Lappeenranta University of Technology
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Chepkoech Ann

FOREST MAPPING A CASE STUDY OF SOUTHERN HIGHLANDS OF TANZANIA

Master's Thesis

Examiners: Professor Tuomo Kauranne
Dr. Virpi Juntilla

Supervisor: Tuomo Kauranne (Tech.)

ABSTRACT

Lappeenranta University of Technology
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Chepkoech Ann

Forest Mapping a Case Study of Southern Highlands of Tanzania

Master's Thesis

2017

29 pages, 12 figures, 1 table.

Examiners: Professor Tuomo Kauranne
Dr. Virpi Juntilla

Keywords: Landsat Thematic Mapper Plus, Sparse Bayesian, Geospatial, Satellite data

Forest mapping is one of the major areas where we try to understand the key elements of forest. Different researchers study this using different geospatial tools and its application in the planning and management of the resources. Always this process needs good and accurate understanding of the various forest parameters. This present study focuses on studying forest plantations of Southern Highlands in Tanzania. The data used was collected in three pilot areas; Njombe, Makete and Kilolo. We study the age of the plantation in relation to satellite data. Landsat 8 Enhanced Thematic Mapper Plus $(ETM+)$ of three different years has been used in collecting the data from the three pilot areas. The Sparse Bayesian method has been applied as a modelling tool in studying the correlation between plantation age and satellite data. Plantation age has been grouped in three classes; young $(0-3)$, growing $(3-8)$ and old (>8). The results shows that there is very little correlation between the age of plantation forest and satellite data.

PREFACE

I would like to take this opportunity first to thank our Almighty for having given me good health during the time of my studies here and the good understanding He blessed me with for my work. My sincere gratitude goes to my supervisor Prof. Tuomo Kauranne for his support and guidance during this period and always ensuring that I produce quality work. Many thanks to Dr. Virpi Junttila for her guide with understanding how the method of Sparse Bayesian works.

Special thanks go to my dear Jeffrey Kipkemboi for always giving me support and encouragement when I faced challenges and always ensuring that I did well here at LUT.

Lastly, I would like to thank my parents, brothers and sisters who have always been giving me encouragement and praying for me. Thank you very much and may the Almighty bless you all.

Lappeenranta, May 25, 2018

Chepkoech Ann

CONTENTS

ABBREVIATIONS AND SYMBOLS

QGIS	Quantum Geographical Information System
GIS	Geographic Information System
NDVI	Normalized Difference Vegetation Index
FAO	Food and Agricultural Organization
PFP	Private Forestry Programm
RS	Remote Sensing
UN	United Nations
ETM+	Enhanced Thematic Mapper Plus
UTU	University of Turku
REDD+	Reducing Emissions from Deforestation and Forest Degradation
NAFORMA	National Forest Monitoring and Assessment of Tanzania
ERDAS	Earth Resource Development Assessment System

1 INTRODUCTION

1.1 Background

Forest refers to a land occupying more than $0.5ha$ consisting of trees measuring $5m$ height and having a canopy cover that is greater than 10% without including land that is being used for agricultural practices and other activities [1]. Forest is one of the important natural resources and it plays a major role especially in balancing the ecosystem. It is also important in the provision of goods and other products for example timber and paper. According to Acharya [2], forests play an important role in maintaining the ecological cycles by controlling the flow of energy and materials. Also, Ochego [3] shows how forests are important as they contribute to agriculture, manufacturing and processing industry and towards the tourism industry which improves the economy of the country. Tanzanian forest is important in terms of employment since it creates job opportunities for almost 3 million people in the country per year [4]. Different countries differ in the way they use their forest resources depending on the type of forest available for them [5].

Tanzania is one of the developing countries that is blessed with most natural resources as shown in [4] with forest covering a greater percentage which contributes towards the development of country. Plantation of forests in the country started long time ago during the colonial rule and up to date people are still practising the activity of planting trees. Total forest plantation coverage is estimated to be $250,000ha$, out of this the government itself owns about $85,000ha$ while the privately owned forest is about $40,000ha$ and the remaining percentage belong to the outgrower schemes [4]. Most part of the planted areas in Tanzania is in the Southern Highlands with the most planted species being pine and eucalyptus. Public forest plantation has in a way changed or remained almost constant for the past 40 years dating back from 2011 due to various constraints [4] while on the other hand, the privately owned forests are increasing at a higher rate. This calls for the need of studying and showing the distribution of forest plantations which can help in the management, planning and allocation of new lands for the plantation which can only be possible using Geospatial tools. These can help in the future during harvesting and planning and also for the preparation of reforestation projects [6].

This practice of forest plantation always needs one to have reliable sources of information which can later on help in the decision making at the planning stage. Remote Sensing (RS) and Geographic Information System (GIS) are some of the well identified tools that can be used in the analysis and in the management of land related data due to their uniqueness of handling such kind of data [7]. Most of the scientists have been able to study the behavior of forests using these tools. According to Kandel [6], satellite RS provides strong and reliable techniques that can be used in monitoring of forests. Also, with the help of time series satellite data, one can be in a position to study the long term changes in the forest cover [8]. Maurya et al. [9], studied monitoring and predicting forest cover in the Bilaspur district of Himachal Pradesh using RS and GIS as a tool with their objective being to monitor the forest cover changes. Junttila [5] stated that the use of RS data is the best since it can cover a given region of interest which can later be used in studying various characteristics of forests. Forests are always facing challenges caused by either cultural, social or economic factors which calls for better understanding of these forces towards improving afforestation [10]. Forests can be identified according to their current state and the way they change over time with their different characteristics being defined as their state while the way in which they change over time being temporal dynamics [11]. Getting to know well these dynamics will help in the management and planning of future. They further stated that for this to be achievable, the use of RS data can be employed as a way of estimating baseline forest status so as to assess the current situation and trend at which the forest is changing.

Studying the age of the forests is also one of the key roles in taking care of the resources especially when people plan to harvest timber. To do this, many researchers use RS satellite data to study these features and by combining it with other geospatial tools and other mathematical modelling approaches, different scientists have been able to study forest mapping in different countries. [12] studied the role of ecological applications of RS by trying to show how important the Landsat imageries are in modelling of forests. [13] Estimated the age and the structure of forests in Western Orengo using different experimental techniques. They did unsupervised classification using Earth Resource Development Assessment System (ERDAS) software and grouping the age factors into different groups; open, semi-open and closed forests. Another study is by [14] where they mapped tropical dry forest by studying the disturbance type and the age of forest using satellite data. They used imageries with no clouds and were able to apply regression analysis with tree normalization.

This study focuses on studying the age of forest plantation in areas of Kilolo, Makete and Njombe and modelling it using the Sparse Bayesian Method [5] with GIS tools and Satellite data. Landsata $8ETM+$ for three years $2013 - 2015$ is applied for the analysis. We are focused on studying the relationship between age using ground data and this satellite data obtained from the three years. With this method, we will assume a simple regression model that can help distinguish the three classes of forests plantations; young, growing and the mature ones.

1.2 Statement of the Problem

Forests are one of the major contributing factors to the development of a country and also towards the improvement of the well being of human beings especially those living in the rural areas. It helps in reducing poverty rates especially for the developing countries due to the fact that it always acts as a water catchment which helps improve agriculture in the regions that are forested or near forests. Therefore according to [6], this implies that the forest resources need to be well managed so as to fulfill the needs of increasing population.

Tanzanian forests are always under pressure due to the expanding and growing population which makes it to be a focus of Reducing Emissions from Deforestation and Forest Degradation (REDD$+$) activities [15]. Tanzania, being one of the developing countries, faces a big challenge in timber production which can only be met by forest planting [16]. Looking at the Southern Highlands of Tanzania, most of the forested lands are planted with the privately owned and commercial forests covering a greater percentage. The problem with this area is that these plantations coverage is not well known [17] and thus, due to this reason, there is a need of studying the changes in forest resources. There is also one major problem that can be encountered with studying the distribution of these forest resources which is the missing information on how large or to what extent it covers, how old the plantation is and other factors that can be related to forest. This later on makes it very hard to have proper planning and distribution of lands for future planning and management.

Generally, forest loss in Tanzania per year is approximately given as $403,000ha$ [1]. Table 1 below shows the estimated change in the forest cover from $1990 - 2010$.

Table 1. A table showing estimated forest loss in Tanzania from $1990 - 2010$ [1]

Forest area (million ha)			
1990	2000	2005	2010
41.5	37.5	35.4	33.4

Many researchers have been concentrating on studying forest cover changes, deforestation and degradation in different countries, less study has been shown on concentrating on forest plantations and studying their age. In addition, studies combining satellite data and using GIS tools together with Sparse Bayesian methods are very rare.

This report focuses on studying the age of the forest plantations in the Southern Highlands of Tanzania by using Landsat band values extracted from satellite images of the three pilot areas and it further explains how the Sparse Bayesian method can be used to study the correlation between these band values and the plantation age. It further elaborates more on how the method is applied to the study.

1.3 Objectives

The objective of this study is to estimate the age of the forest plantation using Sparse Bayesian technique.

1.4 Structure of the Thesis

This study tries to study the forest plantation age for three years $2013 - 2015$ in Southern Highlands of Tanzania using the Sparse Bayesian method. The satellite images used are Landsat Enhanced Thematic Mapper plus ($8\ ETM+$) for the corresponding years.

Introduction (Chapter 1) introduces us to the topic. It gives an outline of the benefits of forests and discusses forest plantation in the study area. The main objective of the study is also provided in this chapter. It further gives a summary of research that has been carried out by other researchers related to the study and various techniques employed including the use of RS and GIS techniques. This chapter also describes the gap that exists with the other studies carried out and tries to show in connection to our objective what has never been covered. This is also where we have the outline of the study given.

Chapter 2 describes the study area and its various physical characteristics that can be connected with the study. It further provides a map that shows the three areas of interest where the data was collected and their location on the entire region of Southern Highlands.

Chapter 3 describes the methods that were employed in the study so as to achieve our objective. It describes materials, data used and the model that was employed in the analysis. The satellite images used were Landsat $8\ ETM+$ collected within a period of 3 years which is from 2013 to 2015. It further gives details on how the sampled points that fell on the plantation masks were extracted using the QGIS tool and how the band values were extracted too. It also gives a brief description of the method used in getting the Normalized Difference Vegetation Index (NDVI).

Chapter 4 presents the analysis and the various outputs. It shows the results for all the methods that were tested to obtain the results.

Chapter 5 gives a summary of the work and draws conclusions from the results provided with regards to the study.

2 STUDY AREA

In this study, we use data that was collected by the Food and Agricultural Organization (FAO) and University of Turku (UTU) in Southern Highlands where they mapped the entire area. According to the Private Forestry Program (PFP) study, most of the biggest plantation industries are situated in the region [17]. They also provided three pilot areas which were in our interest which were situated along the three administrative regions, Mbeya, Njombe and Iringa.

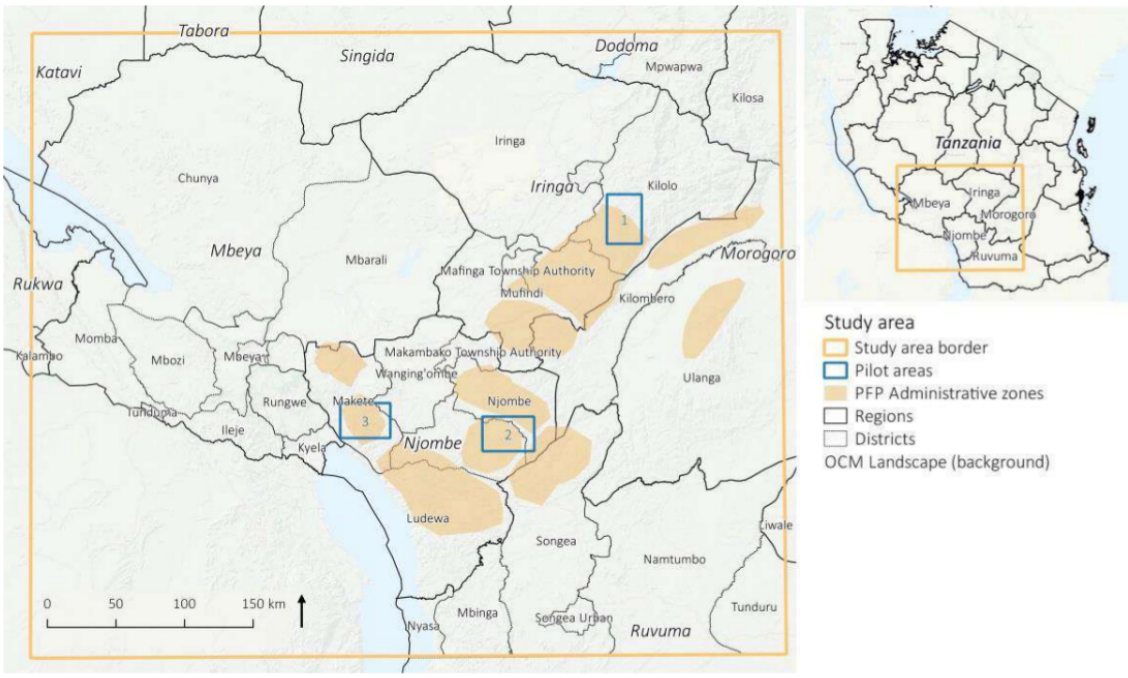


Figure 1. Study Area.

The Southern Highlands of Tanzania lie in between Latitude $6^0S - 12^0S$ and Longitude 29^0E and 38^0E. It consists of four districts namely: Iringa, Mbeya, Rukwa and Ruvuma. According to Mbululo and Nyihilani, the region is made up of volcanic mountains which are partly covered in forest and grasslands [18]. It receives rainfall that are linked with the temperature of the Indian Ocean. They stated that the Southern Highlands experiences a uni-modal type of rainfall pattern which starts in the month of November and ends in the month of April. This type of rain is only experienced in the regions of Mbeya, Rukwa and Iringa.

The region has the highest forest plantation density as shown by PFP in 2016 which shows that commercial and privately operated forests covered $207,000ha$ with most plantations located in Njombe and Iringa $(175,762ha)$ [17]. Ngaga [4] showed that most of the planted species in the region are pine and eucalyptus with pines being dominant species covering 78% and the remaining 22% being other species. The height of these species vary from plantation to plantation [19].

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 Data Description

The data used was based on the study conducted by UTU and FAO of the United Nations (UN) where they mapped the whole area of Southern Highlands between June and December 2016 [17]. They created baseline information regarding the geographical extent of forest plantations of the Southern Highlands. The process involved two levels; the first is where they created plantation maps for the forests and the second is where large sample data was collected to improve the accuracy. The second step had further two levels where first mapping of plantations for the whole area was done at a spatial resolution of $30m$ while the second level was where they mapped forest plantations for the three pilot areas at a higher spatial resolution of $10m$. The data collected from the training sample of forest plantations were done on high resolution using Open Foris Collect Earth. The landcover classification was based on National Forest Monitoring and Assessment of Tanzania (NAFORMA) with plantations based on different characterictics (species, age and density). The data from the three pilot areas and the plantation mask were the ones we employed in our study to make our analysis.

We obtained data in different methods as follows: firstly, we needed to obtain those plots that fall within the plantantion mask that was provided by the study carried out by FAO and University of Turku. This was done by placing the plots on the mask and extracting them in QGIS by using the sparse query tool as seen in figure 2.

Figure 3 shows how the plots were chosen in QGIS for each of the pilot areas.

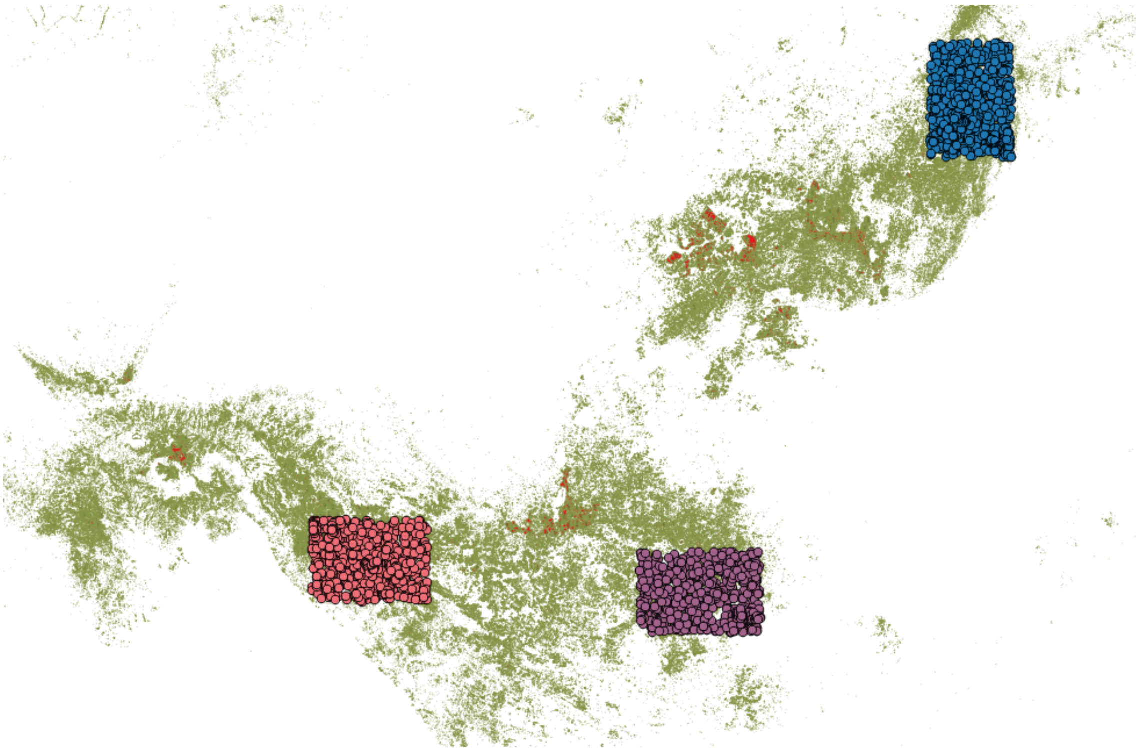


Figure 2. Points on the mask.

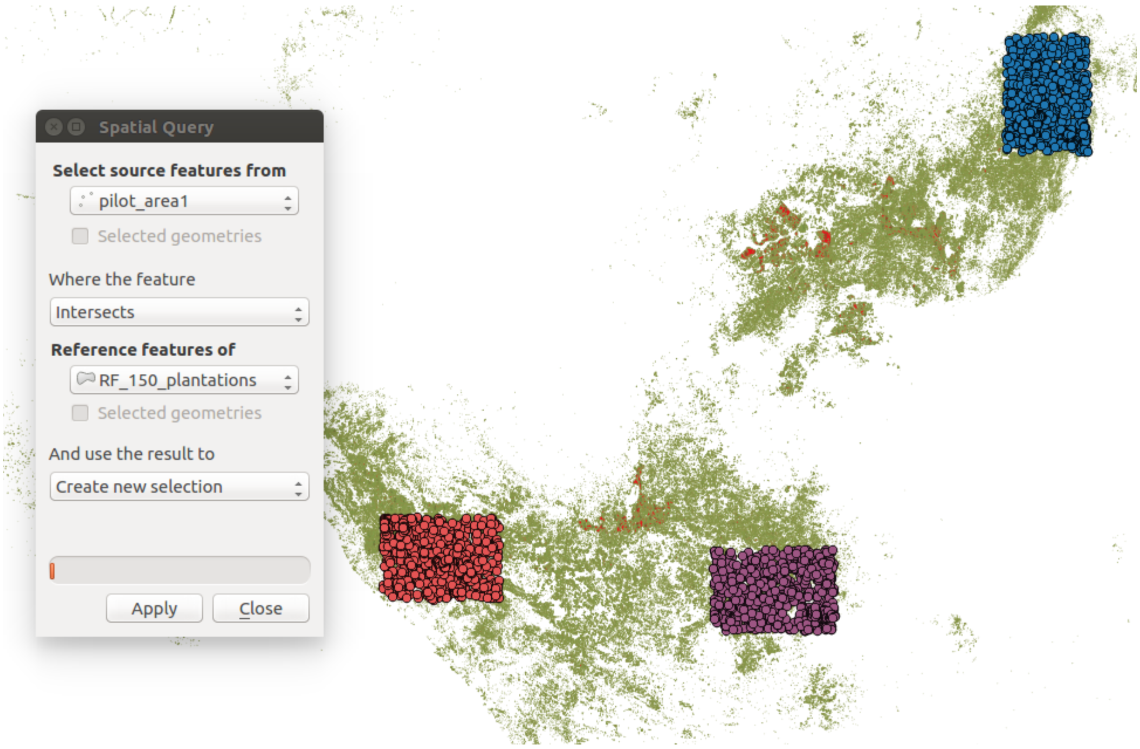


Figure 3. Selecting the plots.

Finally those plots that fall on the plantations were identified and can be seen in figure 4.

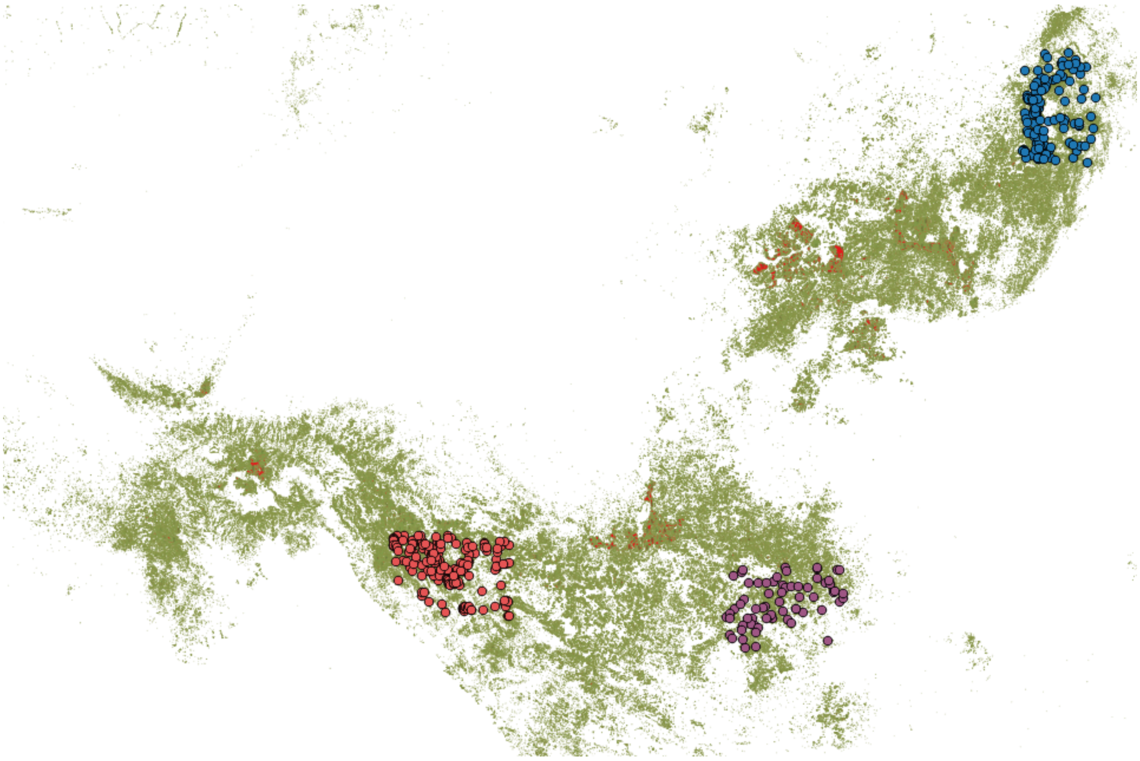


Figure 4. Plots identified.

There were 819 plots in pilot area 1, 657 plots in pilot area 2 and 762 plots in pilot area 3. After running the process to identify the plots falling on plantation, only 135 plots fell on plantation in pilot area 1, 70 plots fell on plantation in pilot area 2 and finally 189 fell on plantation for pilot area 3. After identifying the plots with that fell on the plantation mask, since our objective was to use satellite image data in estimating the age of the plantation, we needed to extract band values using these plots with the satellite image. The satellite images used in this study were Landsat 8 ETM for the years $2013 - 2014$. These were extracted in QGIS using the point sampling tool as shown in figure 5. We were only interested in using bands $2, 3, 4, 5, 6, 7$.

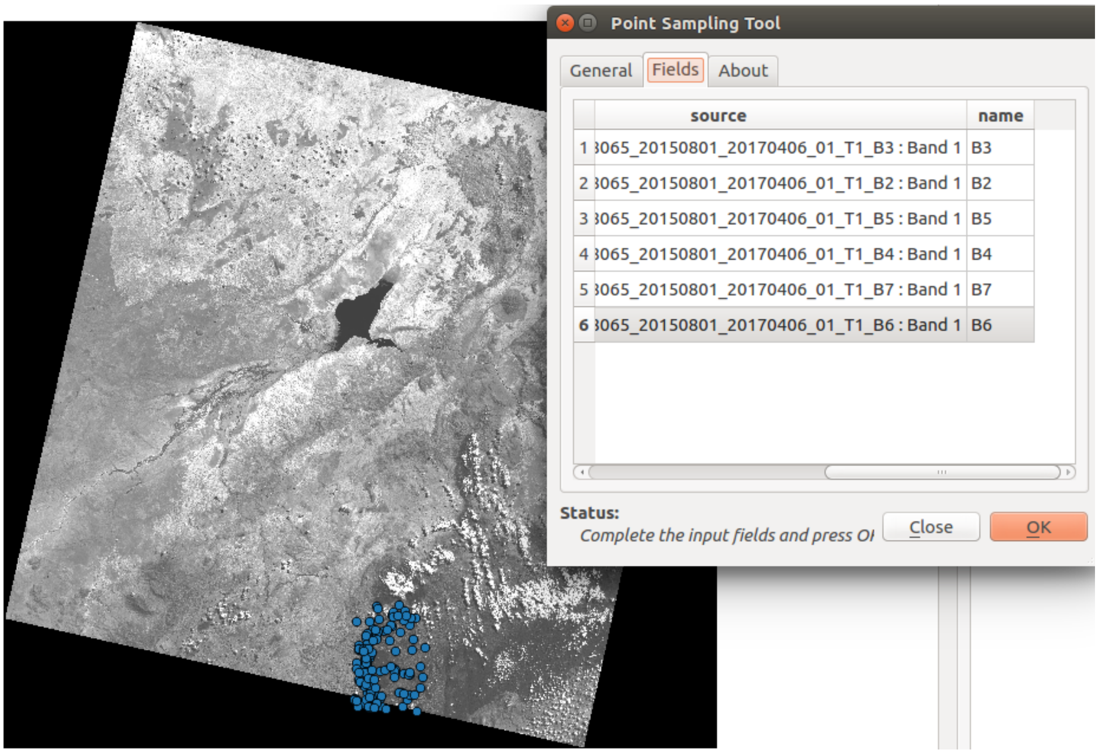


Figure 5. Extracting Band Values.

3.2 Methods

After obtaining the data, several steps were carried out in trying to predict the age of plantations. There were three age classes obtained from PFP [17] that were used in the study. We had about $0-3$ years for the young, $3-8$ years which represented the growing plantations and finally those that were above 8 years which represented the mature plantations. These were collected from the field measurements. The first step was to scale the age to make it to be continous as it has been shown by equation 1 .

- **Scaling the Age**

$$Age = Median\ \ of\ \ age\ \ class\ \ + (X - Y), \tag{1}$$

$$\tag{2}$$

where, X $\Longrightarrow$ Year when satellite image was taken.
Y $\Longrightarrow$ Year when imagery was recorded.

The second step was to get the NDVI. It is always computed from reflectance measurements. It always ranges between -1 and $+1$. To get the NDVI, the formula on Equation 3 was applied and results were used to see if there is any correlation with the plantation age.

$$NDVI = \frac{B_5 - B_4}{B_5 + B_4}, \tag{3}$$

where B_5 is band number 5 and B_4 is the band number 4.

3.2.1 Sparse Bayesian Technique

Sparse Bayesian model considered for the study is based on the model by [5]. We first consider a simple regression model given by Equation 4 whereby we need to estimate the age of the plantation using the sparse Bayesian method based on satellite image data. The simple linear regression is given as,

$$y = f(X, \theta) + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2) \tag{4}$$

where y are the obtained measurements, $f(X, \theta)$ is the model with the design matrix for variables X and θ represents the unknown parameters, and ε is the measurement error.

We need to find θ that gives a good functional mapping between the model $f(X, \theta)$ and the measurements, y takes a function $f(X, \theta)$ that give a good prediction for new data. To estimate θ,we compute an LSQ estimate for the parameters as shown by Equation 5 defined as,

$$l(\theta) = \sum_{i=1}^{n} \left[y_i - f(X_i, \theta)\right]^2. \tag{5}$$

The algebraic expression of the least square estimator of θ is computed in matrix form as,

$$\hat{\theta} = \left(X'X\right)^{-1} X'y.$$

The estimation of parameters under Bayesian regression model is based on the same assumption as of the regression model whereby measurement errors are minimized by calculating the difference between the model and the responses [5]. In the Bayesian setting, the parameter estimated θ is taken as a random variable and the objective is to get the posterior distribution $\pi(\theta|y)$ for those parameters. By applying the Bayes' formula we can write our posterior density as;

$$\pi(\theta|y) = \frac{l(y|\theta)p(\theta)}{\displaystyle\int l(y|\theta)p(\theta)d\theta}, \tag{6}$$

where $l(y|\theta)$ has the measurement error model and it provides us with the probability density of observing the measurements y given θ. $\displaystyle\int l(y|\theta)p(\theta)d\theta$ is our normalizing function and $l(y|\theta)$ is our posterior distribution.

Based on the objective of our study, we want to predict the age of a plantation from satellite image data using the sparse Bayesian method. This is done such that, the linear regression model of the targeted vector (plantation age) with a specified matrix of the proposal variables is formulated such that,

$$y = \theta X + \varepsilon, \tag{7}$$

for the response y_P on every plot $P = 1, \cdots\cdots\cdots, P$ expressed as

$$y = [y_1 \ldots\ldots\ldots\ldots, y_P]^T .$$

Also,

$$\theta = [1_{P,1}, \theta_1, \cdots\cdots\cdots\cdots, \theta_Q]^T,$$

stands for the parameters which measure the specific weight of the model. Moreover,

$$X = [1X_{i1} \cdots\cdots\cdots\cdots, X_{iQ}]^T$$

is a $P \times Q + 1$ design matrix of proposal variables . The measurement error ε is also defined as;

$$\varepsilon = [\varepsilon_1 \ldots\ldots\ldots\ldots, \varepsilon_P]^T .$$

From equation 7, we can write it in the likelihood function given as;

$$p(y|X,\sigma^2) = \Pi_{P=1}^{P} N\left(X\theta,\sigma^2\right) \tag{8}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{P}{2}}} \exp -\frac{1}{2\sigma^2}\left(||y - X\theta||^2\right), \tag{9}$$

where $X\theta$ is the mean of the likelihood estimate and σ^2 being the unknown variance. The error term ε in equation 7 is also normally distributed with mean 0 and variance σ^2. We obtain a good prediction by minimizing the error in the regression model but in the Bayesian setting we maximize the likelihood $l(y|\theta)$ in equation 6. When estimating many parameters, we opt for Bayesian model with a priori distribution defined as;

$$p(\theta|\alpha) = \Pi_{q=0}^{Q} N(0,\alpha_q^{-1}), \tag{10}$$

where apriori mean in each of the weights is given as 0 and the variance defined by hyperparameter $\alpha = [\alpha_0, \cdots, \alpha_Q]^T$.

Bayesian inference: Bayesian inference is always done by getting the posterior distribution given the data. This can be simply done by applying the Baye's rule as shown:

$$p(\theta|y,\alpha,\sigma^2) = \frac{p(y|\theta,\sigma^2)p(\theta|\alpha)}{p(y|\alpha,\sigma^2)}, \tag{11}$$

where $p(y|\alpha,\sigma^2)$ is the normalizing constant $\displaystyle\int p(y|\theta,\sigma^2)p(\theta|\theta)d\theta$ and also, $p(\theta|y,\theta,\sigma^2) \sim N(\theta|\mu,\Sigma)$. The mean and covariance are then computed and defined as;

$$\Sigma^{-1} = A + \sigma^{-2}X^T X, \tag{12}$$

and

$$\mu = \sigma^{-2}\Sigma X^T y. \tag{13}$$

4 RESULTS AND DISCUSSION

The results obtained were based on the data which was used. Results were first tested using Band values as our covariate and plantation Age as our target variable and presented in figure. By using data from equation 1 and 3, the results were as shown in figure 6.

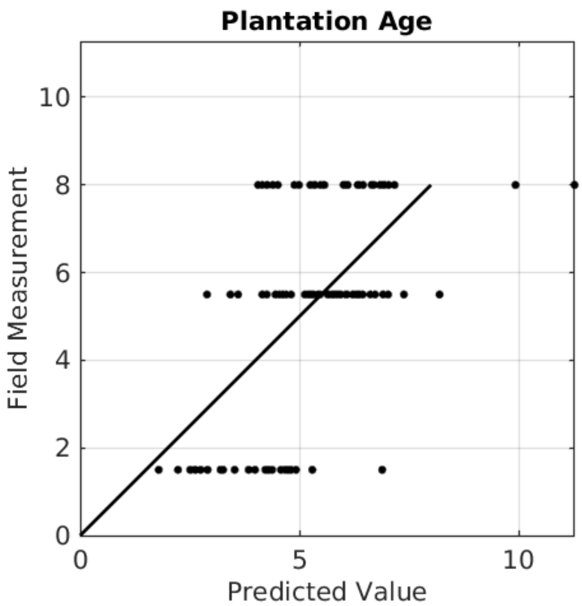
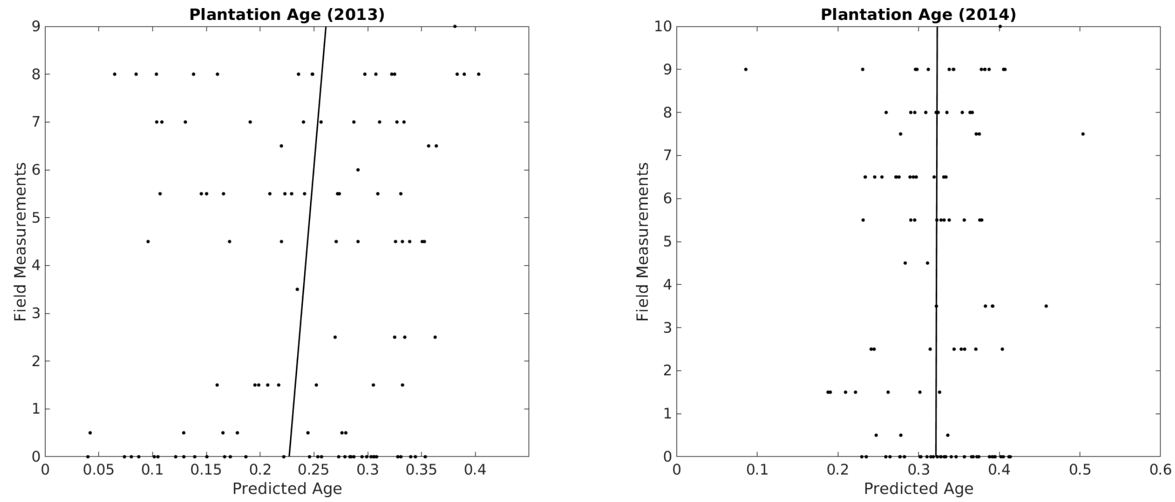


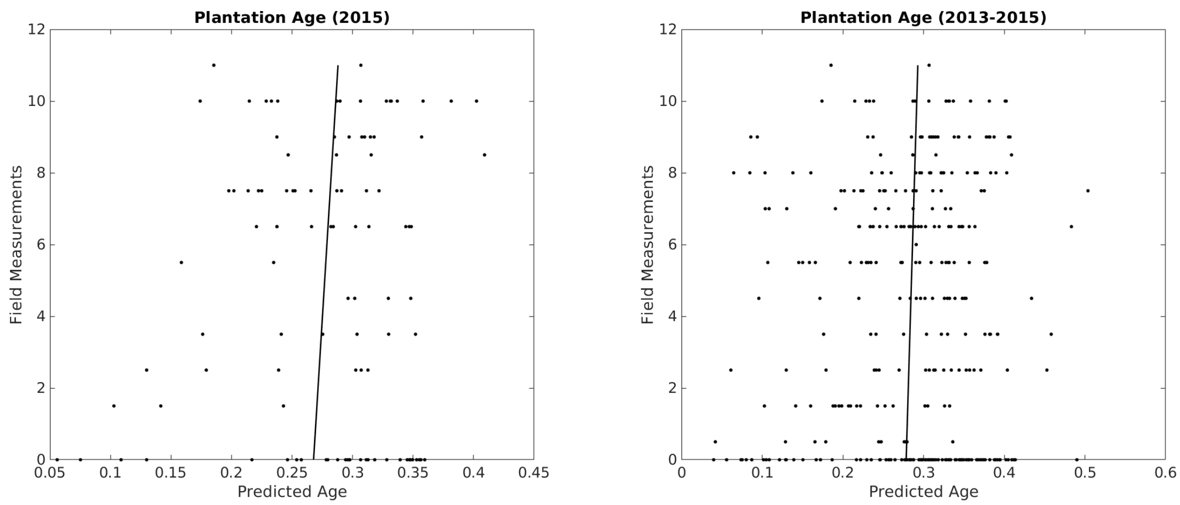
Figure 6. Predicted Age using scaled age and Band Values.

The results do not show if there exists any correlation between the plantation age and the band values. It is not easy to draw conlusions about the relation between the two quantities. What we did was to use the NDVI as our covariate to test if it can show any correlation with the plantation age for each pilot area. First we had to test for each individual year and there after combine them all for pilot area 1.



(a) (b)

Figure 7. Predicted Age for pilot Area 1: (a) 2013; (b) 2014;



(a) (b)

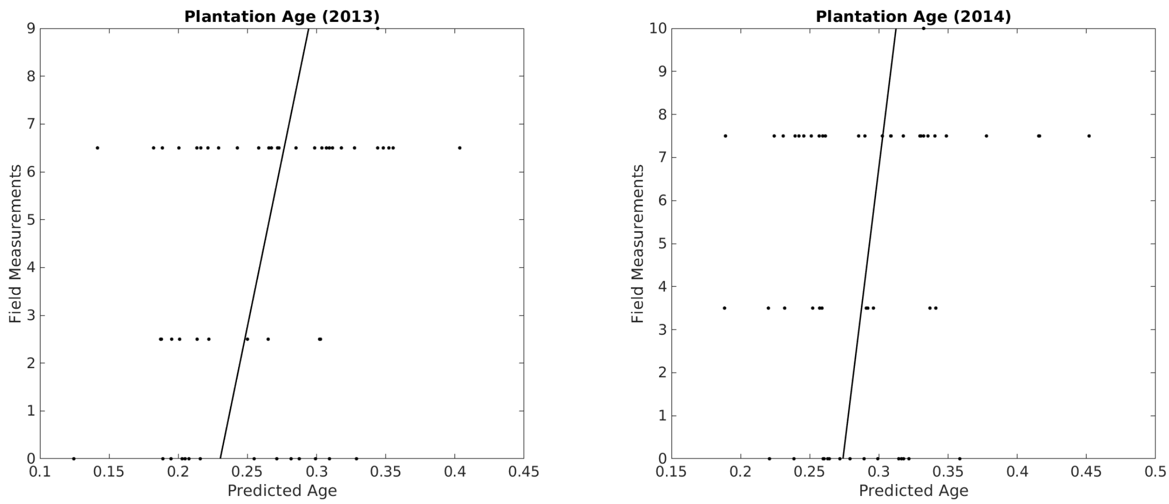
Figure 8. Predicted age for pilot area 1: (a) 2015; (b) 2013-2015;

Figures 7a, 7b and 8a give a summary of the results for each year. We can observe that the correlation between the age and the NDVI is present but very small in figure 7a and the calculated R^2 was 0.017.

Figure 7b showed that there was very little any correlation between the plantation age and the NDVI and the R^2 was 0 in this case.

For the year 2015, there were similar results that were observed which were close to that of the year 2013 and the calculated R^2 was 0.01068 which is close to that of 2013.

We also needed to see how the results would be when we combine all the years . This was clearly shown in figure 8b . We can also observe that there is a slight correlation between the plantation age and the NDVI for all the years. The R^2 that was calculated was 0.003.

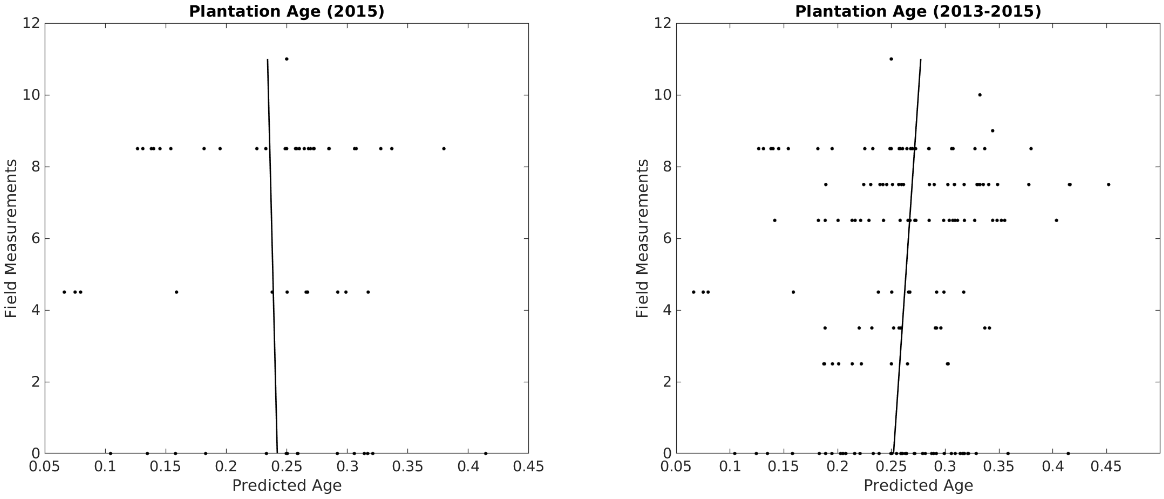


(a) (b)

Figure 9. Predicted age for pilot area 2: (a) 2013; (b) 2014;

Figure 9 shows the predicted age at each plantation plot and the correlation with the NDVI for each year for pilot area 2. We can observe that in figure 9a and 9b there is slight positive correlation for the prediction of the age at each plantation plot with the NDVI. The observations in this pilot area were small and can clearly be seen in the figures. R^2 in 2013 is 0.1200 while in 2014 is 0.0542.

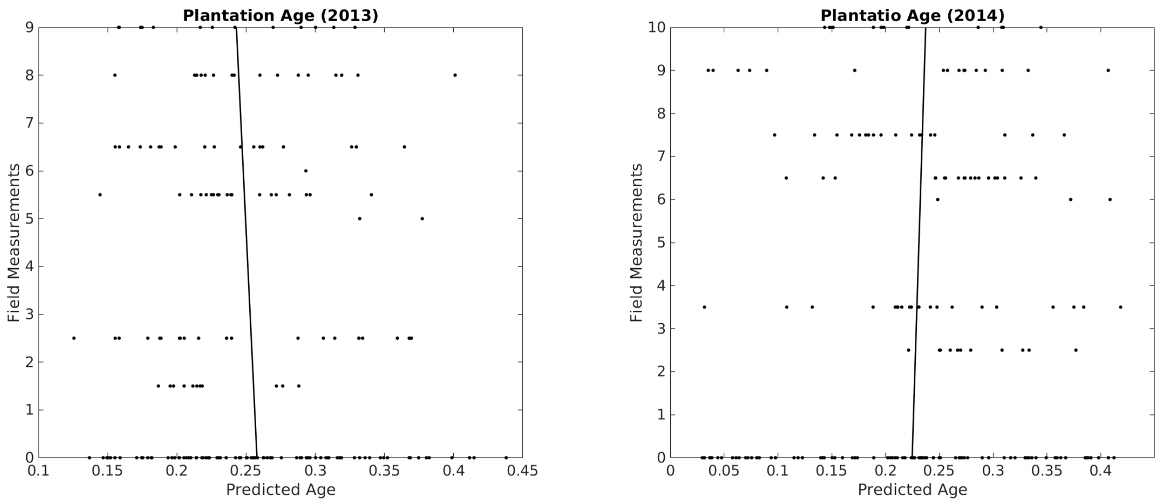
During the year 2015 as seen from Figure 10a, the correlation between the plantation age and NDVI was slightly negative and the calculated R^2 was 0.0013.



(a) (b)

Figure 10. Predicted age for pilot area 2.: (a) 2015; (b) 2013-2015;

Analysis was also done in pilot area 2 for all the years and the results are represented in Figure 10b. It can also be observed that the correlation between the plantation age and NDVI is positive but also very small. Calculated R^2 is 0.0132.



(a) (b)

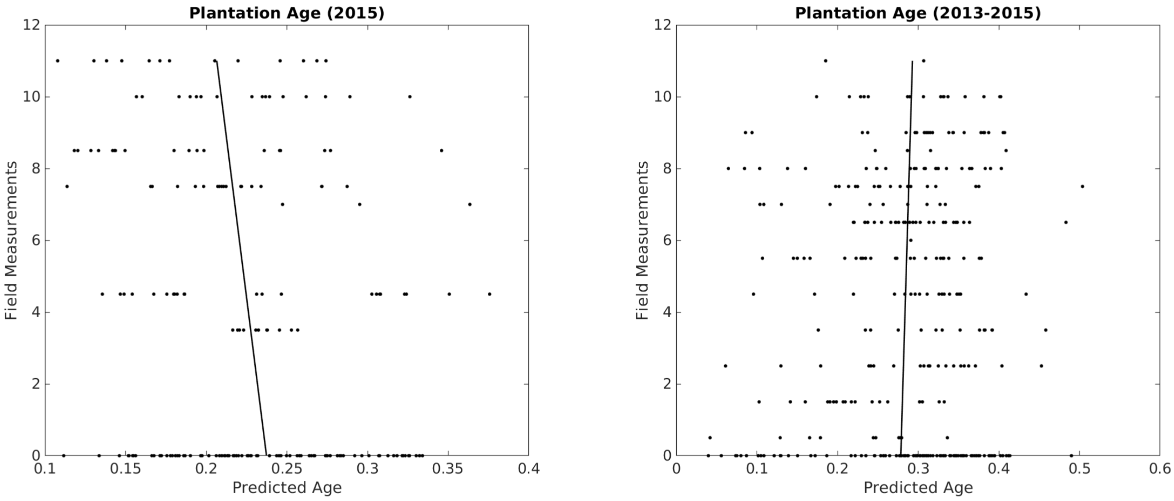
Figure 11. Predicted Age for pilot Area 3: (a) 2013; (b) 2014;

Figures 11a, 11b and 12a shows the results for each year in pilot area 3 that were observed.

Figure 11a shows the correlation between the plantation age and the NDVI for the year 2013. There is a slight negative correlation and our calculated R^2 is 0.0068.

Figure 11b shows the correlation between the plantation age and the NDVI for that year which is slightly positive and calculated R^2 is 0.0022.

There is a slight negative correlation between the plantation age and the NDVI for the year 2015. This is clearly shown in Figure 12a. Calculated R^2 is 0.04.



(a) (b)

Figure 12. Predicted age for pilot area 3.: (a) 2015; (b) 2013-2015;

Figure 12b shows the correlation between the plantation age and the NDVI when all the years are combined that is from 2013 to 2015 in pilot area 3. There is a slight negative correlation and our calculated R^2 is 0.0054.

4.1 Discussion

From the results obtained, in summary, the method seems not to predict well the forest plantation age using the satellite data and the ground data from the three piloted areas. The results showed a very small correlation between the satellite band values and the plantation age. This was also seen from the goodness of fit of the model which was not good. The reason behind this may also be that the quantities measured from the field in the regions may not have been accurate. We may not rely very much on them when it comes to getting information regarding the forest plantation. The measurements may have been incorrect in some areas or contain errors which led to the analysis not being good and it may lead to poor understanding of the area and applicability of the results may not match the requirements.

We might also have had problems with the plantation mask. It may have contained wrong information concerning the plantations because some of the natural forests might have been interpreted as plantations while some plantations might have been left out yet they could have been included. Also, when we were trying to match the ground data collected from the three regions, which were in sample points, some of them that fell partly on plantation and partly outside and might have been ignored and some might have been included when selecting. This also creates a problem because plantations were in polygon forms while sample points were given in point form.

4.2 Future Work

The time for the project was short and therefore there were little approaches for trying out other other methods to obtain better results. We currently do not know if the use of the data with the Sparse Bayesian method was a problem and therefore this calls for trials with other data together with the satellite data and see if we can obtain good results for the prediction of age.

There is a need of predicting the age of the plantation because this is one of the factors that is quite useful for the economic development of a country. This is good especially for timber production and other products obtained from forest plantations. We will still carry out age prediction using the same method but trying with a different source of data with satellite data and see if it can work.

5 CONCLUSIONS

The present study focuses on using GIS and the Sparse Bayesian method to study the age of forest plantation in relation to satellite data. This study has focused on using Landsat $8\ ETM+$ imageries for three years $2013 - 2015$. The images were collected based on the study areas (Kilolo, Njome and Iringa) from USGS earth explorer `https://earthexplorer.usgs.gov`. With this study, various methods were tried to see if we could predict forest plantation age. The first step was to use the band values before normalizing and after normalizing which did not work.

The Sparse Bayesian method has also been described and explained how it works when analyzing forest plantation age and the results were similar even when NDVI was used as an explanatory variable. From the results, we can clearly see that there is very little correlation between the forest plantation age and the satellite data in all the areas. This might have been caused by inaccurate measurements of the data from the field or the plantation mask might have produced unfavourable results.

The Sparse Bayesian method as presented in this study is a typical example technique that can be used to study various characteristics of forest parameters. This method can help us in understanding the nature of forest plantations, which can help in estimating timber volume by studying its age.

REFERENCES

[1] FAO. Global Forest Resources Assessment. Technical report, 2010.

[2] Krishna P Acharya. Twenty-four years of community forestry in nepal. *International Forestry Review*, 4(2):149–156, 2002.

[3] Hesbon Ochego. Application of remote sensing in deforestation monitoring: a case study of the aberdares (kenya). In *2nd FIG Regional Conference Marrakech, Morocco, December*, 2003.

[4] YM Ngaga. Forest plantations and woodlots in tanzania. In *African Forest Forum. Nairobi*, 2011.

[5] Virpi Junttila. *Automated, adaptive methods for forest inventory*. PhD thesis, 2011.

[6] Chintamani Kandel. *Forest cover monitoring in the Bara district (Nepal) with remote sensing and geographic information systems*. PhD thesis, 2009.

[7] Armando A Apan. *GIS applications in tropical forestry*. University of Southern Queenland, 1999.

[8] Jean-Paul Donnay, Mike J Barnsley, and Paul A Longley. *Remote sensing and urban analysis: GISDATA 9*. CRC Press, 2003.

[9] Poonam Singh Maurya, Dharmendra Singh, and Reshu Agrwal. Deforestation monitoring of forest and its predictive modeling: case study of bilaspur forest area, himachal pradesh. *International Journal of Advanced Scientific and Technical Research*, 2(3):288–299, 2013.

[10] Niina Käyhkö, Nora Fagerholm, Bakari S Asseid, and Abbas J Mzee. Dynamic land use and land cover changes and their effect on forest resources in a coastal village of matemwe, zanzibar, tanzania. *Land Use Policy*, 28(1):26–37, 2011.

[11] Asim Banskota, Nilam Kayastha, Michael J Falkowski, Michael A Wulder, Robert E Froese, Joanne C White, Asim Banskota, Michael J Falkowski, Joanne C White, and Robert E Froese. Forest monitoring using landsat time-series data-a.

[12] Warren B Cohen and Samuel N Goward. Landsat's role in ecological applications of remote sensing. *AIBS Bulletin*, 54(6):535–545, 2004.

[13] Warren B Cohen, Thomas A Spies, and Maria Fiorella. Estimating the age and structure of forests in a multi-ownership landscape of western oregon, usa. *International Journal of Remote Sensing*, 16(4):721–746, 1995.

[14] EH Helmer, Thomas S Ruzycki, Joseph M Wunderle Jr, Shannon Vogesser, Bonnie Ruefenacht, Charles Kwit, Thomas J Brandeis, and David N Ewert. Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared landsat and ali image mosaics to characterize avian habitat. *Remote Sensing of Environment*, 114(11):2457–2473, 2010.

[15] Demetrius Kweka, Rachel Carmenta, Maija Hyle, Irmeli Mustalahti, Therese Dokken, and Maria Brockhaus. *The context of REDD+ in Tanzania: Drivers, agents and institutions*, volume 133. CIFOR, 2015.

[16] Rasmus Hundsbæk Pedersen. *The Political Economy of Private Forestry in Tanzania: A Review*. DIIS, 2017.

[17] Private Forest Programme. *Forest Plantation Mapping of the Southern Highlands.* Final report. Iringa, Tanzania., 2017.

[18] Yassin Mbululo and Fatuma Nyihirani. Climate characteristics over southern highlands tanzania. 2012.

[19] Private Forest Programme. *desk study for developing mechanisms and policies that strengthen the private plantation forestry and related value chains.* Final report. Tanzania., 2014.