include the social and economic purposes and contexts for and within which lands

are managed (or left unmanaged), such as subsistence versus commercial agriculture, rented vs. owned, or private vs. public land.

Land Cover Change Detection using Remote Sensing and GIS

Monitoring the locations and distributions of land-cover changes is important for establishing links between policy decisions, regulatory actions, and subsequent land-use activities (Lunetta et al., 2006).

Deforestation as earlier defined is the change of intended land use from forest to non-forest (urban, agricultural, etc.) (FAO, 2010), as such it can be regarded as a form or consequence of land cover change. This change needs to be studied to mitigate negative effects to the environment, therefore, due to technological advancement remote sensing (RS) and geographic information systems (GIS) are 2 important tools for the study and analysis of land cover change.

Remote sensing products have been a key source of information for monitoring land cover changes in the past decades (Lunetta et al., 2002). In remote sensing, change detection has been defined as the identification and location of changes in radiance values between multi-temporal images (Wang, 1993). Moreover, these products can be considered as the single feasible way of consistently monitoring changes in forest cover over time for large geographical regions (Shimabukuro et al., 2014).

Ideally, the following image characteristics are required for studying deforestation:

- Cloud free and clear atmosphere during the time of data acquisition.
- Availability of imagery for the optimum date or dates.
- Spatial resolution fine enough for accurate mapping and course enough so image size is manageable.

- Band selection (band width, placement, and number of bands) optimized to identify features of interest.
- Affordable
- Study area covered on a single image
- Multi-date imagery acquired under identical conditions
- Water level (tide, river, and lake level)
- Consistent and “stable” phenologic state
- Soil moisture
- Repeat interval consistent with project goals
- Same sensor and sun position when images were acquired
- Similar atmospheric conditions

These are all ideal characteristics for the acquisition of satellite imagery, unfortunately getting imagery under these conditions is not always possible, but it does provide a standard to which all image acquisitions should follow.

Table 1

Major Satellite Remote Sensing Data Available (FAO, 2007)

Name	**Spatial Resolution**	**Spectral Resolution**	**Temporal Resolution (Days)**	**Launch (Year)**	**Website**
Landsat	15-80	V/NIR, SWIR, TIR	16	1972	www.landsat.org
SPOT	2.5-20	V/NIR, SWIR	26	1986	www.spotimage.fr
IRS	6-188	V/NIR, SWIR	24	1995	www.isro.org
IKONOS	1-4	V/NIR, SWIR	3	1999	www.spaceimaging.com
QuickBird	0.61-2.44	V/NIR, SWIR		2002	www.digitalglobe.com/

Table 1 (continued)

MODIS	250-1000	V/NIR, SWIR, TIR	1	1999	modis.gsfc.nasa.gov/
VEGETATION	1000	V/NIR, SWIR, TIR	1	1998	www.spot-vegetation.com/
AVHRR	1000	V/NIR, SWIR, TIR	1	1978	www.noaa.gov
MERIS	300	V/NIR, SWIR	3	2002	www.envisat.esa.int
ASTER	15-90	V/NIR, SWIR, TIR	4-16	1999	asterweb.jpl.nasa.gov/
Hyperion	30	V/NIR, SWIR	16	2000	eo1.usgs.gov
ALI	10-30	V/NIR, SWIR	16	2000	eo1.usgs.gov
CBRES	20-260	V/NIR, SWIR	3-26	2003	www.dgi.inpe.br/
JERS/SAR	18	L-band	44	1992	www.eorc.jaxa.jp
Radarsat	8-100	C-band	24	1995	www.rsi.ca

Table 2

General Steps Required to Perform Digital Change Detection, Jensen (2015)

State the nature of the change detection problem	• Specify thematic attribute(s) or indicator(s) of interest. • Specify change detection geographic region of interest (ROI) • Specify change detection time- period (e.g., daily, seasonal, yearly) • Define the classes of interest in a classification system • Select hard and/or fuzzy change detection logic • Select per-pixel or object-based (OBIA) change detection

Table 2 (continued)

Consideration of significance when performing change detection	• Remote sensing system considerations: spatial, spectral, temporal, and radiometric resolution • Environmental considerations: Atmospheric conditions, soil moisture, Phenology, Obscuration conditions, Tides
Process remote sensor data to extract change information	• Acquire appropriate change detection data • Preprocess the multiple-date remote sensor data • Select change detection algorithm • Apply appropriate image classification logic if necessary • Perform change detection using GIS algorithms if necessary
Perform accuracy assessment	• Select method: Qualitative or Statistical • Determine number of samples required by class • Select sampling scheme • Obtain ground reference test information • Create and analyze change detection error matrix
Accept or reject previously stated hypothesis	

Deforestation Detection Research using Remote Sensing & GIS

In the Brazilian Amazon, according to (Souza et al., 2013), remote sensing has been used for mapping selective logging, from local to regional scales, and the approaches range from visual interpretation to automated techniques.

Mengistu & Salami (2007) used Landsat products coupled with GIS techniques to track land use/land cover conversion and modification over large areas in southwestern Nigeria.

Deng et al., (2008), utilized multi-temporal and multi-sensor data (SPOT & Landsat) to identify and quantify land use changes in an urban environment based the principal component analysis (PCA) and hybrid classification methods.

Change Detection Techniques

There are several land cover change detection techniques currently in use within the field of remote sensing. (Jensen, 2015), divides these techniques into 2 groups which are the binary change detection methods which provide "change/no-change" information and the thematic change detection methods which provide "from-to" information.

Table 3

Change Detection Techniques

Binary Change Detection "Change/No-Change" Information:	**Thematic Change Detection "From-To" Information**
• Analog "On-Screen" Visualization Change Detection • Image Algebra Change Detection: band ratioing or band differencing • Multiple-Date Composite Image Change Detection: Supervised/Unsupervised classification, Principal Component Analysis (PCA) • Continuous Change Detection and Classification using Landsat data	• Photogrammetric Change Detection • LiDARgrammetric Change Detection • Post-Classification Comparison Change Detection: Per-Pixel or Object bases image analysis (OBIA) • Neighborhood Correlation Image (NCI) Change Detection • Spectral Change Vector Analysis Change Detection

Predictive Modelling of Land Use/Land Cover Changes

Land use and land cover (LULC) models are essential for the analysis of LULC changes and predicting land use requirements and are valuable for guiding reasonable land use planning and management. There are several models utilized for LULC predictive

modelling, some more popular than others, however each model comes with its own advantages and constraints as such the choice of which model to use is an important decision.

Land cover change modeling means time interpolation or extrapolation when the modeling exceeds the known period (Aitkenhead & Aalders, 2009). Commonly used models for estimating land cover changes are analytical equation-based models (Shamsi, 2010), statistical models (Ralha et al., 2013), evolutionary models (Hyandye, 2015), cellular models (Yang et al., 2012), Markov models (Singh et al., 2015), hybrid models (Subedi et al., 2013), expert system models (Yang, 2002) and multi-agent models (Guan et al., 2011).

The cellular automata – Markov model is one of the commonly used among many LULC modelling tools and techniques, which models both spatial and temporal changes. It has been used in various LULC research with a proven track record (Auwalu et al., 2020).

The robust nature of the CA-Markov model in simulating land use changes is made apparent with the work of (Linling et al., 2011). CA–Markov is a combined Cellular Automata/Markov Chain/Multi-Criteria/Multi-Objective Land Allocation (MOLA) land cover prediction method that adds an element of spatial contiguity as well as knowledge of the likely spatial distribution of transitions to Markov chain analysis. It successfully simulated land use changes in the Fangshan region of China, using land use maps (2001, 2006 and 2008) of the area.

Chen et al., (2018) in the use of the CA-Markov model made use of a hybrid model which combines the long-term predictions of the Markov model and the ability of the Cellular Automata (CA) model to simulate the spatial variation in a complex system using temporal Landsat imagery for the years 1992, 2003 and 2014. This applied in a GIS produced a predicted land use map for the year 2014 and was validated by actual land use results and was

further used to predict land use trends for 2025 and 2036 for Jiangle County in Fujian Province, China.

These studies considered amongst others, utilize a limited number of potential independent factors having normal distributions, which is a basic requirement for using parametric techniques. However, there is a need for studies that include many factors encompassing, as much as possible, all aspects of the socioeconomic (SE) and bio geophysical (BGP) context within which LULCC is taking place (Rather et al., 2020).

This study will be concerned with the predictive power of Random Forest (RF) or Random Decision Forest machine learning model in LULC changes such as deforestation. Random Forest is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees (Ho, 1995).

Despite the improvements in our understanding of the impacts of LULCC on tropical environments, there is still no optimal tool for understanding relationships between deforestation (DF) metrics and SE or BGP factors. Random Forest analysis (Breiman, 2001) is a variable selection technique and has great potential in this respect. RF can identify complex interactive and non-linear response-predictor relationships and has excellent predictive performance (Prasad et al., 2006; Smith et al., 2011). Thus, application of RF analysis to disentangle these sorts of relationships may be particularly useful. RF is used widely in bioinformatics (Cutler & Stevens, 2006), for land cover classification (Gislason et al., 2006) and analysis of medical experiments, for example. Although there were few ecological applications, it has recently gained popularity in this area (Prasad et al., 2006; Fu et al., 2010; Gilbert & Chakraborty, 2011).

A study by (Zanelle et al., 2017) where RF was used to identify relationships between derived Deforestation (DF) and Forest Fragmentation (FF) metrics to Socioeconomic (SE) and Bio-Geophysical (BG) factors in the Brazilian Atlantic Forest of Minas Gerais, Brazil, Random Forest was used to identify relationships between the sets of variables and its performance was compared to traditional multiple linear regression models. The study found that RF modelled relatively-well variance in all metrics used; (the rate of deforestation, the amount of forest and the density and isolation of forest patches), presenting a better performance when compared to the classical approach. RF also identified geographical location and topographic factors as being most closely associated with patterns of DF and FF metrics. RF was better at explaining variations in rates of deforestation, remaining forest, and patch patterns, than the multiple linear regression approach.

Saha et al., (2020) assessed three models of which one is a probabilistic model Binary Logistic Regression (BLR), one is machine learning model Random Forest (RF), and another is hybrid ensemble model Rotational Forest and Reduced Error Pruning Trees (RTF-REPTree). Twelve independent variables (slope, altitude, aspect, distance from forest edges, forest density, distance from river, distance from settlement, distance from road, population density, settlement density, proximity to agricultural land, agricultural land density) have been selected for modelling and evaluation of the performance of the models. Before using these factors, multi-collinearity test has been performed that suggests the independence and suitability of the variables. The results showed that all models BLR, RF & RTF-REPTree have good capability of assessing deforestation probability, among the RTF-REPTree had the highest accuracy level followed by random forest (RF).

Another RF study useful in this research involves the work by (Cushman et al., 2017) where RF and logistic regression are compared in a multi scale approach to model forest risk

between 2000 – 2010 as a function of topographical variables and landscape structure, and application of the highest performing model to predict the spatial pattern of forest loss risk between 2010 and 2020. The study found RF significantly outperformed logistic regression and the inclusion of landscape structure variables significantly improved the prediction. (Grinand et al., 2020), used machine learning algorithm and high-resolution imagery to provide spatial forecasts of deforestation, land degradation and regeneration over a 17,000 square kilometer region of Madagascar. Response variables were selected using a stratified random sampling scheme and empirical studies conducted to select predictor variables. Random Forest gave a better predictive result compared to other predictive techniques with elevation and distance to forest edge being the most important predictor variables.

Chapter 3: Methodology

Study Area

Cross River National Park (CRNP) (figure 4) is located between Latitude 5˚ 05' and 6˚ 29' N, and Longitudes 8˚ 15' and 9˚ 30' E, in the extreme south-eastern corner of Nigeria, in Cross River State. It covers a total area more than 4000 km^2, consisting of primary moist tropical rainforest ecosystem in the North and central parts, and montane mosaic vegetation on the Obudu Plateau. It is Nigeria's last Great Rainforest Reserve, and the closest to the Mangrove Swamps on the coastal region.

The CRNP is managed by the Nigeria National Parks Service an agency of the Federal Ministry of Environment and along with Korup National Park in the Republic of Cameroon is an important biotic reserve which contains one of the oldest rainforests in Africa. It is also one of the 25 United Nations acclaimed biodiversity hot spots in the World. Some portions of the park lie in the Guinea-Congolian region of the lowland rainforest refugia with closed canopy and scattered emergent trees which reach a height of between 40 and 50 meters. The Park exists in two distinct, non-contiguous divisions; Oban and Okwangwo (NNPS, 2020).

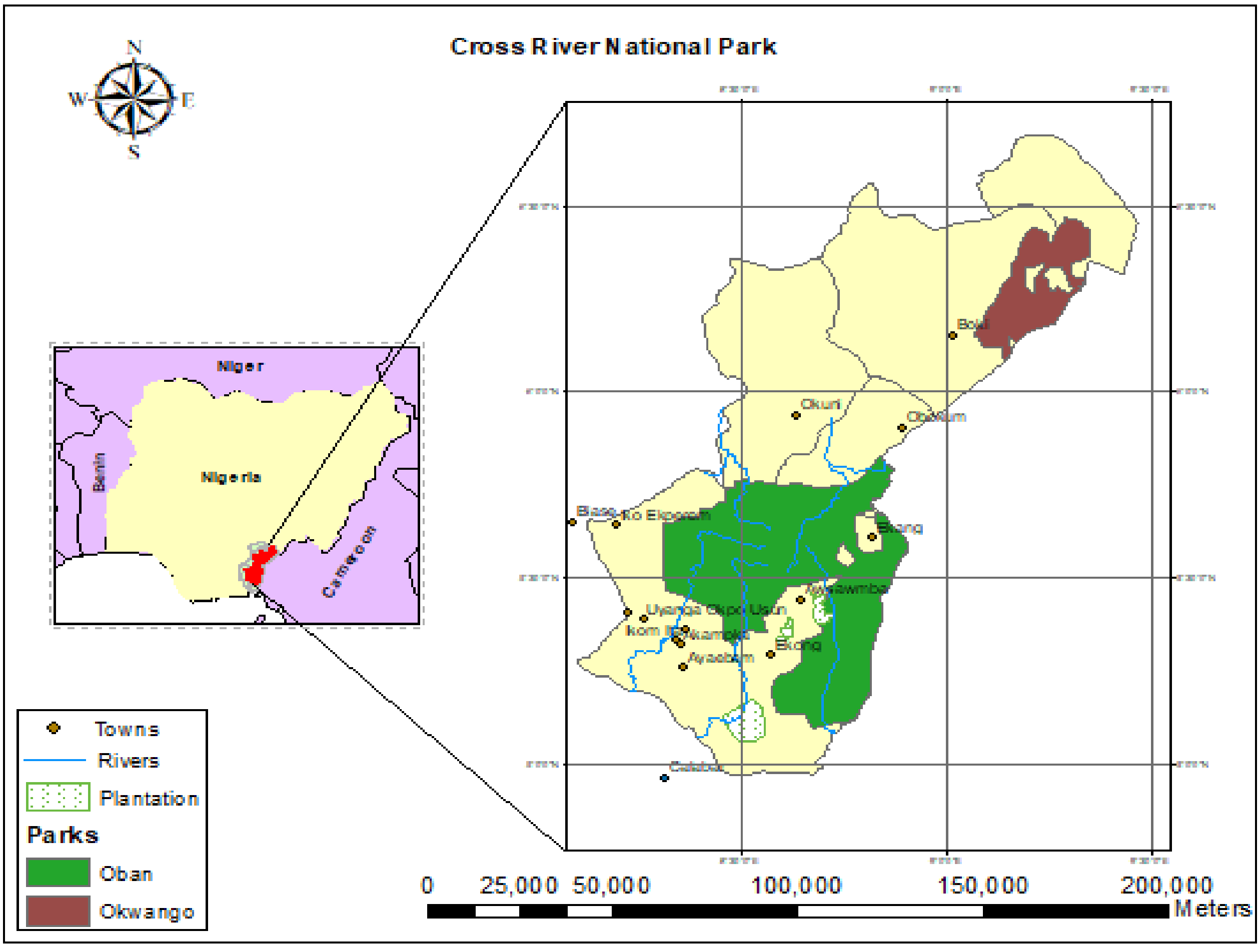
The Park contains large areas of lowland rain forest (covering all of Oban Division and part of Okwangwo Division) as well as an unbroken elevational gradient of lowland to submontane forest in the Okwangwo Division. This gradient rises from 150 m above sea level in the valleys of Cross River tributaries to 1,700 m on the edge of the Obudu Plateau. Parts of the central Oban Hills rise above 500 m, with one peak reaching approximately 1000m. Although the Obudu Plateau itself has high rainfall, the Okwangwo Division in general has lower annual rainfall than Oban, with a longer dry season and therefore a different forest structure. Among the many biologically significant features of CRNP are its

small population of Cross River gorillas (below the Obudu Plateau in the former Boshi Extension Forest reserve, which was created as a gorilla sanctuary in 1958), and its population of Preuss's red colobus monkeys (in the Oban Division, northeast of Ekonganaku towards Korup) (Oates et al., 2004).

Studies have revealed that vegetation here has evolved over 60 million years ago, there are 119 species of mammals in the Park which include 18 out of the 23 species of monkeys found in Nigeria (representing 78% of Nigeria's total), 48 species of Fish and 950 species of butterflies (90% of Nigeria's totals) among a wide array of insects (NNPS, 2020).

Figure 5

The Cross River National Park Showing the Study Area in Green



Oban Division: This is the larger of the two divisions and is the focus for this study with an area about 3000 km^2 in size and very high in biodiversity concentration. It is ecologically contiguous with Korup National Park in the Republic of Cameroon separated only by the international boundary between Nigeria and Cameroon. About 1,568 plant species (77 of which are endemic to Nigeria); 75 mammals including forest elephant, Chimpanzee, Drill and Buffalo; 382 birds including the olive-green ibis and 42 snake species have all been documented in the Park. It is also rich in epiphytic ferns and orchids (NNPS, 2020).

Figure 6

Wildlife at the Cross River National Park, Photo Credit World Wildlife and Rbrausse/Creative Commons

The Oban division operates in two Ranges, Oban East, and Oban West, for management and administrative convenience.

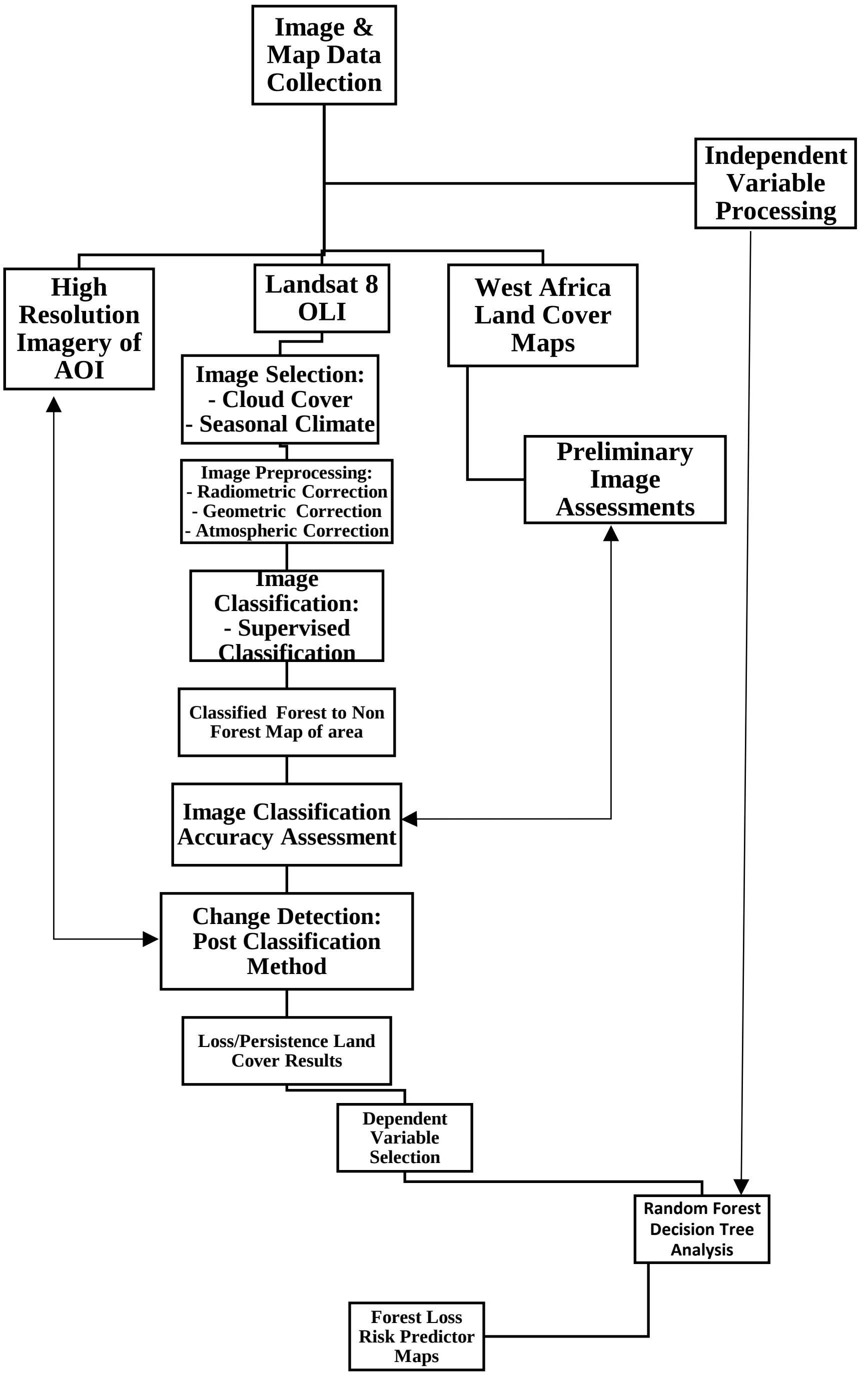
The Cross-River State Superhighway

The Cross River State Government in Nigeria is proposing to construct a “Cross River Superhighway” that would bisect critical remaining areas of tropical rainforest in southeastern Nigeria (Mahmoud et al., 2017). The route, proposed by Cross River State Government, would slash more than 100 kilometers (60 miles) through intact forest. Half of this length is to run through a national park sheltering the critically endangered Cross River gorilla (*Gorilla gorilla diehli*); the proposal claimed an astonishing 20–kilometer (12-mile) buffer on either side of the route.

Environmentalists, NGOs, and local communities have fiercely resisted the project, launching a series of legal challenges and numerous petitions to state and national authorities. Environmental impact assessments for the project have been rejected three times by Nigeria’s Federal Ministry of Environment for failing to meet required standards; the assessments have been found to be comprehensively inadequate — missing baseline data, engineering specifications, and required environmental safeguards — as well as failing to properly consult affected communities (Unah, 2019). Yet, the project is still ongoing even when alternative proposals have been made to the government on the route of the highway.

Figure 7

Flowchart of the Methodology



Selection of a methodology for this research is a very subjective process since there are several change detection methods within the field. However, by falling back to the main overarching research question: what are the *spatial* and *temporal* patterns of deforestation at

the Cross-River National Park, it can be deduced that a "from-to" change detection technique would be most useful in answering the research question as information on not only the extent of the land cover change is required, but also what the land-use or cover was, and what it turned into (Kamusoko & Aniya, 2008).

Previous research methods undertaken in land cover change detection and deforestation using remote sensing, GIS and machine learning formed the basis of the methodology I employed on this study.

Such studies include (Boori et al., 2016), who applied supervised classification – maximum likelihood algorithm in ArcGIS 10.2 to detect land use/cover changes in Eastern Siberia using multispectral satellite data obtained from Landsat 7 and 8(OLI). The study classified the various land use in 5 classes: settlement, vegetation, water/ice, wasteland, and wetland. These classes were used to delineate land cover changes over a 15-year period. (Kamusoko & Aniya, 2008), is another study where a post-classification change detection technique is used to reveal different trends in land use/cover changes over two periods (1973-1989 and 1989-2000). A hybrid supervised/unsupervised classification was carried out, with an ISODATA algorithm used for all other classes except bare land and settlement. Initially using training areas in only a supervised classification the study revealed problems with the classifier whereby agricultural areas having close spectral signatures with settlement areas/roads filled with by gravel were classified together, hence, a hybrid classification method was utilized to solve the problem.

In a report for the Wildlife Conservation Society (Okeke & Imong, 2018) assessed patterns of deforestation at the Oban division of the Cross River National Park and adjoining community forests using satellite land cover data obtained from 2000 Landsat 7 ETM images and 2018 Landsat 8 OLI images. Using the supervised maximum likelihood classifier in

Erdas imagine they were able to accurately determine the rate of deforestation between that period.

Using machine learning, remote sensing, and GIS (Cushman et al., 2017), predicted the future extent of deforestation across Borneo. Multi-date land cover maps were processed, and in a GIS processed as layers and fed into 2 machine learning applications: random forest and logistic regression as response variables. Several predictor variables were collected and used to create a probability map of Borneo showing extent of predicted deforestation.

PART I

Land cover change detection

For this study, 4 scenes covering the area of interest will be downloaded with path 187 and row 56 scene parameters. The sensor from which these images were obtained is the Landsat 8 (Operational Land Imager) with a spatial resolution of 30 meters from the USGS Earth Explorer website (https://earthexplorer.usgs.gov/).

The datasets are Landsat collection-1 level-2 products or surface reflectance products. These improve comparisons between multiple images over the same region by accounting for atmospheric effects such as aerosol scattering and thin clouds, which can help in the detection and characterization of Earth surface change. (USGS, 2020). The images being level 2 data have been ortho-rectified with geometric and radiometric corrections already applied. This makes it easier for image-to-image comparison in change detection analysis.

Further raster data sets to be applied to this study are GeoTiff landcover maps of the sub-Saharan region of West Africa. This dataset is produced by the Earth Research Observation and Science (EROS) Center of the United States Geological Survey under the West Africa land use dynamic project. It is an effort by the USGS to map land use and land

cover, characterize trends in time and space and understand their effects on the environment in West Africa (USGS, 2020).

These land cover maps represent fully classified covers of sub–Saharan West Africa, and the A.O.I is fully covered by the land cover map. The regional dataset is created at a 2-Kilometer resolution with 24 LULC classes using the Mangambi classification which applies mainly to the discerptions and understanding of vegetated land cover types, though the maps integrate vegetated and no vegetated surfaces. The LULC mapping covers 3 time periods 1975, 2000 and 2013 and are developed using the Rapid Land Cover Mapper (RLCM).

For this research LULC dataset will be combined with the downloaded Landsat GeoTiff images for the year 2020. As such the make-up of the datasets would be:

- 1975 - already classified land cover map
- 2000 - already classified land cover map
- 2013 - already classified land cover map
- 2020 – geotiff image to be classified using a hybrid classifier

The choice on which anniversary date the images are acquired is totally dependent on the seasonal phenological cycle of the Cross-River rain forest in Nigeria. Using approximately near-anniversary images greatly minimizes the effects of seasonal phenological differences that may cause spurious change to be detected in the imagery (Jensen, 2015), so the images are selected at periods of comparable characteristics like atmospheric conditions, sun elevation, and time of the year. Nigeria like all other tropical lands has only 2 climatic seasons, these are the dry and wet season. The rainy season runs from late March and ends in early October while the dry season begins end of October and ends early March. In a rainy season, heavy cloud cover greatly decreases the image quality this tends to make images taken in the dry season more reliable (Yichun, et al., 2008). Also, during the hot and dry

season such rain forests attain their maximum greenness thereby making it ideal for remote sensing capture (University of Illinois, 2016).

Image Processing

These are the computer processes employed to extract information from remotely sensed imagery by analyzing the variations in the individual picture elements (pixel) of a scene. For this study we shall make use of Erdas Imagine version 16.5 (2018) and ArcGIS 10.7.1 for most image processing procedures.

Pre-Processing

Image pre-processing would mostly be concerned with geometric correction of all raster datasets available for this study. Since the images are all Landsat collection 1 – level 2 data, ortho-rectification, geometric and radiometric correction have already been applied to the images. However, to maintain uniformity and follow best practices all raster datasets will be geometrically corrected to the 2020 Landsat 8 OLI image obtained from the USGS. This image is ortho-rectified and geometrically corrected to the UTM projection and will be used to reference the other images. Using Erdas Imagine, a minimum of 45 ground control points (GCP) is selected for this correction, this is selected at the primary image (the Landsat 8 imagery) and used to correct the secondary images to the primary.

A first order polynomial transformation or linear transformation is used to attain a root mean square error less than +/- 0.5. This is most important for any change detection method especially post-classification as inaccurate geometric rectification in both date images will be present in the final change detection thematic map (Jensen, 2015).

Image Classification Schemes

Image classification was focused on the 2020 Landsat images, it is broadly grouped to conform to the 1975, 2000 and 2013 classified images obtained from the West African land

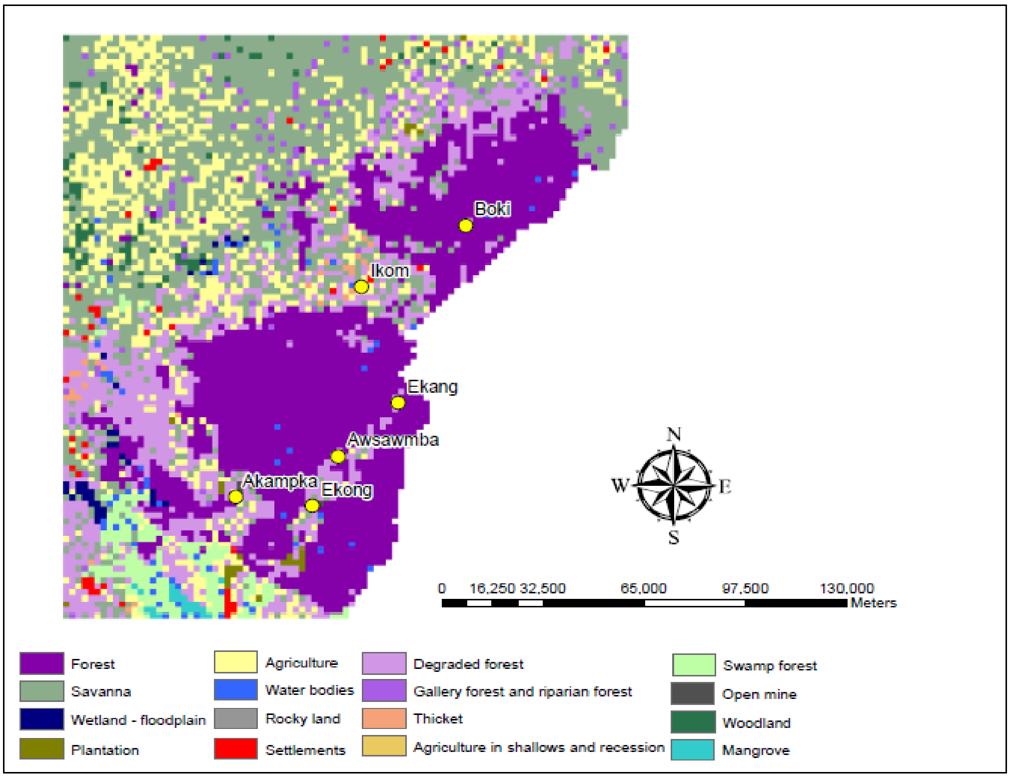
use dynamic project. For this study, the International Geosphere-Biosphere Program (IGBP) Land-Cover Classification System modified for the creation of MODIS land cover type products (Friedl et al., 2010) is be utilized.

The classifier is reduced to the following 4 classes:

1. Forest: Dominated by woody vegetation with a cover > 60% and height > 2m. Almost all trees remain green all year. Canopy is never without green foliage.
2. Agriculture/Sparse Land: Lands covered with temporary crops followed by harvest and a bare-soil period or Sparse woodland or scattered trees about 100 m apart and Exposed soil, sand, rocks, or snow and < 10% vegetated cover during the year .
3. Urban/Settlement: Land covered by buildings and other human-made structures.
4. Water Bodies: Oceans, seas, lakes, reservoirs, and rivers. Can be fresh or salt water.

Figure 8

1975 West Africa Land Use Land Cover Time Series Dataset of the Area of Interest.



Post-Classification Comparison Change Detection

A supervised classification of the Landsat 8 2020 images was carried out using the supervised maximum-likelihood algorithm. However, as there were noticeable signature extension problems due to the use of the supervised classification an ISODATA unsupervised classifier was employed to resolve some spectral signature confusions of the classifier, as such a hybrid of both classifiers is utilized to accurately classify the spectral surface characteristics. This is the foundation for making a “from-to” change detection analysis.

To select training sites for the maximum-likelihood classifier, a combination of high resolution near date imagery and the most recent topographic maps of the area of interest is utilized, this is referred to as an on-screen training site selection method.

After classification, a post classification comparison change detection was utilized in detecting temporal change in surface spectral characteristics. There are 2 methods for post-classification comparison, which include:

a) Per-pixel analysis

b) Object-based image analysis (OBIA)

The pixel-to-pixel analysis is used for this study, for this a simple intersect of the Date 1 image over the Date 2 image is performed in ArcGIS and a change detection matrix is developed from the overlay. The intersect will produce a change image map of different spectral/brightness values of date 1 and date 2 images and subsequently produce several from – to change matrices for each set of temporal intersects. For instance, change matrices would be produced for the 1975 – 2000, 2000 – 2013 and 2013 – 2020 change maps. Hence, 27 possible change sequences are produced as well as 4 no change elements for each change period.

A conversion of the land cover maps from raster to polygons enabled the calculation of the geometry for each class, to display and quantify the extent and spatial distribution of land use/cover changes.

Accuracy Assessment.

The accuracy assessment is performed on the individual classified maps used in the change detection study. Error matrices were developed for the 5 classified map dates whereby each classified image is compared with control data which in this case would be a high-resolution imagery of the area of interest.

The elements of the error matrix accuracy assessment include overall accuracy, producer's accuracy, user's accuracy, and kappa coefficient.

Table 4

Sample Confusion Matrix for Map Classification Assessment

		Ground Reference Test Information Class 1 to k (j columns)					
		1	2	3		k	Row total X_{i+}
Map class 1 to k (i rows)	1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	….	$X_{1,k}$	X_{1+}
	2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	….	$X_{2,k}$	X_{2+}
	3	$X_{3,1}$	$X_{2,3}$	$X_{3,3}$	….	$X_{3,k}$	X_{3+}
	.	.	.	.	….	.	.
	k	$X_{k,1}$	$X_{k,2}$	$X_{k,3}$	….	$X_{k,k}$	X_{k+}
	Column total X_{+j}	X_{+1}	X_{+2}	X_{+3}	….	X_{+k}	N

where:
. Cell entry, X_{ij}, is the proportion of area mapped as class i and labeled class j in the reference data.
. The row marginal, X_{i+}, is the sum of all xij values in row i and represents the proportion of area classified as class i.
. The column marginal, X_{+j}, is the sum of all xij values in column j and represents the proportion of area that is truly class j. The diagonal, X_{ii}, summarizes correctly classified pixels.
. All off diagonal cells represent misclassified pixels.

Formulae for the calculation of the overall, producers, and user's accuracy is listed below:

Overall accuracy:

$$\frac{\sum_{i=1}^{k} x_{ii}}{N}$$

Producer's accuracy: $\frac{x_{jj}}{x_{+j}}$

User's accuracy: $\frac{x_{ii}}{x_{i+}}$

Kappa Analysis and Coefficient of Agreement:

The Kappa analysis is a discrete multivariate technique of use in accuracy assessment (Jensen, 2015). The coefficient of agreement $\hat{K}$ is a measure of agreement or accuracy between remote sensing derived classification map and the reference data as indicated by a) the major diagonal, and b) the chance agreement, which is indicated by the row and column totals or the "marginal" (Congalton & Green, 2009; Paine & Kiser, 2003).

$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+j})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+j})}$$

PART II

Forest Loss Risk Prediction

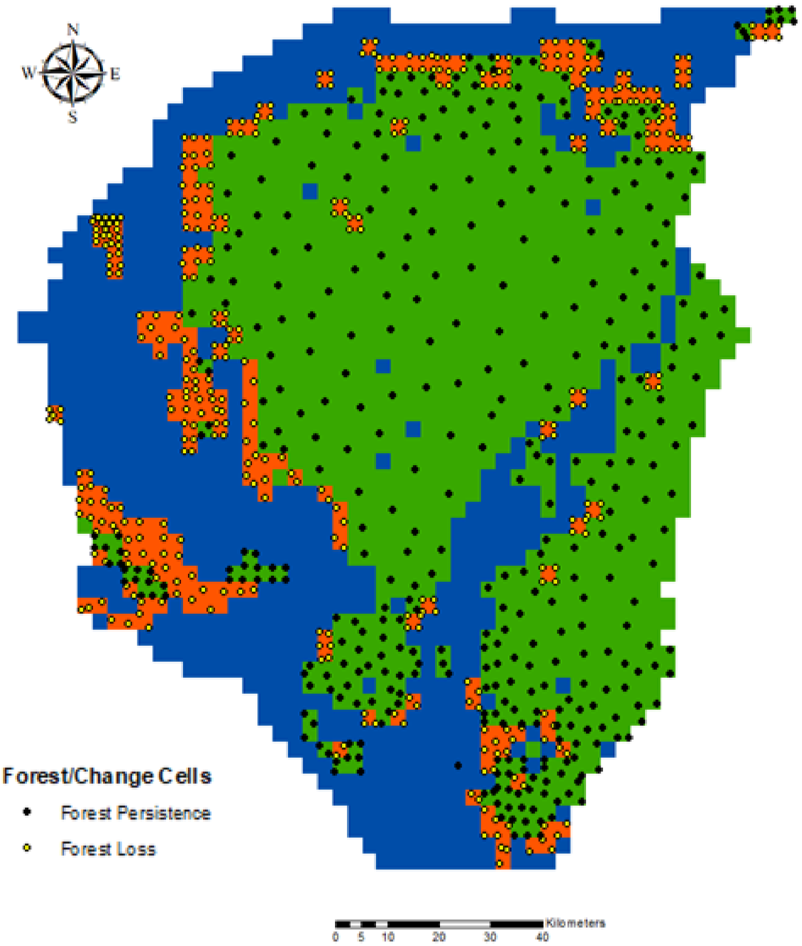
For this part of the study the response variables are be obtained from forest and non-forest land cover map classes created in part I of the study and used for the deforestation analysis, the analysis is based on combining all forest & non-forest classes into a binary forest/non-forest map for each time series. The binary maps created are for the following time series: 1975-2000, 2000 – 2013 and 2013 – 2020, with each series having a binary GIS layer of points depicting 0 for "persistent forest" and 1 for "forest loss". Maps are produced for all pixels that were loss or persistence for each time series; 1975 – 2000, 2000 – 2013, and 2013 – 2020 and used as the source of the loss and persistence cells for model training and model

assessment. Non-Forest pixels for each time series between 1975 – 2020 are not utilized as part of the model training or assessment.

From the work of (Chen et al., 2018) in the selection of the number of occurrence or non-occurrence response variables, it was discovered that an imbalance in the selection would create a bias in the prediction model and fit. In an imbalanced sample the resulting model could be deceptive, over-predicting in the majority class and under-predicting in the minority. Hence, it was important to find a ratio that produces unbiased estimates of forest loss risk. It is on this premise that a random selection of equal loss to persistence cells are selected for this study to be used to train the algorithm.

Figure 9

Response Variable Selection



Predictor Variable Selection

To create the deforestation probability models, selection of the most important independent variables is imperative. Independent variable selection is predicated on the need to adequately reflect socio-economic/human and environmental landscape factors that are thought to drive deforestation rates. Some assumptions that reflect the drivers include:

1. Topography: high risk areas are at low elevation and low risk areas at higher elevation
2. Consideration of areas with past forest loss
3. Large human population parameters influence high resource consumption.
4. Distance to important features such as roads and rivers which drive settlement establishment and expansion.
5. Soil types which sustain plant growth.

The predictor variables selected are divided into 2 classes, environmental and anthropogenic factors which are most viable in predicting the model.

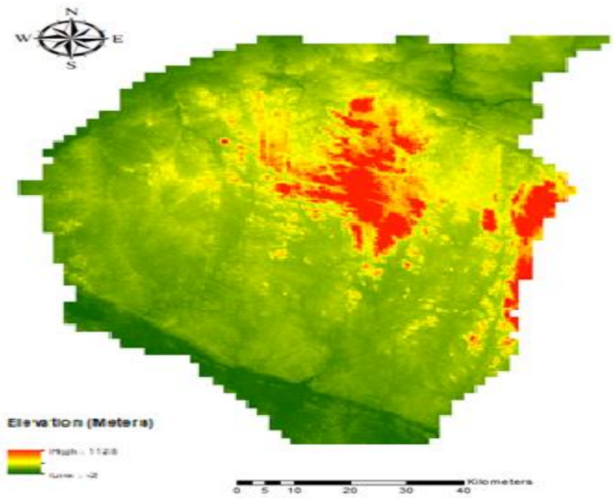
Physical Variables

Elevation:

Elevation is an important factor in the determination of the deforestation probability model, (Cushman et al., 2017) hypothesized in their study that the risk of deforestation increased as elevation decreases as an inverse relationship. In the higher altitude, deforestation has mainly occurred by the action of physical factors such as weathering, aeolian erosion, landslide, and other physical causes but in lower altitude it is caused by mainly anthropogenic interferences like grazing, wood smugglings, agricultural land expansion and human occupation related activities (Saha et al., 2020). The elevation map was created from the Shuttle Radar Topography Mission (SRTM) 30 meters digital elevation model created as a joint venture of the National Imagery and Mapping Agency (NIMA) and National Aeronautics and Space Administration (NASA)

Figure 10

Elevation of Study Area

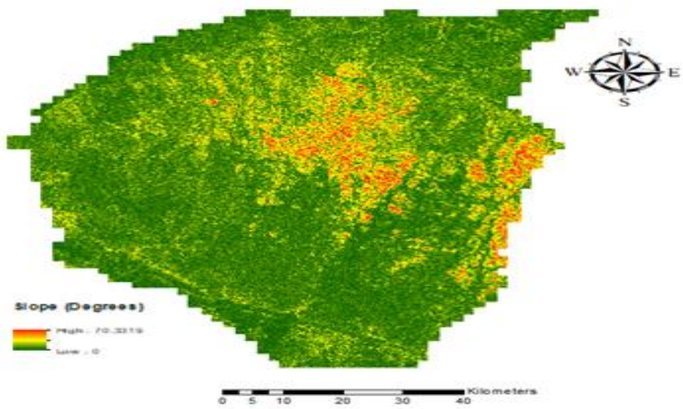


Slope Angle:

The slope angle is a measure of change in elevation, it has an indirect relationship with forest cover loss. A study by (Sandel & Svenning, 2013) showed that tree cover and slope show a strong, general, and global association, suggesting that slopes indeed consistently act as refuge for trees. In contrast, human impacts are most pronounced on relatively flat terrain. The association of trees and slopes was particularly strong where human impacts are high, supporting the hypothesis that the effect is anthropogenic in origin.

Figure 11

Slope Angle Raster of the Area of Interest

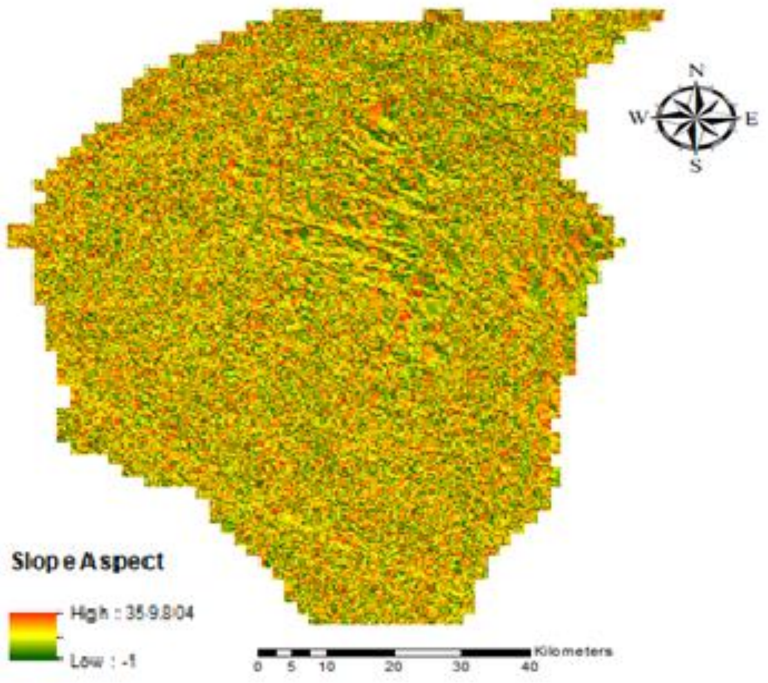


Aspect:

The slope aspect is the compass direction that a terrain surface faces, it determines the amount of sunshine and precipitation that impacts the terrain. (Sandel & Svenning, 2013), also found that both standing tree cover and tree cover are most pronounced on sloped terrain, while in areas of flat terrain or little or no slope the tree cover is highly impacted by human activities.

Figure 12

Slope Aspect Raster



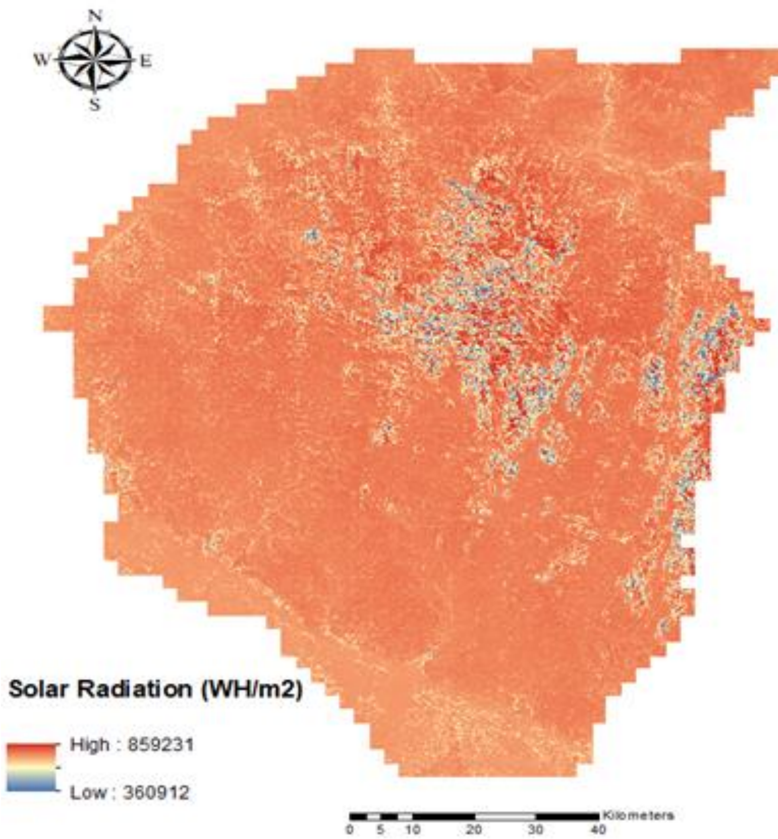
Solar Radiation:

This is the amount of sunlight or solar radiation received in the Earth's atmosphere or surface, it is a vital part of the plant/forest photosynthesis cycle and is measured in watt hours per square meter (WH/m^2). The structure of tropical rainforests is very dependent on competition for sunlight as such the denseness of tree cover is directly proportional to amount

of solar radiation received in the rainforest. An output raster of the total amount of solar insolation for each pixel of the area of interest is created.

Figure 13

Solar Radiation Raster



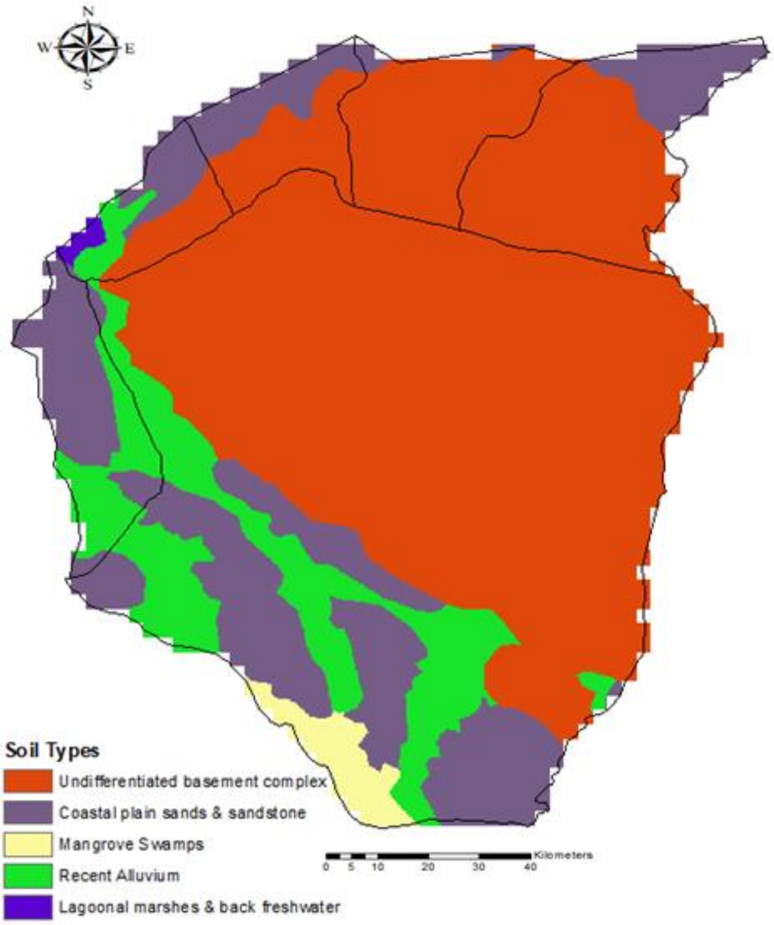
Soil Type:

Soils are formed by the gradual process of weatherization of underlying rock material and decomposition of organic matter, in tropical climes vegetation itself has a great influence in the formation and nature of soil (Peace Corp. Information Collection & Exchange, 1990). However, deforestation is most likely to occur in and around tropical forests with soil suitable for agriculture in terms of soil fertility and hydrological conditions (Zanelle et al., 2017), a

good example is the conversion of large parts of tropical forests in Borneo for oil palm cultivation, a phenomenon which is being replicated at the Cross River National Park.

Figure 14

Soil Type Map for Area of Interest



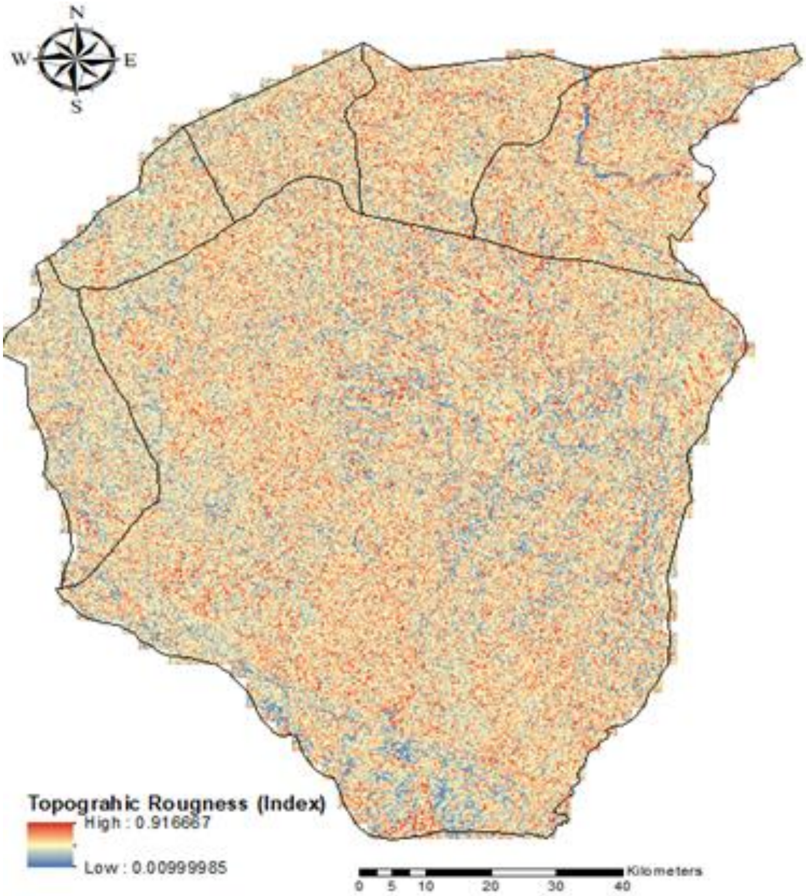
Topographic Roughness Index:

This is the spatial variability in elevation (Amatulli et al., 2018), to express the amount of elevation difference between adjacent cells of a DEM. It calculates the difference in elevation values from a center cell and the eight cells immediately surrounding it, then it squares each of the eight elevation difference values to make them all positive and averages the squares. The topographic roughness index is then derived by taking the square root of this

average (Riley et al., 1999). It is hypothesized that the risk of forest cover loss decreases with increased topographic roughness.

Figure 15

Topographic Roughness Index



Human/Anthropogenic Variables

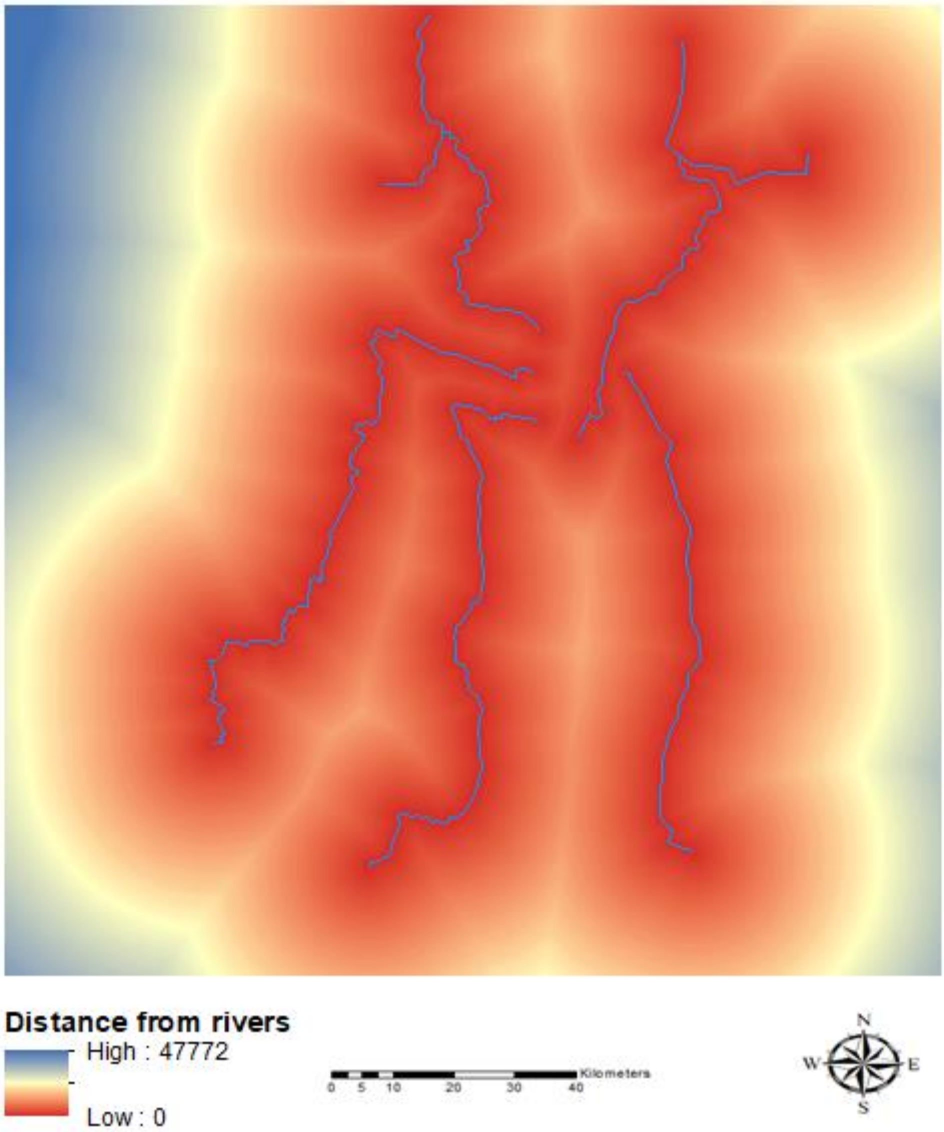
The following human variables are utilized for this study:

Distance from rivers:

Water is an important factor for the support of vegetation and for transportation of timber. The rivers that run through the area of interest such as Rivers Cross, Afi and other smaller rivers are used to create a raster of the Euclidean distance of each pixel to the rivers.

Figure 16

Raster of Distance from Rivers

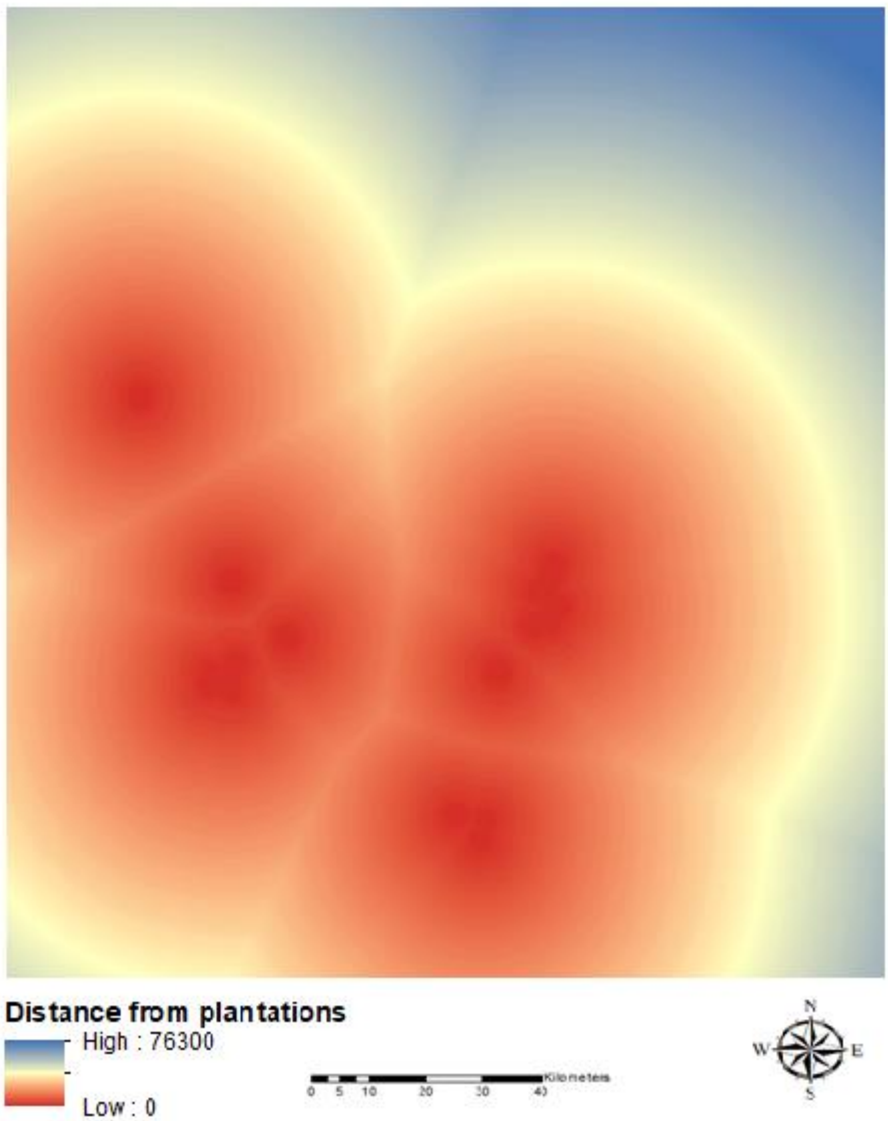


Distance from large plantations:

The Cross River National Park shares a boundary with some large plantations producing oil palm, pineapple, rubber, and other cash crops, some of which have extended into the forest park already. The risk of forest loss is hypothesized to be greater with proximity to these large plantations.

Figure 17

Raster of Distance from Plantations

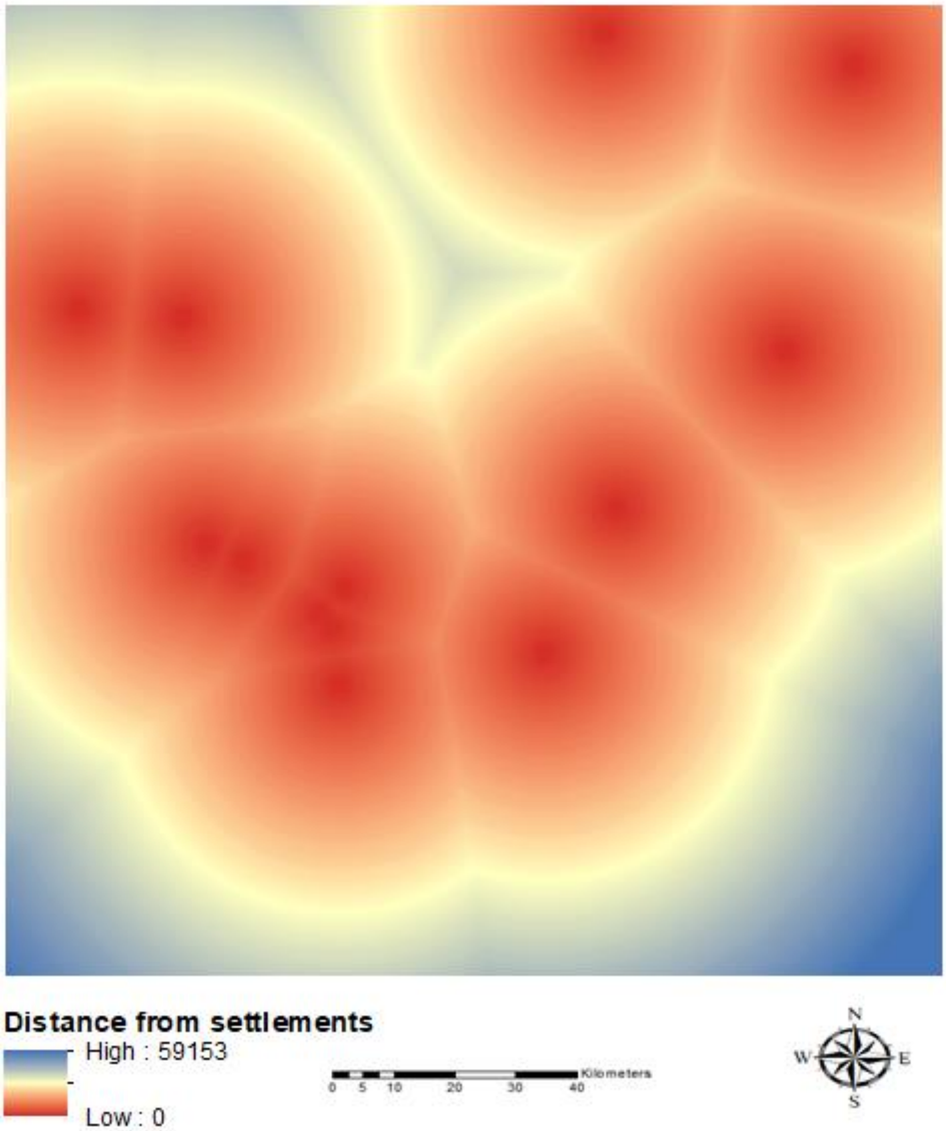


Distance to major towns/settlements:

Due to expansion of settlements around the park and the nature or occupation of the communities around the parks being agrarian as well as chain saw loggers both illegal and legal, it is deduced that the risk of forest loss will increase with a lower distance to the settlements.

Figure 18

Raster of Distance from Settlements

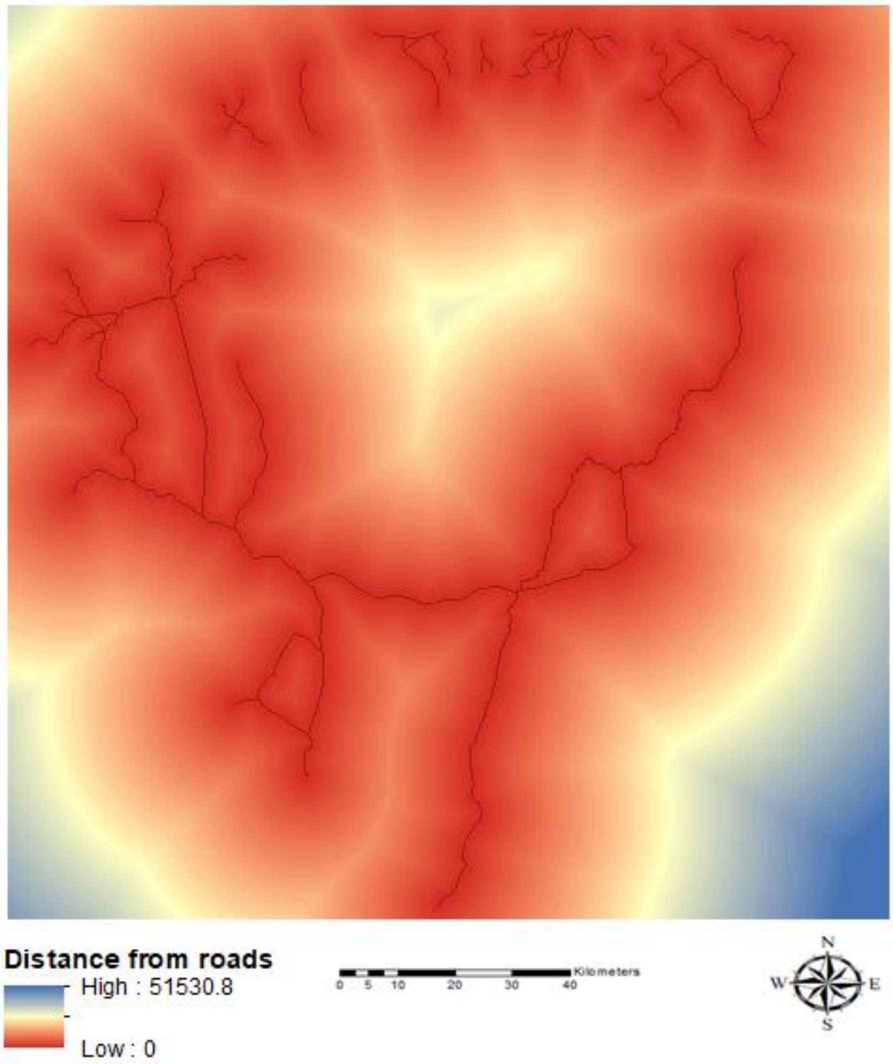


Distance from roads:

Deforestation is much higher in areas closer to roads and rivers, the proximity to transportation networks particularly rapidly growing unofficial road networks is a major proximate driver of deforestation unless mitigated by strict protection laws. (Barber et al., 2014). It is hypothesized that the closer the forest cover to a road network the higher the risk of deforestation.

Figure 19

Raster of Distance from Roads



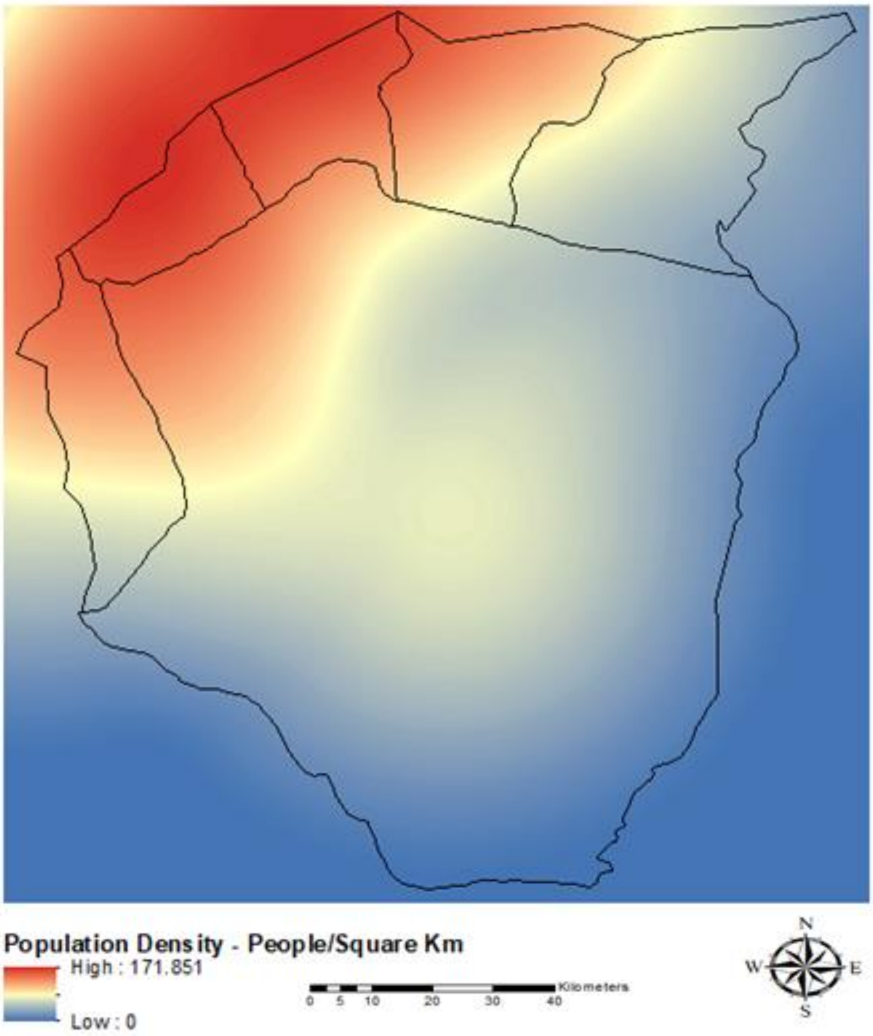
Population Density:

Human population growth and development increases the rate of deforestation in biodiverse hotspots. In areas which relatively low human development index the rates of deforestation are remarkably higher as there is increased pressure by humans on forest resources (Jha & Bawa, 2006). There is a direct relationship between population and deforestation as the higher the population density around the biodiverse landscape the higher the rates of deforestation. For this study the population density for each Local Government

Area (LGA) contained in interest was obtained from the National Populations Commission (NPC) and used to create a density raster of the 6 LGA's in the area. Due to reliable data constraints, the population data utilized is the National Population Commission and National Bureau of Statistics 2016 projections for each LGA for the 2020 model.

Figure 20

Population Density Raster of the Area of Interest



Model Creation using Random Forest Regression Analysis

The random forest regression tree analysis is used in this paper to predict the risk of forest loss for time periods 1975 – 2000, 2000 – 2013 and 2013 – 2020. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).

The response variable and predictive variables split nodes (m) are required to build a random forest, numerical and categorical variables can be taken as predictive variables, Breiman (2001), posited that random forest largely eliminates generalization error and by increasing the number of trees, generalization error is further reduced also RF estimates the importance of predictive variables through the OOB error measurement (Breiman & Cutler, 2004). The RF model does not create the risk of overfitting as all the individual trees are completely independent and hence it can resist the overtraining, outliers in predictors and handle missing values (Breiman & Cutler, 2004).

For this study equal numbers of response variables (change/persistence) were selected for each period, and the predictor variables for each pixel point was extracted using ArcGIS Desktop 10.7.1. The points, its independent variable extracted data and each raster produced for the predictor variables were imported into random forest using R to train and run the model. The number of trees was set to 4000, as this would produce more stable models and covariance importance estimates, though it requires more memory and a longer running time.

Model Evaluation

To evaluate model performance, a 30% validation dataset is applied to determine the Area Under Curve (AUC) of the Receiver Operating Characteristics (ROC) of each model. The ROC curve was originally developed for operators of military radar receivers during

World War II, hence its name. It is a useful tool for evaluating the performance of diagnostic tests and more generally for evaluating the accuracy of a statistical model (e.g., logistic regression, linear discriminant analysis), (Zou et al., 2007)

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate at various threshold settings. The true positive rate is also known as sensitivity, recall or probability of detection while the false positive rate is also known as the probability of false alarm and is calculated as 1-specificity. Specificity and sensitivity are statistical measures of the performance of a binary classification test. (MathWorks, n.d.)

The AUC is an overall summary of diagnostic accuracy. AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy, an AUC of <0.5, indicates that the test does worse than chance (Hanley & McNeil, 1982).

$$\text{Sensitivity} = \text{TPR} = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - \text{FNR}$$

$$\text{Specificity} = \text{FPR} = \frac{FP}{N} = \frac{FP}{FP+TN} = 1 - \text{TNR}$$

P = condition positive (the number of real positive cases in the data)

N = condition negative (the number of real negative cases in the data)

TP = true positive (equivalent with hits)

TN = true negative (equivalent with correct rejection)

FP = false positive (underestimation or false alarm)

FN = false negative (overestimation or miss)

Chapter 4: Results

Image Classification and Change Detection of the CRNP 1975 - 2020

Using data of classified raster obtained from the West Africa Land Use Land Cover Time Series dataset, alongside image classification of Landsat 8 OLI imagery of the area of interest, the following change detection results are obtained.

1975 Land Use Land Cover Analysis

Analysis of the classified image of the area of interest in 1975 yielded the following results for the 4 land classes:

Table 5

Land Classes Area and Percentage of Total Area for Year 1975

Classes	**Area (km^2)**	**Land Percentage**
Forest	5170.693608	68.48
Agriculture/Sparse Land	2320.720592	30.74
Water	53.731967	0.71
Settlement	5.123108	0.07
Totals	**7550.269275**	**100.00**

Figure 21

Land Cover Classes - 1975

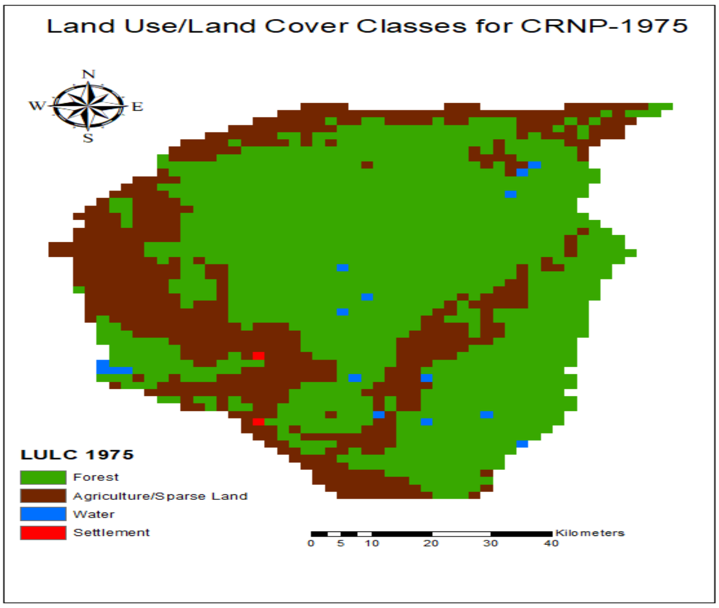


Figure 22

Land Classes of the Area of Interest in Km^2 - 1975

2000 Land Use Land Cover Analysis

Analysis of the four (4) land classes obtained from the USGS West Africa Land Cover Time Series Dataset is shown below:

Table 6

Land Classes Area and Percentage of Total Area for Year 2000

Classes	**Area (Km^2)**	**Land Percentage**
Forest	4812.64415	63.77
Agriculture/Sparse Land	2672.035762	35.41
Water	53.731967	0.71
Settlement	8.038816	0.11
Totals	**7546.450695**	**100.00**

Figure 23

Land Classes of the Area of Interest in Km^2 - 2000

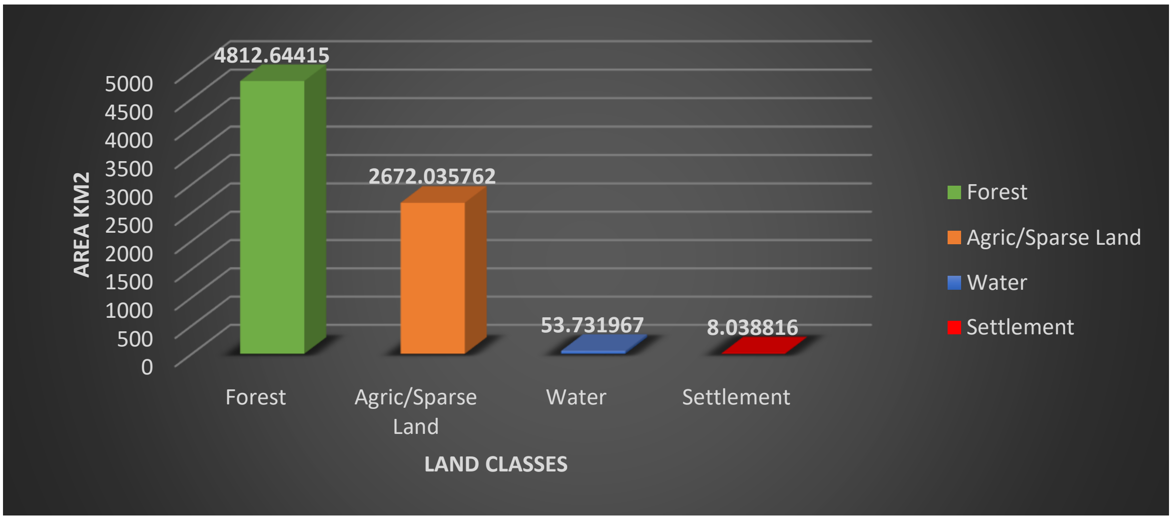
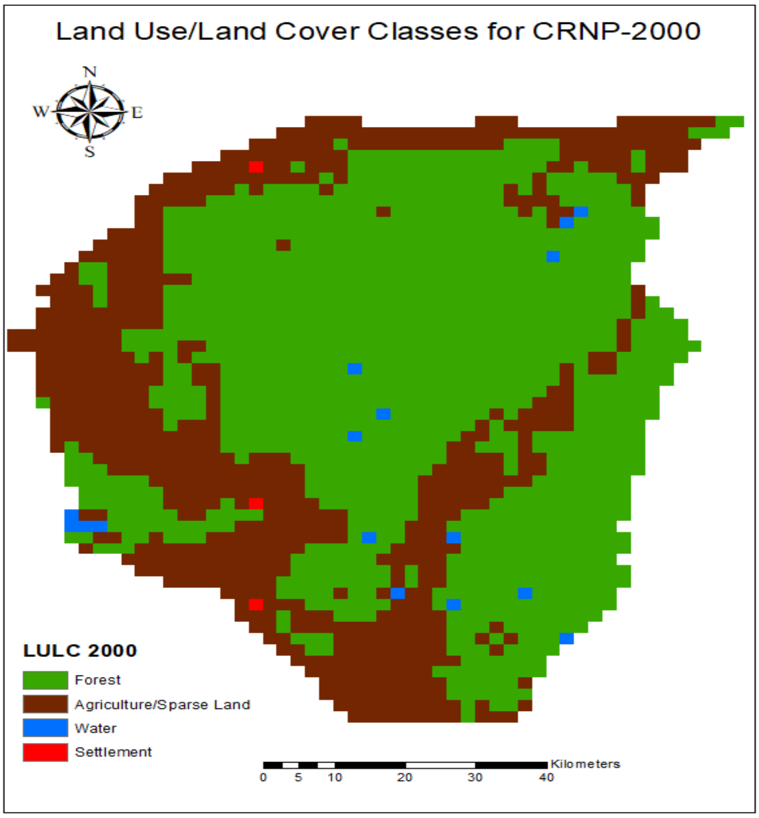


Figure 24

Land Cover Classes - 2000



2013 Land Use Land Cover Analysis

Analysis of the four (4) land classes obtained from the USGS West Africa Land Cover Time Series Dataset is shown below

Table 7

Land Classes Area and Percentage of Total Area for Year 2013

Classes	**Area (Km2)**	**Land Percentage**
Forest	4059.135329	53.78
Agriculture/Sparse Land	3400.327759	45.05
Water	57.642645	0.76
Settlement	30.677997	0.41
Totals	**7547.78373**	**100.00**

Figure 25

Land Classes of the Area of Interest in Km^2 - 2013

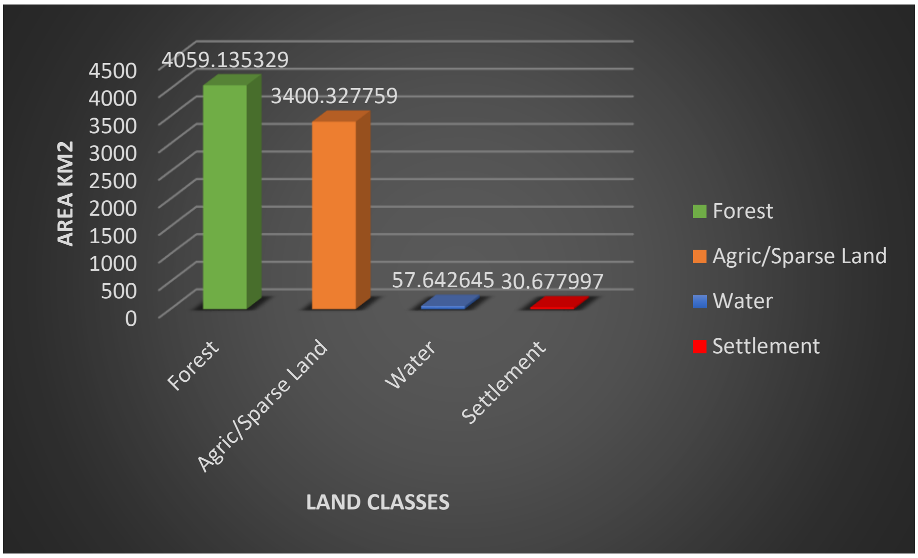
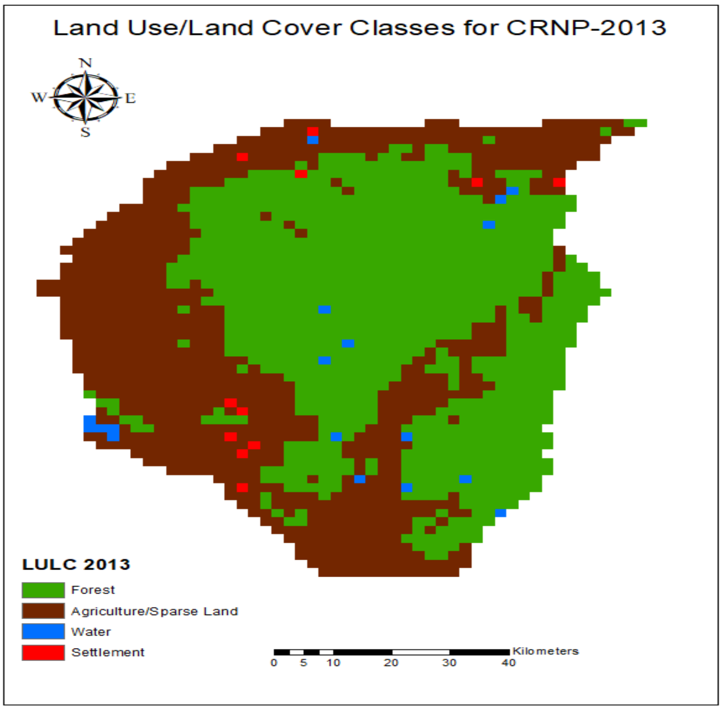


Figure 26

Land Cover Classes - 2013



2020 Land Use Land Cover Analysis

Table 8

Land Classes Area and Percentage of Total Area for Year 2020

Classes	Area (Km2)	Land Percentage
Forest	3595.399858	47.58
Agriculture/Sparse Land	3891.867386	51.50
Water	14.451544	0.19
Settlement	55.284225	0.73
Totals	**7557.003013**	**100.00**

Figure 27

Land Classes of the Area of Interest in Km^2 - 2020

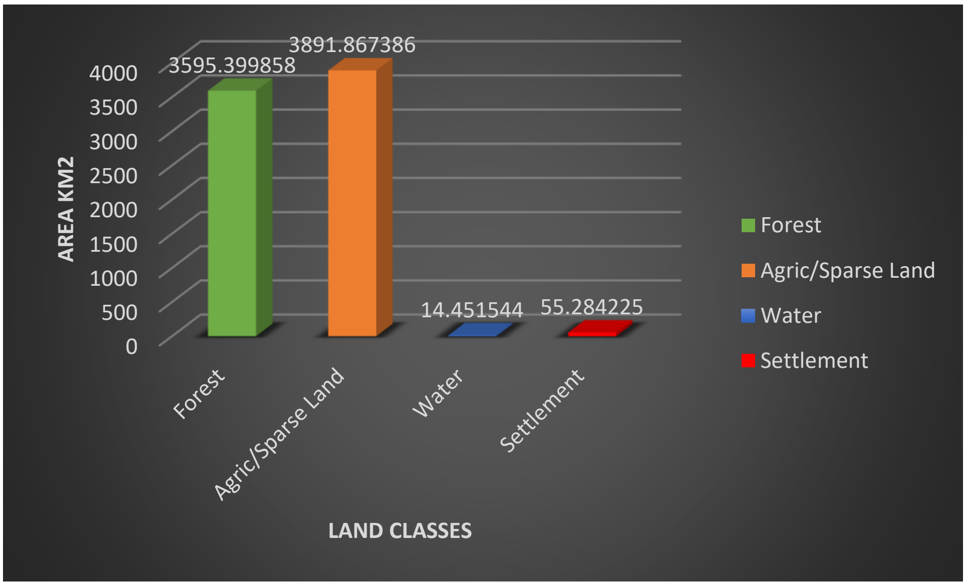


Figure 28

Land Cover Classes – 2020

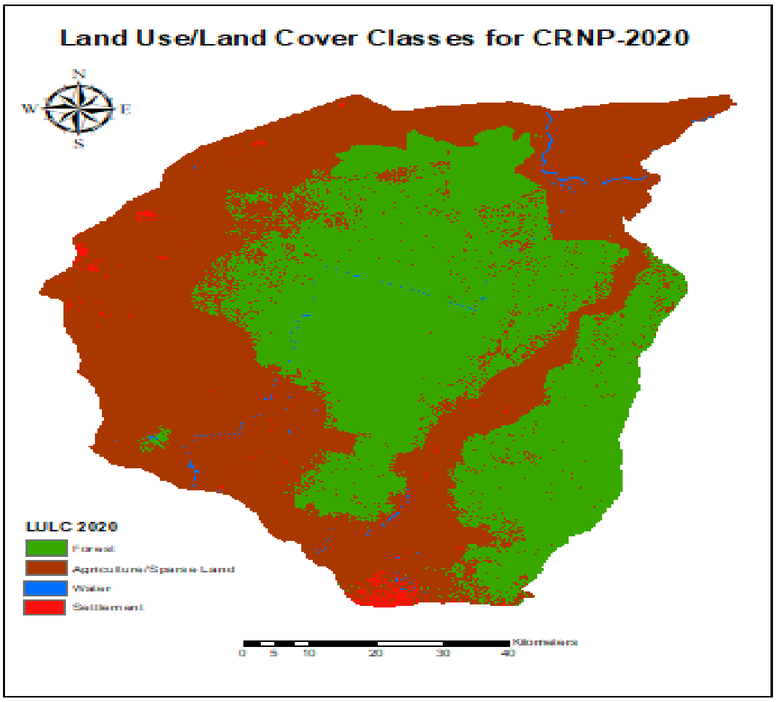
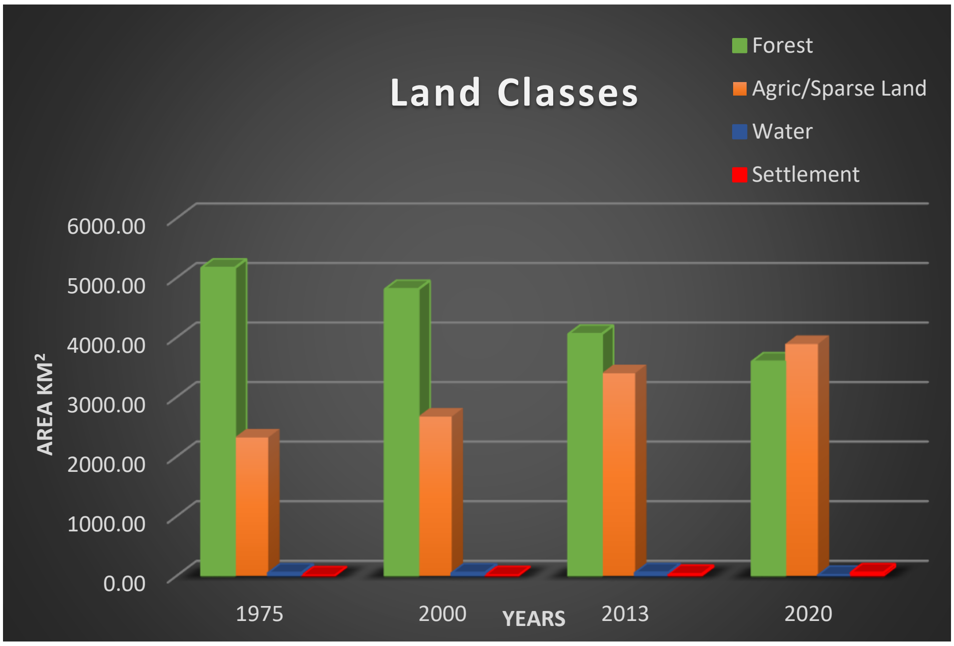


Figure 29

Land Classes over a 45 - Year Period



Land Use/Land Cover Accuracy Assessment – 2020

A classification accuracy assessment is carried out on the 2020 land classification map obtained from classifying a Landsat 8 OLI image if the area of interest. An error matrix is produced (Appendix A) and used to calculate the user accuracy, producer accuracy, overall accuracy, and kappa coefficient (Appendix 2-3) for the classification. All the four (4) classes have user accuracies above 50% with Forest at 94.12%, Agriculture/Sparse Land at 85.7%, Water at 68.75% and Settlement at 62.5%. For the producer's accuracy settlement and water have 100% accuracies while Forest and Agriculture/Sparse Land have accuracies of 72.72% and 70.58 respectively, the overall accuracy for the classification is calculated at 80% and the Kappa coefficient 72.5%.

Land Cover Change Detection within CRNP

Using the per-pixel post classification change detection analysis, the following results are derived from the change maps created:

LULC change detection 1975 – 2000

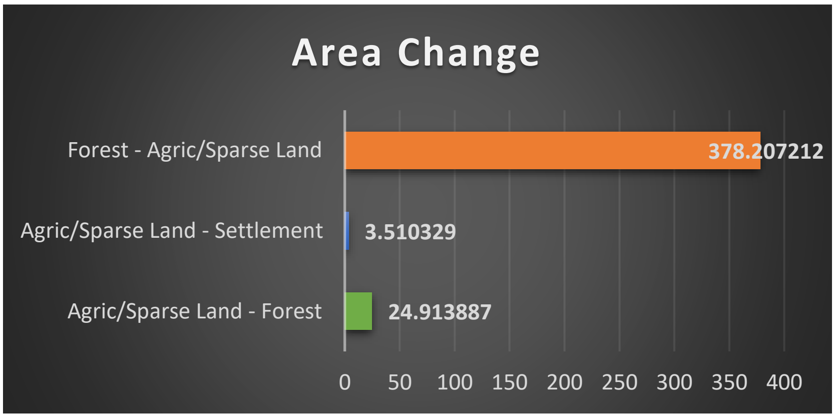
Table 9

Significant Class Changes 1975-2000

Change (1975-2000)	**Area Change**
Agriculture /Sparse Land - Forest	24.913887
Agriculture /Sparse Land - Settlement	3.510329
Forest - Agriculture/Sparse Land	378.207212

Figure 30

Change in Area Period 1975 – 2000 (Km^2)



LULC change detection 2000 – 2013

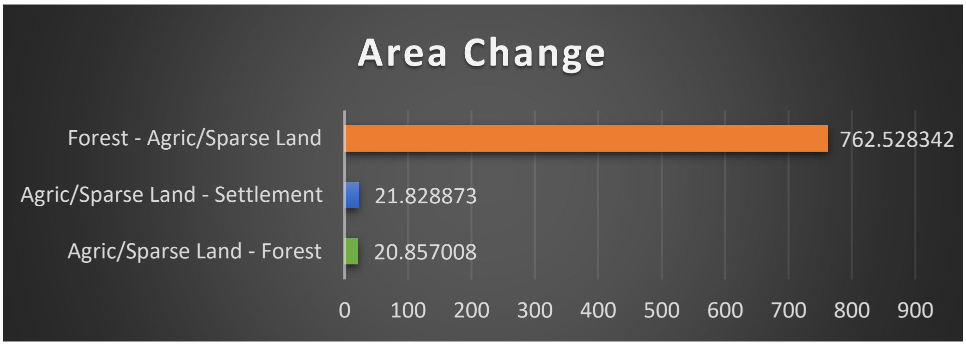
Table 10

Significant Class Changes 2000-2013

Change 2000 - 2013	**Area Change**
Agriculture /Sparse Land - Forest	20.857008
Agriculture /Sparse Land - Settlement	21.828873
Forest - Agriculture /Sparse Land	762.528342

Figure 31

Change in Area Period 2000 – 2013 (Km^2)



LULC change detection 2013 – 2020

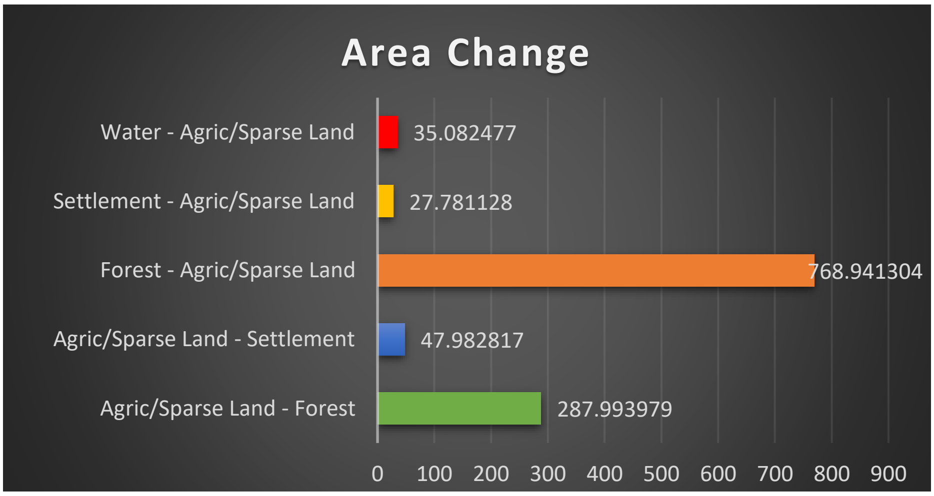
Table 11

Significant Class Changes 2013-2020

Change 2013 - 2020	**Area Change**
Agric/Sparse Land - Forest	287.993979
Agric/Sparse Land - Settlement	47.982817
Forest - Agric/Sparse Land	768.941304
Settlement - Agric/Sparse Land	27.781128
Water - Agric/Sparse Land	35.082477

Figure 32

Change in Area Period 2013 – 2020 (Km^2)



Deforestation probability assessment using random forest (RF)

The deforestation probability analysis of the Cross River National Park was carried out in 3 time series (1975-200, 2000-2013 and 2013-2020) to create 3 different deforestation probability assessments. The dependent factor is expressed as a binary variable of persistent forest 0 and forest cover change 1, while the independent variables are extracted from generated raster, both applied to model deforestation probability using random forest in R.

Multi-collinearity test of independent variables

To test for correlation between the multiple independent variables the Variance Inflation Factor (VIF) and Tolerance are derived for each individual variable using IBM SPSS, these indicators provide the correlation and strength of correlation between predictor variables in a regression model (Statology, 2019). From the results listed in the table below it is determined that there is no multi-collinearity problem in the variables as each of the have a VIF of <5 and tolerance > 0.1.

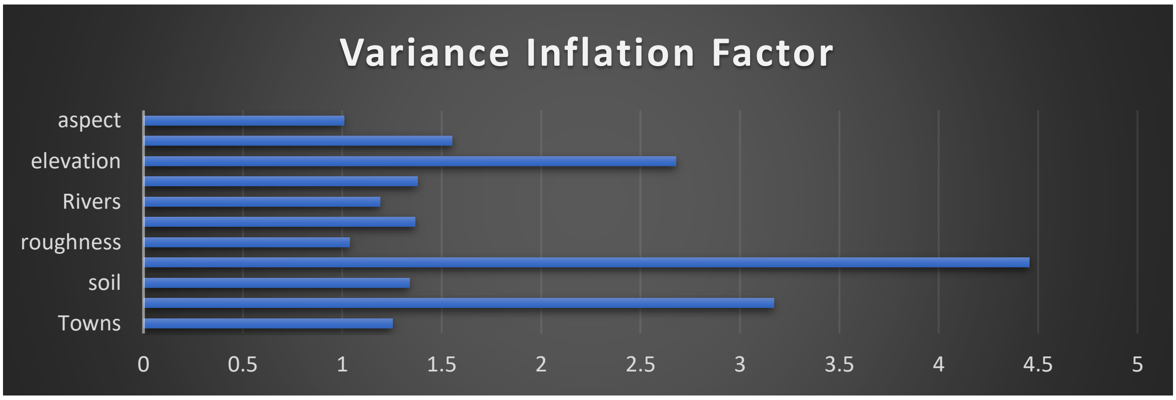
Table 12

Multi-Collinearity Test Data for Independent Variables

Variable	Tolerance	VIF
Towns	0.797644	1.253693
Solar	0.315377	3.170809
Soil	0.747073	1.338557
Slope	0.224418	4.455978
Roughness	0.965028	1.036240
Roads	0.73254	1.365113
Rivers	0.840917	1.189178
Plantation	0.725083	1.379152
Elevation	0.373207	2.679477
Population	0.644277	1.552127
Aspect	0.991389	1.008686

Figure 33

Variance Inflation Factor Plot



Deforestation Probability Analysis in the 1975-2000 period

For the period between 1975-2000, the random forest regression analysis using the binary response and predictor variables produced the following statistical results:

- Mean of squared residuals: **0.08230612**
- % Var explained: **67.08**

Table 13

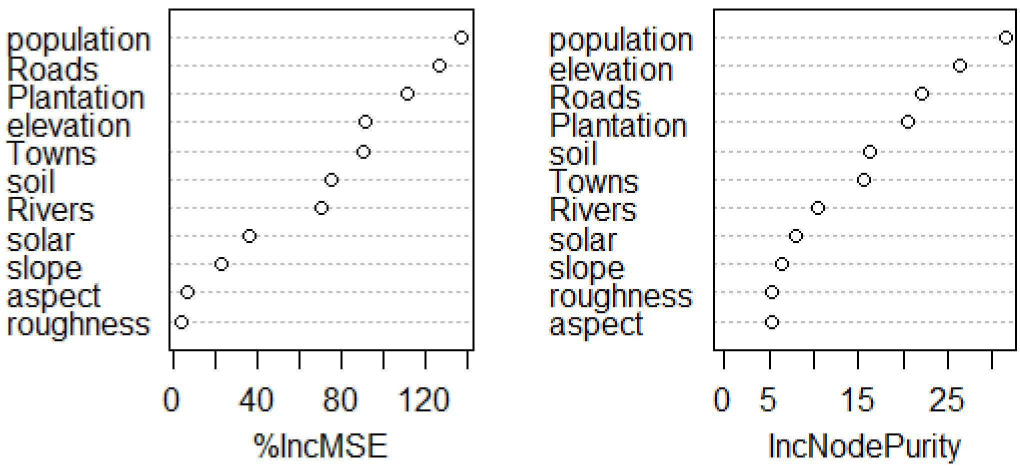
Importance of Independent Variables using the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 1975 to 2000

Variable	%IncMSE	IncNodePurity
Solar	30.0012402	7.536835
Slope	22.5022202	6.45786
Elevation	96.6704845	28.907097
Aspect	-0.2504866	5.110581
Rivers	68.4907557	10.432099
Roads	106.8424562	18.2959
Towns	82.4310427	13.831677
Plantation	115.1850086	20.727274
Topographic Roughness	7.9463509	5.497474
Soil Type	84.89268	19.463513
Population Density	135.6407216	31.55175

The %IncMSE parameter is known as the Mean Decrease Accuracy, it shows how much the model accuracy decreases if that variable is left out, while the IncNodePurity parameter is a measure of variable importance based on the Gini impurity index used for the calculating the splits in trees. The higher the value of mean decrease accuracy or mean decrease gini score, the higher the importance of the variable to our model. (Datacamp Inc, n.d.).

Figure 34

Plot of Importance of Independent Variables showing the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 1975 to 2000



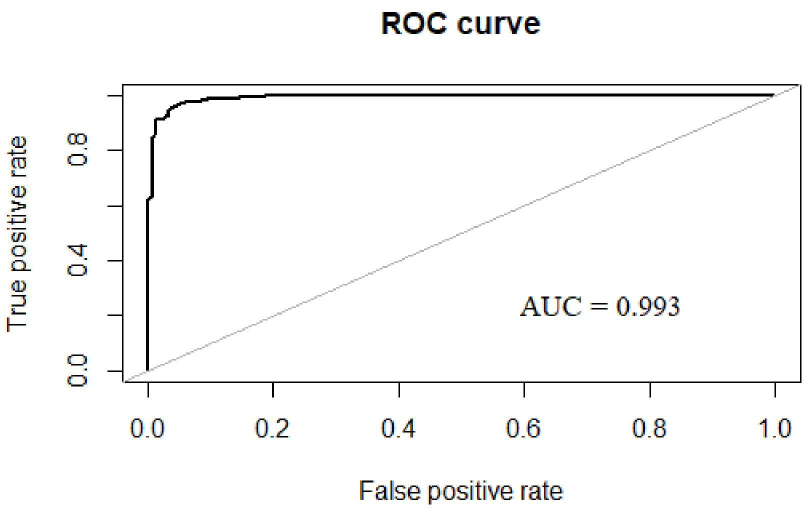
From the variable importance charts and table, focus will be placed on the %IncMSE as this is a more robust and informative measure. Population density alongside distance to roads, plantations are the most important variables in predicting this model, these are all anthropogenic factors, as such the model indicates that socio-economic factors are the greatest drivers of forest loss in the area. It is important to note that the most important physical factor within this period is elevation, in fourth place.

Model Assessment

Using the area under the curve (AUC) of the receiver operating characteristics (ROC) as earlier discussed, the AUC for the model is 0.993 which indicates a good fit of the model with the training data set, it does much better than the average having a higher positive rate.

Figure 35

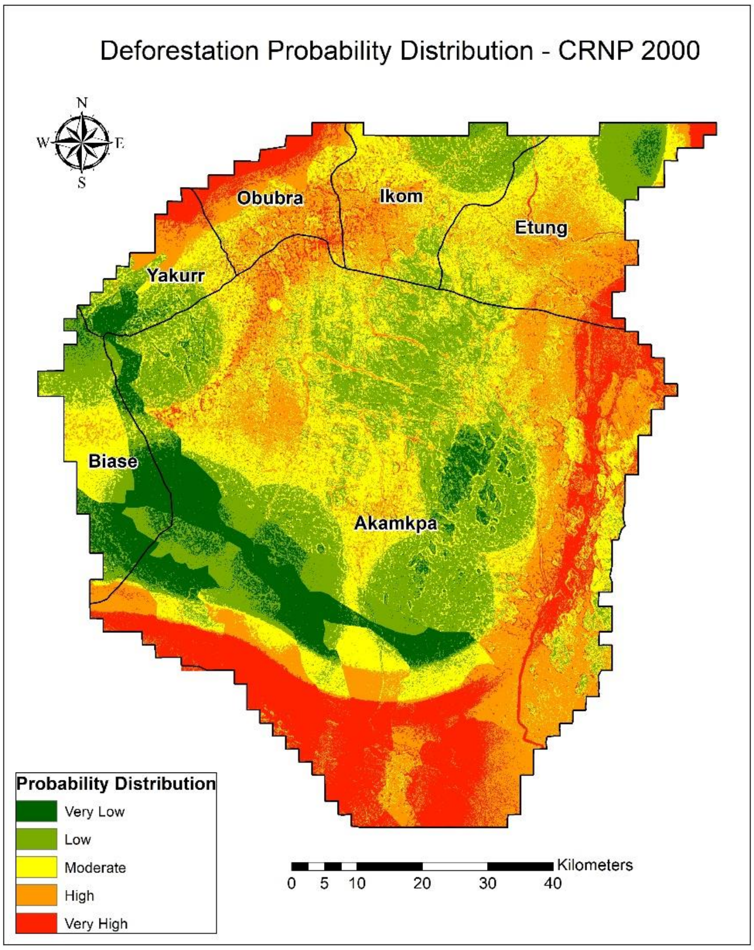
ROC Curve and AUC for Period 1975 - 2000



The deforestation probability map for the period 1975-2000 is displayed below, this was generated using the predict function in R to predict the response value of a new set of observations created by the model. The plot of the new response variable is displayed as the probability map below.

Figure 36

Deforestation Probability Model – 1975 To 2000



The categories, pixel depth and percentage are listed below for the probability map above:

Table 14

Risk Categories and Percentage of Total Area for Model 1975 – 2000

Category	Pixel	Percentage of total area
Very Low	707978	8%
Low	2051237	24%
Moderate	2242454	27%
High	2119304	25%
Very High	1268759	15%
Total	**8389732**	**100%**

The probability distribution of the map reveals areas of very high risk in the south and southeastern areas stretching into the forest cover classes of the park itself, another area which has a very high risk of forest cover loss stretches from the north western axis towards the central highlands of the forest. However, the center of the forest which is has a much higher elevation is predominantly dominated by low-risk cells.

Deforestation Probability Analysis in the 2000-2013 period

For the period 2000-2013, the random forest regression analysis using the binary response and predictor variables produced the following statistical results:

- Mean of squared residuals: **0.1422866**
- % Var explained: **43.09**

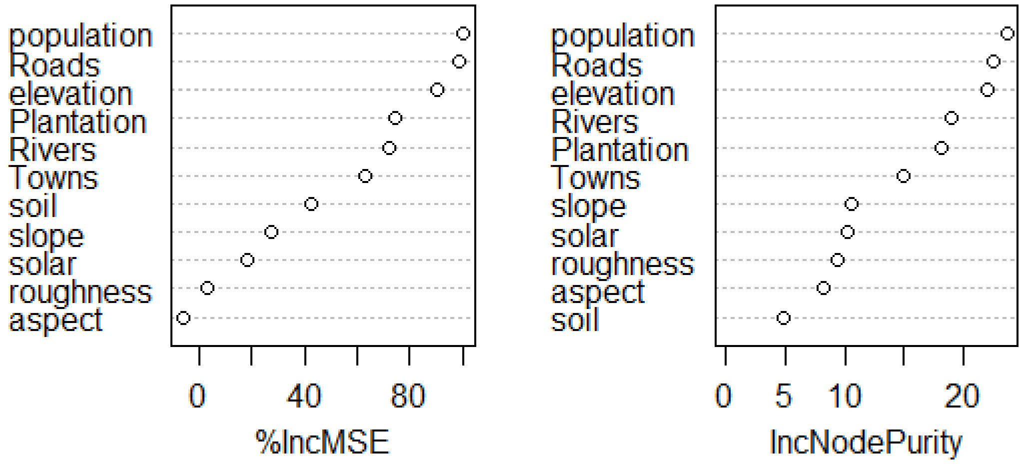
Table 15

Importance of Independent Variables using the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 2000 to 2013

Variable	%IncMSE	IncNodePurity
Solar	20.613879	9.292148
Slope	24.544177	9.885352
Elevation	84.237827	19.856121
Aspect	6.335175	9.130460
Rivers	77.611720	19.355171
Roads	100.406483	23.818125
Towns	64.790544	15.461724
Plantation	81.576744	19.104737
Topographic Roughness	1.477697	8.775994
Soil Type	50.507293	5.766446
Population Density	105.132433	23.993522

Figure 37

Plot of Importance of Independent Variables showing the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 2000 to 2013



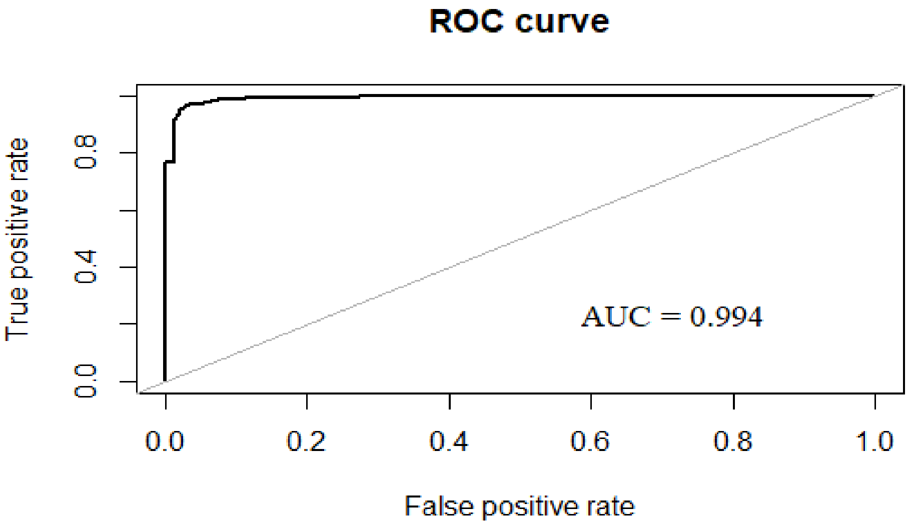
The percentage of variance explained by the variables is calculated at 43.09%, a reduction in comparison to the previous model, however, like the previous model, population density alongside distance to roads, plantations are the most important variables in predicting this model, once more socio-economic factors are the greatest drivers of forest loss in the area. Most physical variables except for elevation did poorly in this model, in a similar fashion to previous model.

Model Assessment

The area under the curve (AUC) of the receiver operating characteristics (ROC) of the model is **0.994** which indicates a good fit of the model with the training data set, it does much better than the average, having a higher positive rate.

Figure 38

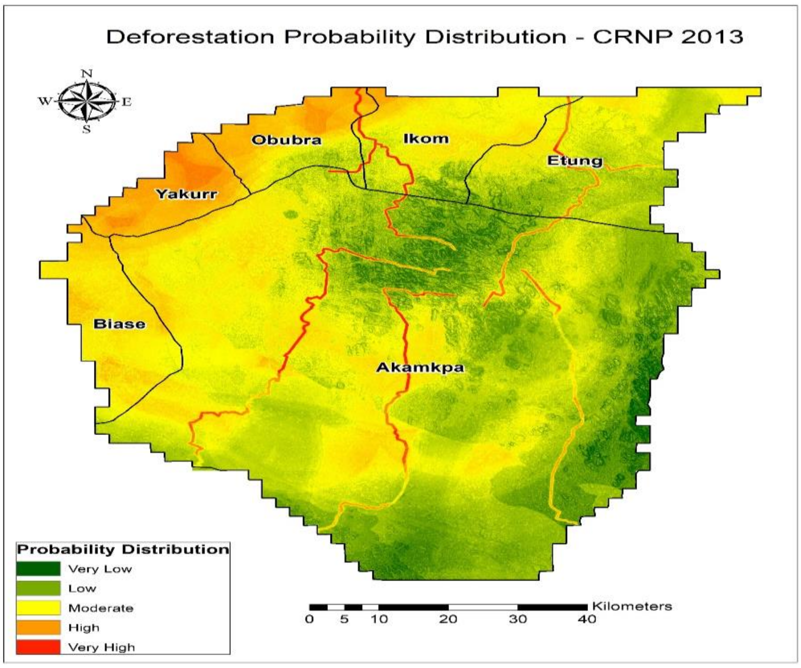
ROC Curve and AUC for Period 2000 – 2013



Using the results from the model, a new set of predicted response variables is plotted to show areas of very high, high, moderate, low, and very low risk for the area of interest for the period beyond 2000-2013.

Figure 39

Deforestation Probability Model – 2000 to 2013



Although this model bears several similarities to the previous model, there are some differences that stand out, for instance the south, southeastern sector towards the republic of Cameroon has low to very low risk probability of forest loss, and the south eastern sector contains the area of forest cover that adjoins the Korup National Park in Cameroon. However, the central part of the park's forest cover towards the northern sector still has an area dominated by low-risk cells, while the north west to western axis covering local governments like Yakurr, Biase and Obubra is dominated by high-risk cells. There are notably no predominant very high-risk cells for this model.

Table 16

Risk Categories and Percentage of Total Area for Model 2000 – 2013

Category	Pixels	Percentage of total area
Very Low	1387658	17%
Low	2891485	34%
Moderate	2963422	35%
High	1074698	13%
Very High	72469	1%
Totals	**8389732**	**100%**

Deforestation Probability Analysis in the 2013-2020 period

For the period 2013-2020, the random forest regression analysis using the binary response and predictor variables produced the following statistical results:

- Mean of squared residuals: **0.09089773**
- % Var explained: **63.64**

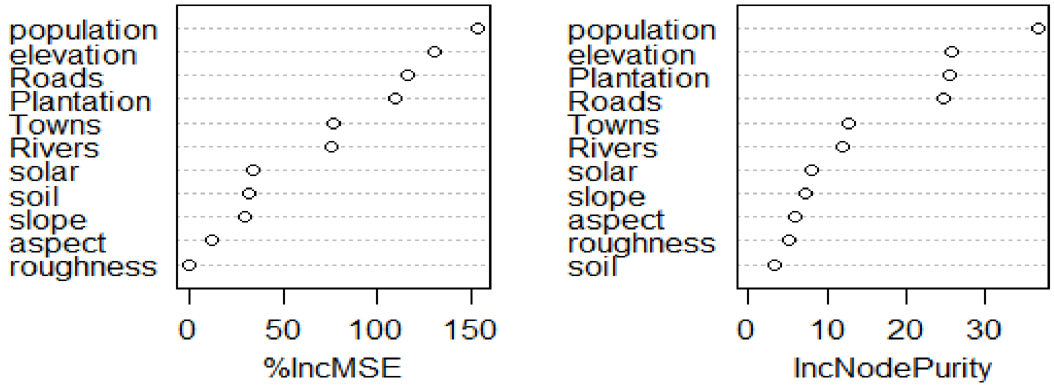
Table 17

Importance of Independent Variables using the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 2013 to 2020

Variable	%IncMSE	IncNodePurity
Solar	34.1479640	8.003302
Slope	29.0243226	7.129393
Elevation	131.1485075	25.781791
Aspect	11.9771946	5.989033
Rivers	75.6432031	11.943034
Roads	115.9904569	24.676363
Towns	76.4652407	12.843857
Plantation	109.5986990	25.468842
Topographic Roughness	-0.5596185	5.245921
Soil Type	31.7749155	3.187714
Population Density	154.0140671	36.731290

Figure 40

Plot of Importance of Independent Variables showing the Mean Decrease Accuracy (%IncMSE) and Gini Impurity Index (IncNodePurity) – 2013 to 2020



The percentage of variance explained by the variables is calculated at 63.64%, higher than the previous model. The predictor variable elevation has the 2nd highest influence on the model unlike in the previous models, with population largely standing out as the most influential in all models. As with all other models' socio-economic factors remains the

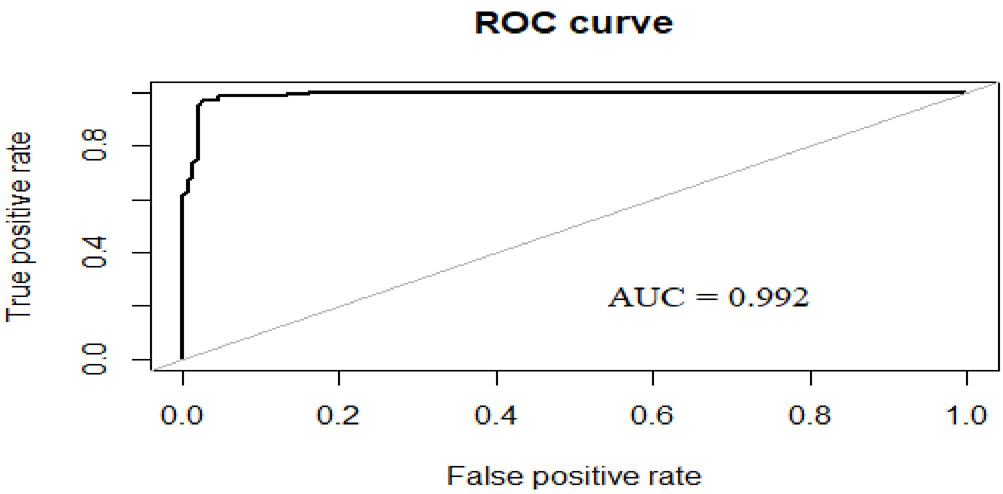
greatest drivers of forest loss in the area while physical factors have lesser impact on influencing forest loss as per the model.

Model Assessment

The area under the curve (AUC) of the receiver operating characteristics (ROC) of the model is **0.992** which indicates a good fit of the model with the training data set, it does much better than the average, having a higher positive rate.

Figure 41

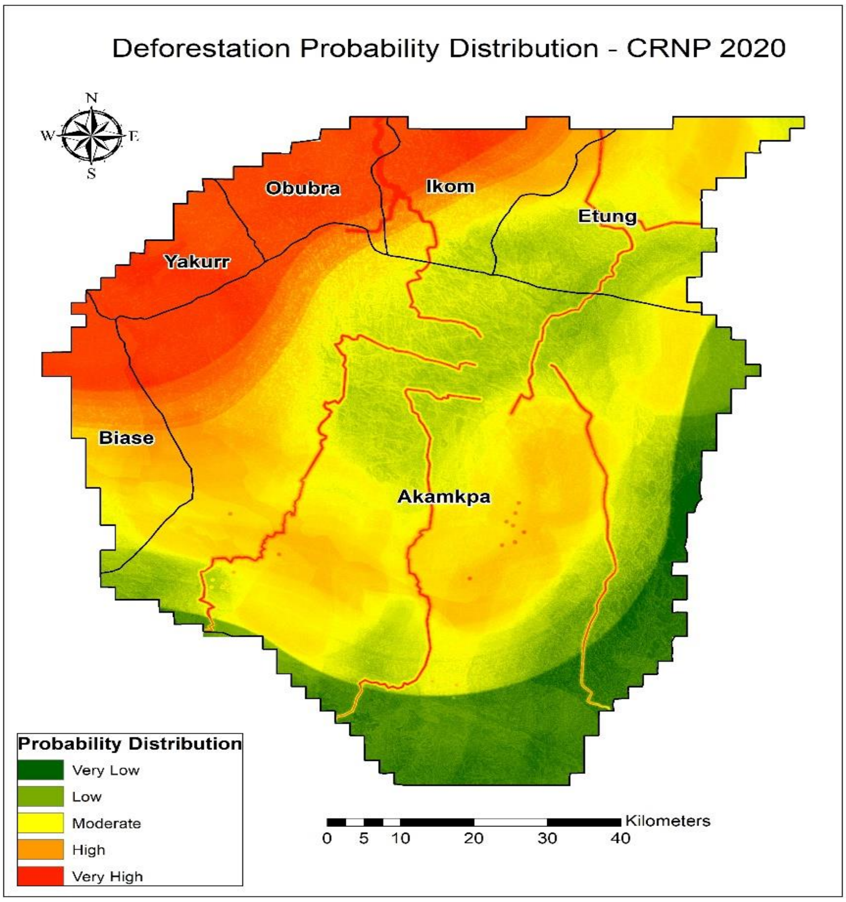
ROC Curve and AUC for Period 2013 – 2020



The 2013-2020 model was used to predict a deforestation probability map of the area of interest, the map bears similarities to the 2000-2013 model and differences as well.

Figure 42

Deforestation Probability Model – 2013 to 2020



Although this model bears several similarities to the previous model, there are some differences that stand out, the map has large parts in the Northwest to Western sector that are dominated by very high-risk pixels, this same area in 2013 was dominated by high-risk pixels, meaning the risk of that area has increased. This area of very high-risk pixels radiates out to the central area of the forest cover. The forest cover class is uniformly surrounded by high and moderate pixels, while the central highlands part of the forest is dominated by low-risk pixels. The South to Southeastern part of the AOI is dominated by very low risk pixels, this borders the Korup National Park in Cameroon.

The probability distribution of the model as categorized into very low, low, moderate, high, and very risk, from the table below it shows the model has more areas of moderate risk, this model has the highest percentage of high-risk pixels compared to the others, this can be seen in the north western sector of the probability map.

Table 18

Risk Categories and Percentage of Total Area for Model 2013 - 2020

Category	Pixels	Percentage of total area
Very Low	1098680	13%
Low	2231080	27%
Moderate	2705950	32%
High	881794	11%
Very High	1472228	18%
Totals	**8389732**	**100%**

Table 19

Area in Km^2 and Percentage of Forested Land Predicted by Models to be affected by Deforestation Considering the Moderate to Very High-Risk Categories.

	2000		2013		2020	
Deforestation Probability Categories	**Area of predicted impact (km^2)**	**Percentage of total forest**	**Area of predicted impact (km^2)**	**Percentage of total forest**	**Area of predicted impact (km^2)**	**Percentage of total forest**
Very Low	252.41	5%	863.08	21%	533.64	15%
Low	1033.90	21%	1557.67	38%	1525.90	42%
Moderate	1486.79	31%	1355.73	33%	1060.16	29%
High	1494.47	31%	180.57	4%	247.24	7%
Very High	467.50	10%	34.56	1%	159.95	4%
Total Forested Area	**4812.64**		**4059.14**		**3595.40**	

Chapter 5: Discussion

Land use changes, especially in the form of deforestation is the second largest anthropogenic source of atmospheric carbon dioxide emissions after fossil fuel combustion. This in turn leads to the myriad of environmental problems the world faces today (CO2 Human Emissions (CHE) Project, 2017).

This study sets out to examine the land cover changes at the Cross River National Park and predict the probability of forest loss in relation to several independent variables across the national park. For the change detection analysis all classified imagery of the area of interest were reduced to four (4) classes: Forested Land, Agriculture/Sparse Land, Water and Settlement. The study used a post classification change detection method to assess the land cover changes within the area, from the analysis it revealed significant changes especially within the forest cover class over the 45-year period of this study.

The trend of forest cover loss shows an area loss of approximately 1,580Km^2 from 1975 to 2020. From information gathered on the field through interviews and news publications this change is attributable to the establishment of new oil palm and pineapple plantations, the expansion of old oil palm and rubber plantations, some of the plantations include the Dansa pineapple plantation (owned by the Dangote Group), Calaro, Biase, Eyop and Ibiae plantations (owned by PZ Wilmer), Kwa Falls, Oban Oil Palm, Ayip Eku, Obasanjo Farms, Nsadop Oil Palm and Borum plantations.

A report by (Ojo et al., 2017) highlights the immense effect these large-scale conversions of forested land has had on the people in the inhabiting the areas occupied by the plantations, the people are now in a sense impoverished as the best soils for subsistent cultivation are taken away which leads to more land grabbing by the communities ultimately reducing further the forest resource.

From the change detection analysis, it is observed that areas around the western, southwestern, and south eastern corridors bounding the forest class have been significantly affected with the agricultural/sparse land class pushing forward to the forest class in these areas. The presence in these areas of many roads and logging trail paths identified through high-resolution imagery can also be identified as a factor in this change. From 1975 to date, it is observed that most of the community forests surrounding the main forest reserve have totally disappeared, while settlements and population have continued to grow exponentially.

Mapping the probability of forest loss

Due to the impact deforestation has on the environment, it is imperative that studies on its future extent and impact must be carried out to preserve this important resource. Having precise data on areas or zones that are susceptible to forest loss can help stakeholders and planners prepare adequately to mitigate this risk.

In this study 3 models have been created for 3 different time series 1975-2000, 2000-2013 and 2013-2020. The random forest regression tree analysis was used to model the probability of forest loss applying both dependent and independent variables, the dependent variable was a binary layer of forest change against forest persistence, while 11 different independent variables are applied in the model.

Model 1 (1975 to 2000)

The first model for the time series 1975-2000 utilized the binary dependent variables and 11 independent variables. The model has population density, distance to roads, distance to nearby plantations and elevation as the top 4 drivers of deforestation.

A look at some historical events in the area of interest explains why these factors are important as the years between 1975 to 2000 showed a double increase in the size of the population (see appendix 3), that would increase resource exploitation. There was also a

significant growth in road networks within this period, opening areas previously inaccessible, especially areas on the edges of the forest. This period also witnessed the creation of various government funded agricultural programs such as “Operation Feed the Nation” instituted by the military government of Nigeria in 1976 and the “Green Revolution Program” in 1980, among other programs within this period granted access to both forested and non-forested land for large agricultural purposes. This is confirmed visually in the probability model showing the south to southeastern sector having areas very high-risk pixels, these are areas heavily affected by rubber, pineapple (within the park) and oil palm plantations. In terms of elevation as an independent variable of deforestation, this factor restricts access to forest resources, as it made it difficult for chain saw loggers to access timber and farmers to grow crops as such the higher the elevation the more likely the probability of low to very low risk pixels.

The predicted trend from the 1975-2000 model as displayed in table 16, shows a forest at a high risk of deforestation considering all factors applied. 72% of the forest is at risk of deforestation in areas ranging from moderate risk to very high risk and this has been proven with the loss of forest resources in land area of approximately 1500km^2 from 1975 to 2020.

Model 2 (2000 to 2013)

The second model is visually different from model 1 in terms of the location of the very low to very high-risk probabilities on the map. This model has no significant very high-risk probabilities and differs visually to the first model. Like the first model Population Density, Roads and Plantation are the top factors influencing forest cover changes, while the influence of Elevation in limiting forest loss has risen in importance in the model. This period has a much higher rate of deforestation (762km^2 of land lost) than the previous period. The

biggest threat during this period was illegal logging in the rain forest, added with lax government control over timber smuggling it created an avenue for wanton destruction of the forests. It was so bad that the then governor of the Cross River State Liyel Imoke had to institute an immediate 2-year logging ban in 2008, seizing illegal timber and impounding logging trucks (Tropical Forest Group, 2011). Population Density remains an important factor, as land is being taken in for agricultural purposes the communities which still rely on the land for survival will look towards the forested land for resources. Also, the network of roads around the borders of the forest and into the forest itself grant access to illegal loggers and the communities to access the forest. A visual assessment of the probability map shows that areas around Obubra, Biase and Yakurr LGA have a high risk of forest loss which spreads easterly to the forest itself, with all other areas surrounding the forest area covered with moderate risk due to dense network of roads in all these areas. The north central area of the forest has low risk pixels, these are the areas with the highest elevation which limits access to forest resource unlike the fringes or borders which are at a lower elevation.

Model 3 (2013 to 2020)

The final model, from 2013 – 2020 shows an alarming trend of growth in very high-risk areas from the previous model. Most areas classified as high risk in the previous model have now been designated as very high risk and stretches further into the forest cover class, thereby reducing areas designated as low risk. This model has the highest level of very high-risk areas at 18% and from the change detection analysis, this was the period with the highest level of deforestation at 768km2 in this study. This period saw a lot of expansion and consolidation of existing and new plantations as foreign and domestic investment in the agricultural sector took place. Some of the plantation take overs started in the previous period, however, their effects were felt in this period as companies like PZ Wilmer bought

and expanded their oil palm land holdings in the area to about 50,000 hectares (Emmanuel Etim, 2015). The Guardian Nigeria in its environment feature report of October 2019 titled “Communities urge Cross River government to reverse ban on logging” highlighted the failure of the logging ban as instituted by the Cross River State Government due to very high levels of corruption and accountability, so bad that the levels of illegal logging and timber smuggling have exceeded former years when the ban was not in place. The article quoting community leaders of the area requested for different tactics to sustainable forest governance in the state (Aniete, 2019).

From this model as reflected in table 16, 40% of the forest cover area is predicted to fall between the moderate to very high risk, this is an equivalent of 1467.35km^2 of forested land lost to deforestation. In a 10-year span it means that annual rate of 146.7 km^2 of forest will be lost yearly and if this is carried forward, then in a time frame of 25 years the forest could be completely lost.

Results from the 3 models show that human factors are the greatest drivers of forest loss within the area of interest compared to physical factors. This is consistent with previous studies by (Enuoh & Ogogo, 2018; Okeke & Imong, 2018; Cross River State/UN REDD+ Program, 2017), which lists a variety human and socio-economic factors such as agricultural expansion, poor public forest management, unsustainable/illegal logging, infrastructure development among others as drivers of forest loss at the CRNP. It is worth noting that human factors outperformed almost all physical factors in the models except for elevation, this contrasts with studies done by (Cushman et al., 2017; Zanelle et al., 2017; Saha et al., 2020) which are in areas far removed from sub-Saharan Africa.

Population density proved to be the most important factor in all 3 models which I found quite surprising when compared other variables such as elevation, proximity to

agricultural plantations, road networks and large towns, this was unanticipated as previous research mentioned above had highlighted the uncontrolled commercial exploitation of forest resources as well as elevated topography of the area which makes the local government area in which the park is located have the lowest population density (see appendix 5). However, studies by (FAO, 2007; Dilip & Dimacha, 2012) empirically showed a direct relation between population density and deforestation in developing economies of the world as such the increase in population and a lax forest conservation policy would make population density an important variable in forest loss or gain.

Inversely while population density has impacted the forest land cover class of the lowland areas within the forest, it is important to note that elevation has kept the forest class in areas of the park with high elevation largely intact, it remains to be seen if road networks are extended into these areas which could then lead to forest losses in these inaccessible areas. Besides elevation other physical factors such as slope, solar radiation, aspect, and topographic roughness did not provide a statistically significant relationship to forest loss or persistence, this is consistent with previous studies mentioned above.

Soil type as a variable driver for forest loss or persistence was also moderately significant in 2 out of 3 models, I had expected that this variable would be very significant as a factor of land cover change/persistence due to the high demand in arable land for agricultural purposes, but it still fell well below the human factors and elevation.

Conclusion

From this study, deforestation at the Oban division of the Cross River National Park in Nigeria has been assessed, with a focus on different physical and anthropogenic factors that drive changes in forest cover. The findings particularly show a direct relationship between population and deforestation in a developing country such as Nigeria, and that

population increase is a primary factor of deforestation. The results of the analysis compare positively with several studies and reports done by the United Nations Food and Agricultural Organization (F.A.O) on deforestation in developing economies. The models produced also highlighted other significant drivers of forest loss such as elevation, proximity to agricultural plantations, proximity to road networks, large settlements, and rivers. As such studies of human demography and socio-economic factors as contributory factors of deforestation is imperative in a developing country such as Nigeria. Stakeholders and planners are encouraged to use such studies to engage communities within and around these forest enclaves for control and preservation of our forest resources for generations to come.

Limitations and Future Studies

The most important limitation on this study was the inability to obtaining *in situ* data from the area of interest to verify all findings obtained by satellite imagery. The use of high-resolution imagery was used as a substitute especially in areas of random point's selection for the dependent variable. Also, demographic figures are hard to obtain in Nigeria as census data is still disputed due to irregularities.

The number of independent factors for deforestation probability assessment are probably a lot more than the scope of this paper can manage, it may be important for future studies to explore more factors such as landscape metrics or atmospheric factors among others to create better models in the study.

Land classification of historical and near present day imagery, readily available for analysis will go a long way to speed and encourage studies into land cover changes. It is hoped in future, research bodies such as EROS or ESA could help with that.

References

Aitkenhead, M. & Aalders, I., 2009. Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. J. Environ. Manag., pp. 236-250.

Amatulli, G., Domisch, S., Tuanmu, M. N., Parmentier, B., Ranipeta, A., Malczyk, J., Jetz, W., 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific data, Volume 5.

Aniete, A., 2019. Guardian Nigeria. [Online] Available at: https://guardian.ng/property/communities-urge-cross-river-government-to-reverse-ban-on-logging/ [Accessed 9 6 2021].

Auwalu, Faisal Koko, Wu, Yue, Ghali, Abdullahi Abubakar, Roknisadeh, Hamed, Akram, Ahmed, Noman Alabsi., 2020. Monitoring and Predicting Spatio-Temporal Land Use/Land Cover Changes in Zaria City, Nigeria, through an Integrated Cellular Automata and Markov Chain Model (CA-Markov). Sustainability.

Barber, C. P., Cochrane, M. A., Souza Jr., C. M. & Laurance, W. F., 2014. Roads, deforestation, and the mitigating effect of protected areas in the Amazon. Biological Conservation, Volume 177, pp. 203-209.

Boori, M. S., Choudhary, K., Kupriyanov, A., Sugimoto, A., Kovelskiy, V., 2016. Monitoring Land Use/Cover Change Detection through Remote Sensing and GIS Techniques in Eastern Siberia, Russia. s.l., s.n., pp. 971-978.

Breiman, L., 2001. Random Forests. Machine Learning, pp. 5-32.

Breiman, L. & Cutler, A., 2004. Random Forests. [Online] Available at: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_manual.htm [Accessed 12 November 2020].

Brooks, T., Mittermeier, R. A., Mittermeier, C. G. & Da Fonseca, G. A. B., 2002. Habitat Loss and Extinction in the Hotspots of Biodiversity. Conservation Biology, p. 16.

Chen, L., Sun, Y. & Sajjad, S., 2018. Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. Plos One.

CO2 Human Emissions (CHE) Project, 2017. Main sources of carbon dioxide emissions. [Online] Available at: https://www.che-project.eu/news/main-sources-carbon-dioxide-emissions [Accessed 4 6 2021].

Congalton, R. & Green, K., 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd Ed. In: Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd Ed. Boca Raton: CRC Press, p. 183.

Cross River State/UN REDD+ Program, 2017. Cross River State REDD Strategy, s.l.: Food and Agricultural Organization.

Cushman, S., Macdonald, E., Landguth, E. M. Y. & Macdonald, D., 2017. Multiple-scale prediction of forest loss risk across Borneo. Landscape Ecology, pp. 1581-1598.

Cutler, A. & Stevens, J., 2006. Random forests for microarrays. In: Methods in Enzymology. s.l.:AP, pp. 411: 422-432.

Datacamp Inc, n.d. introduction-to-machine-learning-in-r. [Online] Available at: https://campus.datacamp.com/courses/introduction-to-machine-learning-in-r/how-much-will-i-earn?ex=6 [Accessed 05 06 2021].

Deng, J., Wang, K., Deng, Y. & Qi, G., 2008. PCA based land use change detection and analysis using multitemporal and multisensor satellite data. International journal of Remote Sensing, pp. 4823-4838.

Dilip, N. C. & Dimacha, D., 2012. Population Increase and Deforestation: A Study in Kokrajhar District of Assam, India. International Journal of Scientific and Research Publications, 2(10).

Ellis, E., 2007. Land-use and land cover change and climate change. The encyclopedia of earth, p. 4.

Emmanuel Etim, 2015. Cross River Watch-Cross River Oil Palm Sector Success Story and More That Can Be Done. [Online] Available at: https://crossriverwatch.com/2015/08/cross-river-oil-palm-sector-success-story-and-more-that-can-be-done-by-emmanuel-etim/ [Accessed 6 9 2021].

Enuoh, O. & Ogogo, A., 2018. Assessing Tropical Deforestation and Biodiversity Loss in the Cross River Rainforest of Nigeria. Open Journal of Forestry, Volume 8, pp. 393-408.

FAO, 2007. Manual on Deforestation, Degradation and Fragmentation using Remote Sensing and GIS, Rome: F.A.O.

FAO, 2010. Global Forest Resources Assessment, Rome: United Nations Food and Agriculture Organization.

Friedl, Mark A., Damien, Sulla-Menashe, Bin Tan, Annemarie, Schneider, Navin, Ramankutty, Adam, Sibley, Xiaoman, Huang., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment, pp. 168-182.

Fu, W., Liu, S. & Degloria, S., 2010. Characterizing the "fragmentation-barrier" effect of road networks on landscape connectivity: A case study of Xishuangbanna, Southwest China. Landscape Urban Plan, pp. 95:122-129.

Gilbert, A. & Chakraborty, J., 2011. Using geographically weighted regression for environmental justice analysis: cumulative cancer risks from air toxics in Florida. Social Science Research, pp. 40: 273-286.

Gislason, P. O., Benediktsson, J. A. & Sveinsson, J., 2006. Random Forest for land cover classification. Pattern Recognition Letters, pp. 294-300.

Grinand, C, Vieilledent, G, Razafimbelo, T, Rakotoarijaona, J-R, Nourtier, M, Bernoux, M., 2020. Landscape-scale spatial modelling of deforestation, land degradation, and regeneration using machine learning tools. Land Degrad Dev., pp. 1699-1712.

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., Hokao, K., 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. Ecol. Model., pp. 3761-3772.

Hanley, J. A. & McNeil, B. J., 1982. The meaning and use if the area under a reciever of operating characteristic (ROC) curve. Radiology, Issue 143, pp. 27-36.

Ho, T. K., 1995. Random Decision Forests. Montreal, s.n., pp. 278-282.

Hyandye, C., 2015. GIS and Logit Regression Model Applications in Land Use/Land Cover Change and Distribution in Usangu Catchment. American Journal of Remote Sensing, pp. 6-16.

Jensen, J. R., 2015. Introductory Digital Image Processing: A Remote Sensing Perspective 4th Edition. s.l.:Pearson Series.

Jha, S. & Bawa, K., 2006. Population growth, human development, and deforestation in biodiversity hotspots. Conserv Biol., p. 906.

Juergen Blaser, C. R., 2006. Reduction of GHG Emissions through avoiding Deforestation and Forest Degradation (REDD): Committing forests as carbon reservior, Accra: United Nations Food and Agricultural Organization.

Kamusoko, C. & Aniya, M., 2008. Hybrid classification of Landsat data and GIS for land use/cover change analysis of the Bindura district, Zimbabwe. International Journal of Remote Sensing, pp. 97-115.

Linling, Sang, Chao, Zhang, Jianyu, Zhang, Dehai, Zhu, Wenju, Yun, 2011. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Mathematical and Computer Modelling, pp. 938-943.

Lunetta, Ross S., Ediriwickrema, Jayantha, Johnson, David M, Lyon, John G, McKerrow, Alexa., 2002. Impacts of vegetation dynamics on the identification of land-cover change in a biologically complex community in North Carolina, USA. Remote Sensing of Environment, pp. 258-270.

Lunetta, Ross S, Knight, Joseph F, Ediriwickrema, Jayantha, Lyon, John G, Worthy, Dorsey L., 2006. Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, pp. 142-154.

Mahmoud, Mahmoud I., Sloan, Sean, Campbell, Mason J., Alamgir, Mohammed, Imong, Inaoyom., 2017. Alternative Routes for a Proposed Nigerian Superhighway to Limit Damage to Rare Ecosystems and Wildlife. Tropical Conservation Science.

MathWorks, n.d. Detector Performance Analysis Using ROC Curves. [Online] Available at: https://www.mathworks.com/help/phased/ug/detector-performance-analysis-using-roc-curves.html [Accessed 1 June 2021].

Mengistu, D. A. & Salami, A. T., 2007. Application of remote sensing and GIS inland use/land cover mapping and change detection in a part of south west Nigeria. African Journal of Environmental Science Technology, pp. 99-109.

Myers, N., 1980. Conversion of Moist Forests. National Academy of Sciences, p. 250.

Myers, N., 1993. Tropical Forests: The main deforestation fronts. Environmental Conservation, 01 03.p. 9.

National Geographic Society, 2020. Resource Library - Biodiversity: National Geographic. [Online] Available at: https://www.nationalgeographic.org/encyclopedia/biodiversity/#:~:text=Biodiversity%20is%20a%20term%20used,bacteria%2C%20animals%2C%20and%20humans.

NNPS, 2020. UnitParks: Nigeria National Park Service. [Online] Available at: http://nigeriaparkservice.org/?p=140

Oates, J. F., Bergl, R. A. & Linder, J. M., 2004. Africa's Gulf of Guinea Forests: Biodiversity Patterns and Conservation Priorities. 1 ed. Washington DC: Conservation International: Center for Applied Biodiversity Science.

Ojo, G.U., Offiong, R.A., Akhaine, S.O., Abiodun, Baiyewu-Teru, Fidelis, Allen., 2017. Oil palm plantations in forest landscapes: impacts, aspirations and ways forward in Nigeria, Wageningen: Tropenbos International.

Okeke, F. & Imong, I., 2018. Analysis of Land Cover Change in the Oban Division of Cross River National Park and Adjacent Forest Areas, s.l.: Wildlife Conservation Society: Nigeria Program.

Paine, D. P. & Kiser, J. D., 2003. "Chapter 23: Mapping Accuracy Assessment," Aerial Photography and Image Interpretation, 2nd Ed.,. In: "Chapter 23: Mapping Accuracy Assessment," Aerial Photography and Image Interpretation, 2nd Ed.,. New York: John Wiley & Sons, pp. 465-480.

Peace Corp. Information Collection & Exchange, 1990. Reforestation in the Pacific Islands - A manual for Peace Corps Volunteers Involved in Community Forestry Projects in Papua New Guinea, Tonga, Fiji, Western Samoa and others. Philadelphia: Rogers, Golden & Halpern, Inc.

Prasad, A., Iverson, L. & Liaw, A., 2006. Newer tree classification and techniques: forests random prediction bagging for ecological regression. Ecosystems, pp. 181-199.

Ralha, C., Abreu, C., Coelho, C., Zaghetto, A., Macchiavello, B., Machado, R. A., 2013. Multi-Agent Model System for Land-Use Change Simulation. Environmental Modelling Software, pp. 30-46.

Rather, T., Kumar, S. & Khan, J., 2020. Multi-scale habitat modelling and predicting change in the distribution of tiger and leopard using random forest algorithm. Scientific Reports, p. 10.

Riley, S. J., DeGloria, S. D. & R., E., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of Sciences, Volume 5, pp. 1-4.

Saha, Sunil, Saha, Mantosh, Mukherjee, Kaustuv, Arabameri, Alireza, Thao Thi Ngo, Phuong, Gopal, Chandra Paul., 2020. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. Science of the Total Environment.

Sandel, B. & Svenning, J., 2013. Human impacts drive a global topographic signature in tree cover. Nature Communications, 4(2474).

Sealey, K., Binder, P. & Burch, K. P., 2018. Financial credit drives urban land-use change in the United States. Anthropocene, pp. 42-51.

Shamsi, S., 2010. Integrating Linear Programming and Analytical Hierarchical Processing in Raster-GIS to Optimize Land Use Pattern at Watershed Level. Journal of Applied Science & Environmental Management, pp. 81-85.

Shimabukuro, Yosio E, Espírito-Santo, Fernando D.B, Gloor, Manuel, Keller, Michael, Malhi, Yadvind, Saatchi, Sassan, Nelson, Bruce, Junior, Raimundo C. Oliveira, Pereira, Cleuton, Jon Lloyd, Frolking, Steve, Palace, Michael, Valdete Duarte, Abel Monteagudo Mendo., 2014. Size and frequency of natural forest disturbances and the Amazon forest carbon balance. Nature Communications.

Singh, S.K., Mustak, S., Srivastava, P.K., Szabó, S., Islam, T., 2015. Predicting Spatial and Decadal LULC Changes Through Cellular Automata Markov Chain Models Using Earth Observation Datasets and Geo-information. Environ. Process., pp. 61-78.

Smith, S., Ellis, N. & Pitcher, C., 2011. Conditional variable importance in R package extendedForest, s.l.: s.n.

Souza, Jr, 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., Barlow, J., 2013. Ten-Year Landsat Classification of Deforestation and Forest Degradation in the Brazilian Amazon. Remote Sensing.

Statology, 2019. How to Calculate Variance Inflation Factor (VIF) in R. [Online] Available at: https://www.statology.org/variance-inflation-factor-r/ [Accessed 3 6 2021].

Subedi, P., Subedi, K. & Thapa, B., 2013. Application of a Hybrid Cellular Automaton—Markov (CA-Markov) Model in Land-Use Change Prediction: A Case Study of Saddle Creek Drainage Basin, Florida. Appl. Ecol. Environ. Sci., pp. 126-132.

Tropical Forest Group, 2011. LOGGING MORATORIUM CROSS RIVER STATE NIGERIA. [Online] Available at: http://www.tropicalforestgroup.org/logging-moratorium-cross-river-state-nigeria/ [Accessed 9 6 2021].

Unah, L., 2019. In Nigeria, a highway threatens community and conservation interests. [Online] Available at: https://news.mongabay.com/2019/07/in-nigeria-a-highway-threatens-community-and-conservation-interests/#

University of Illinois, C. o. A. C. a. E. S. (., 2016. Science Daily: Rainforest greener during 'dry' season. [Online] Available at: www.sciencedaily.com/releases/2016/07/160726124011.htm

USGS, 2020. Landsat Science Products: USGS. [Online] Available at: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-surface-reflectance?qt-science_support_page_related_con=0#qt-science_support_page_related_con

Wang, F., 1993. A knowledge-based vision system for detecting land changes at urban fringes. IEEE Transactions on Geoscience and Remote Sensing, 31(1), pp. 136-145.

Yang, X., 2002. Satellite Monitoring of Urban Spatial Growth in the Atlanta Metropolitan Area. Photogramm. Eng. Remote Sens, pp. 725-734.

Yang, X., Zheng, X.-Q. & Lv, L.-N., 2012. Aspatiotemporal model of land use change based on ant colony optimization Markov chain and cellular automata. Ecol. Model, pp. 11-19.

Yichun, X., Zongyao, S. & Mei, Y., 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology, pp. 9-23.

Zanelle, L., Folkard, A. M., Blackburn, G. A. & Carvalho, L. M., 2017. How well does random forest analysis model deforestation and forest fragmentation in the Brazilian Atlantic forest?. Environ Ecol Stat, pp. 529-549.

Zou, K. H., O' Malley, J. A. & Mauri, L., 2007. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. AHA Journals, Issue 115, pp. 654-657.

Appendix A

Class	Settlement	Water	Agriculture/Sparse Land	Forest	**Total (User)**
Settlement	5	0	3	0	**8**
Water	0	11	1	4	**16**
Agriculture/Sparse Land	0	0	12	2	**14**
Forest	0	0	1	16	**17**
Total (Producer)	**5**	**11**	**17**	**22**	**55**

Appendix A: Confusion Matrix showing the total user and producer pixels of the classification scheme applied to the Landsat 8 OLI imagery.

Appendix B

Class	Producer Accuracy	User Accuracy
Settlement	100.00%	62.50%
Water	100.00%	68.75%
Agriculture/Sparse Land	70.58%	85.71%
Forest	72.72%	94.12%

Overall Accuracy	80.00%
Kappa Coefficient	72.50%

Appendix B: Producer, user and over accuracies derived from the confusion matrix of the classification scheme. The producer and user accuracy are provided for each class while the overall accuracy and kappa coefficient are provided for the whole classification scheme.

Appendix C

Year	Population	Yearly % Change	Yearly Change	Median Age	Fertility Rate	Density (P/Km²)	Urban Population
2020	206,139,589.00	2.58%	5,175,990.00	18.1	5.42	226	107112526
2019	200,963,599.00	2.60%	5,088,916.00	17.9	5.67	221	102805995
2018	195,874,683.00	2.62%	5,001,439.00	17.9	5.67	215	98610801
2017	190,873,244.00	2.64%	4,913,003.00	17.9	5.67	210	94525016
2016	185,960,241.00	2.66%	4,822,793.00	17.9	5.67	204	90546177
2015	181,137,448.00	2.71%	4,526,850.00	17.9	5.74	199	86673094
2010	158,503,197.00	2.68%	3,927,636.00	17.9	5.91	174	68949828
2005	138,865,016.00	2.58%	3,316,233.00	18	6.05	152	54288918
2000	122,283,850.00	2.53%	2,867,103.00	17.9	6.17	134	42627440
1995	107,948,335.00	2.54%	2,547,177.00	17.7	6.37	119	34785545
1990	95,212,450.00	2.64%	2,329,933.00	17.4	6.6	105	28276132
1985	83,562,785.00	2.62%	2,027,830.00	17.5	6.76	92	21434266
1980	73,423,633.00	2.99%	2,009,867.00	18	6.76	81	16139321
1975	63,374,298.00	2.51%	1,478,431.00	18.3	6.61	70	12535584

Appendix C: Nigerian population growth estimations 1975 – 2020 obtained from the Nigerian National Population Commission.

Appendix D

Local Government Area	**1991**	**2006**	***2016**	**Area (Km^2)**	**Population Density (2016)**
Akamkpa	118,472	149,705	200,100	4,586	43.6
Biase	101,121	168,113	224,700	1,302	172.6
Etung	...	80,036	107,000	969.4	110.4
Ikom	...	163,691	218,800	1,794	121.9
Obubra	134,225	172,543	230,600	1661	138.8
Yakurr	...	196,271	262,300	659.6	397.7

Appendix D: Population figures for the 6 Local Government Areas in the AOI (* 2016 projected population as the 1991 and 2006 figures are disputed till date)