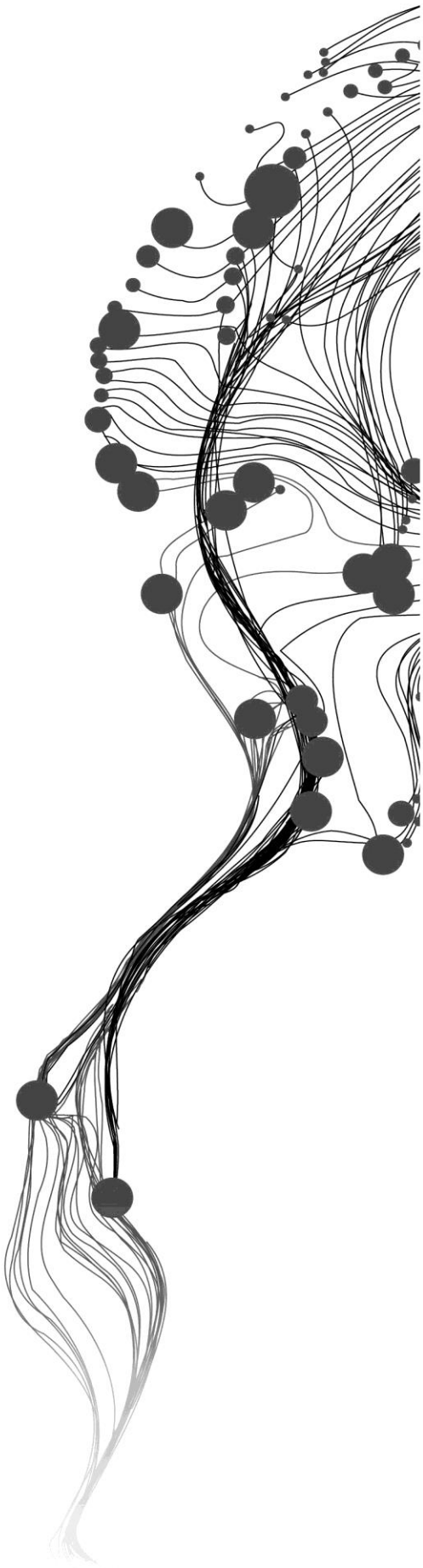


LiDAR REMOTE SENSING OF ABOVEGROUND BIOMASS USING A PLOT-BASED APPROACH IN THE TROPICAL FOREST OF NEPAL: A COMPARISON OF REGRESSION AND GEO-STATISTICAL APPROACH

OLUSOLA ADEFURIN

Enschede, the Netherlands, June 2012

Thesis submitted to the faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-Information Science and Earth Observation.

Specialisation: Natural Resources Management

SUPERVISORS:
Dr. Iris van Duren
Dr. David Rossiter

THESIS ASSESSMENT BOARD:

Chair: Dr. Y. A. Husssin
External Examiner: Dr. N. Kerle (Faculty of Geo-Information Science and Earth Observation, ITC)

DISCLAIMER

ABSTRACT

Tropical forests contribute significantly to the global carbon balance, attaining high rate of carbon sequestration through their huge densities. In order to develop a better understanding and quantification of carbon stocks and flux dynamics, estimation of aboveground biomass becomes very crucial. Remote sensing technologies have proven superiority over other methods because of their extensive coverage capability making estimation of aboveground biomass possible over broad spatial scales. To achieve broad scale mapping with remote sensing, statistical relationship between sensor-extracted tree parameters and field measurements are used. With the advancement in remote sensing technology, laser scanning evolved with the possibilities of acquiring three dimensional information of the forest structure.

The aim of this study is to demonstrate the use of low density data (less than 1 point/m^2) for the estimation of aboveground biomass in the tropical forest of Nepal (Kayarkhola watershed in Chitwan Province) while comparing between regression and geo-statistical approach (ordinary kriging and regression kriging)..

The plot-based approach was adopted and 81 metrics were extracted using LiDAR's elevation and intensity values including canopy density computations. Intensity metrics were excluded from further regression analysis because of their poor relationship with aboveground biomass. Data reduction technique (PCA) was used to select independent and uncorrelated LiDAR metrics. Seven predictors were selected including height maximum, height average and absolute deviation, height L-moment (L2), Height L-moment skewness, 40^{th}, 80^{th}, and 95^{th} height percentile. For this study, it was possible to employ only the 95^{th} height percentile for predicting aboveground biomass without additional variables. The model of 95^{th} height percentile and aboveground biomass showed a moderately strong relationship with aboveground biomass, with an R^2 of 0.54.

The performance of the two approaches was assessed using their RMSE values and ME estimates. Using regression kriging, this study showed an improvement in the accuracy prediction of aboveground biomass with a lower RMSE of 0.20 LN(Mg/ha) and ME of 0.00023 LN(Mg/ha). Regression analysis resulted in an RMSE 0.68 LN(Mg/ha) and ME of 0.42 LN(Mg/ha). Regression kriging showed an improvement in the estimation because of its ability to account for some variations in aboveground biomass.

ACKNOWLEDGEMENTS

I would like to cease this opportunity to thank the Almighty God for bringing me thus far and for his faithfulness. My sincere gratitude goes to the Netherland Government and the Netherlands organization for international cooperation in higher education (NUFFIC) for providing the financial support for the MSc programme. I also want to thank Mr. Ogundele for his support towards my MSc study.

I wish to thank my first supervisor, Dr. Iris van Duren, for her comments and feedbacks. Furthermore, I wish to thank my second supervisor, Dr. D. G. Rossiter, for his support and expert contribution to the geostatistical aspect of my thesis. Sincere appreciation goes to the course director, Dr Michael Weir, for his guidance and support throughout the MSc period.

I wish to acknowledge the International Centre for Integrated Mountain Development (ICIMOD), Asia Network for Sustainable Agriculture and Bio-resources (ANSAB), and Federation of Community Forest Users' Nepal (FECOFUN), for providing necessary facilities and manpower that enhanced the fieldwork processes. They provided the platform through which the research was made possible. Special acknowledgement goes to the Forest Research Assessment (FRA) project, Nepal, funded by Finland and Arbonaut Limited, Finland for providing the LiDAR data, the main data on which my research was centred on.

Also, I wish to acknowledge Hammad Gilani of ICIMOD for his active coordination of activities during the course of my fieldwork in Nepal, Basanta Gautam of Arbonaut Limited for sharing knowledge about LiDAR data and Khamarrul Azahari Razak for his helpful assistance in the processing of my LiDAR data. To my NRM colleagues, we shared nice moment together, I appreciate you all.

I am indebted to my family for their invaluable love, encouragement, and support to me throughout my MSc, in particular, Adebowale Adefurin (generous brother), for his priceless involvement during the course of my thesis. I am so blessed to have you all as my family. To my Mum, I want to say, you are the best Mum in the world. Thank you for your encouragement innumerable times. Your words kept me going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ANSAB	Asia Network for Sustainable Agriculture and Bio-resources
ICIMOD	International Centre for Integrated Mountain Development
FECOFUN	Federation of Community Forest Users' Nepal
NORAD	Norwegian Agency for Development Cooperation
CFUGs	Community Forest User Groups
CFs	Community Forests
RADAR	Radio Detection and Ranging
LiDAR	Light Detection and Ranging

.

1. INTRODUCTION

1.1. Background

Forest ecosystem is an important factor in maintaining carbon balance in the atmosphere. According to FAO, (2010) forests cover about 31% of the total land area of the world. Among the terrestrial ecosystems, forests accumulate a larger portion of anthropogenic carbon generated mainly through fossil fuel burns (Canadell & Raupach, 2008). Naturally, forest ecosystems absorb and store atmospheric carbon dioxide (CO_2) over long periods of time. By virtue of this function, forests regulate the earth's changing climate by sucking up excess CO_2 in the atmosphere. The rate by which a tree stores carbon can be attributed to the rate of by which carbon dioxide (CO_2) is utilised by the absorptive structure of the tree's woody tissues, which is required and used in photosynthetic processes, tree growth and subsequently forest productivity. Growing trees transform atmospheric carbon into woody biomass through photosynthesis. By this process, CO_2 is captured and removed from the atmosphere (CO_2 sink). By estimation, for every ton of carbon stored in a tree 3.67 tons of CO_2 is removed from the atmosphere (Hunt, 2009). Therefore, based on its natural function, the world's forest base has captured and stored more atmospheric CO_2 compared to the amount in the atmosphere (Hunt, 2009). As a result, forests constitute an important part of the terrestrial ecosystem by virtue of the role it plays in mopping up the world's atmospheric CO_2.

Despite the importance of the forest, acting as a carbon sink and mopping up CO_2 concentration in the atmosphere, it can also be detrimental to the carbon cycle phenomena in the atmosphere. Natural and man-made activities carried out on the forest resources such as deforestation, forest degradation, forest fires and pest and disease outbreaks puts immense pressure on the world's forest. This culminates into the simultaneous release of carbon stocks into the atmosphere that have been accumulated over long periods (CO_2 source). The Fourth Assessment Report of IPCC indicates that "the forestry sector, mainly through deforestation, accounts for about 17% of the global greenhouse gas emissions making it the second major source of greenhouse gases after the energy sector" (IPCC, 2007).

Invariably, forests absorb CO_2 naturally through the process of photosynthesis but exudes CO_2 through indiscriminate and destructive man-made activities, regenerating CO_2 back into the atmosphere (Brown, 2002). Sustainable management and appropriate maintenance of existing

forests, as well as increasing forest area through afforestation programs, could make a substantial impact on mitigating the increasingly high rise of atmospheric Co_2.

The recognition of the dual role of the forest as a carbon source and sink has attracted the attention of international bodies enabling them rise to the challenges of tackling global climate change. This led to the conception of international policies and agreement such as the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto protocol. UNFCCC intends to capitalize on this climate protection role of forests through the introduction of a financial and economic compensation mechanism. The mechanism adopted by the UNFCCC is called 'REDD' (Reducing Emission from Deforestation and Degradation). UN-REDD is a financial mechanism that offers financial incentives (carbon credit) to forest-rich developing countries encouraging them to reduce national deforestation rates and associated carbon emissions to below a set baseline (Gibbs et al., 2007). Conversely, the developed countries compensate for their CO_2 emission by purchasing carbon credits from forest-rich developing countries (signatories) that are interested and committed to safeguarding their forest endowment from destruction (Angelsen et al., 2009).

Effective REDD implementation will require accurate biomass and carbon estimation and the continuous monitoring of the forest carbon pool (Englhart et al., 2011). To achieve these requirements, robust methods for quantifying forest carbon storage over time are essential (Gonzalez et al., 2010). This would enable signatory countries demonstrate whether or not they are able to maintain or increase their forest carbon stock and reduce emissions from deforestation, and as a result qualify for carbon credits.

The target countries for the REDD scheme are the forest-rich developing countries. Most of these forest-rich countries are located in the tropics and with good potentials to sequester high levels of carbon due to the high forest density. Tropical forests represent a very significant portion of the global forest covers and also constitute major threat through high levels of deforestation and forest degradation (Brown, 1993) leading to increased carbon emission levels. The high rate of deforestation is attributable to population growth and high poverty levels in these developing countries (Hunt, 2009). This consequently leads to high dependence on forest resources and quest for additional space for agricultural production to meet the food supply and economic demands of the growing populace.

Nepal, a developing country, is blessed with vast expanse of forest land area. In Nepal, the forest area covers about 40% of its total land mass. This qualifies Nepal as a prospective country for mitigating the adverse impact of climate change (Oli & Shrestha, 2009). However, the high

dependence on forest resources by the local communities pose a threat to the existing forest resources which is gradually been depleted through deforestation and forest degradation. In order to alleviate the situation, forest management was handed over to local communities since the 1970's (Dhital, 2009). Currently, about 25% of the total forest area are been managed under the Community Forest User Groups (CFUG) (Dhital, 2009). Review of studies show that through the community management approach, Nepal has witnessed both a reduction in deforestation rate and higher carbon sequestration rate (Dhital, 2009; Oli & Shrestha, 2009).

The benefits of the community forestry approach includes restoration of degraded sites, biodiversity conservation, sustained supply of forest products to rural people and most importantly, promoting local level institutions for resource management (Dhital, 2009; Oli & Shrestha, 2009). Therefore, community forest management is considered as one of the strategies for controlling emission rates under the REDD policy (Dhital, 2009; Oli & Shrestha, 2009).

Nepal is currently a signatory to the UN-REDD policy scheme and benefits from the Forest Carbon Partnership Facility for REDD project implementation. To sustain the UN-REDD scheme, it is necessary to baseline the amount forest carbon through aboveground biomass estimation in order to ascertain a country's performance. There are on-going REDD pilot programme in selected CFUGs to baseline the amount forest carbon in Nepal and this study is undertaken as part of the programme. In addition, this study will assess the use of plot-based approach and LiDAR metrics in estimating aboveground biomass.

1.2. Aboveground biomass estimation

To understand and quantify carbon stocks and fluxes, continuous assessment, monitoring, and management of aboveground biomass are very essential to quantifying biomass changes. The biomass of a tree, as defined by Walker et al., (2011), "is the weight of its living plant tissue". It includes both the section of the tree above the ground (leaves, branches, and stems) commonly referred to as 'Aboveground biomass' and below the ground (roots) referred to as 'belowground biomass'. Aboveground biomass is considered the most important component of trees because it is the part mostly affected by forest deforestation and degradation (Brown, 1997). Aboveground biomass cannot be measured directly from the field but modelled from other parameters that are measured directly from the field, for example diameter at breast height (DBH) and tree height.

1.3. Methods for estimating above-ground biomass

The techniques used to estimate aboveground biomass (AGB) have improved over the years, especially with the advent and evolution of remote sensing data. According to the review studies

carried out by Lu, (2006) and Gibbs et al, (2007), the approaches for estimating aboveground biomass can be summarized into three namely; (I) field measurement based, (II) GIS based and (III) remote sensing based methods.

The field based method formed the basis of aboveground estimation. The most straightforward approach is by harvesting all trees in a certain area and then drying and weighting the woody biomass (Gibbs et al., 2007). Though this method seems accurate, it is very destructive, time consuming and expensive. The approach is not realistic and practicable for regional and large scale biomass estimation (Gibbs et al., 2007; Patenaude et al., 2004). This direct and destructive approach has been further improved upon to extrapolate the estimation to a much larger area by building allometric models.

Allometric models built from harvested trees are related with forest inventory data of tree parameters to produce national level biomass estimates. Although several studies have been done using national forest inventory data, for example, Brown et al., (1989), national forest inventory data is very expensive to collect and maintain on a continuous bases (Gibbs et al., 2007). Inventory based method have the potential to generate biomass estimate with low uncertainty. Most times, sampled inventory data are inadequate for the national scale in view and varying sampling design are mostly employed due to the inconsistent terrain of the region (Brown, 1997; Brown & Gaston, 1995). Another possible, but seldom used, ground measured field data technique is the use of geostatistical techniques to generate the spatial distribution of biomass and carbon. Sales et al., (2007) demonstrated the use of RADAMBRASIL forest inventory data and interpolation with geostatistics in mapping forest biomass in Randonia, Brazil.

GIS based method involves the use of ancillary data such as elevation, precipitation etc. The limitation of this approach include lack of quality ancillary data, static nature of the ancillary data, difficulty with acquiring up-to-date ancillary data and the indirect relationship with aboveground biomass (Lu, 2006). Certain geostatistical techniques such as Kriging with an External Drift (KED) can also be used with ancillary data for biomass estimation (2007).

Remote sensing approach involves employing statistical relationship between satellite and airborne-extracted tree parameters and vegetative indices (with a correlation with aboveground biomass), and field based measurement. Field measurements are converted into aboveground biomass with the use of an allometric equation.

Based on either of these approaches for estimating aboveground biomass, it is obvious that forest biomass cannot be estimated directly (Goetz & Dubayah, 2011; Goetz et al., 2009; Rosenqvist et al., 2003). Remote sensing based approach requires supplementary ground based data for a two-fold analysis with the purpose of calibrating and validating spatial estimates of forest biomass predicted from sensor derivatives (Goetz & Dubayah, 2011). To generate spatial estimates of forest biomass for a large region, remotely-sensed data can be associated with field based aboveground biomass through regression model or geostatistical model.

Remote sensing techniques have proven superiority over the rest of the methods because it is cheaper, covers a larger area, and easier to repeat data collection (Gibbs et al., 2007; Goetz & Dubayah, 2011; Goetz et al., 2009; Lu, 2006; Patenaude et al., 2005). Due to its extensive coverage capability, it can be used to map the spatial heterogeneity of the ecosystem over local, national or global scale over time. The use of remote sensing is very crucial for monitoring carbon stocks and fluxes and the outcome, appropriate within the framework of international policies such as UNFCCC, will influence policies designed to reduce greenhouse gases and aid the implementation of the REDD program (Clark et al., 2011).

Remote sensing (RS) technologies using satellites have advanced with the development of new sensors and high resolution optical images (Goetz & Dubayah, 2011). High resolution optical data have been applied in forestry studies such as mapping vegetation types, land cover classification and mapping aboveground biomass (Asner et al., 2010; Fuchs et al., 2009). Optical data from satellite-based remote sensing technology is adequate in estimating aboveground biomass in homogenous and less dense forest but not appropriate for forest with dense canopy (Holmgren & Thuresson, 1998). However, It suffers from a number of limitations such as the challenges of acquiring high quality data due to cloud cover interfering with satellite signals (Gregory, 2009), most especially in the tropical forest region. Furthermore, there is also the problem of low saturation level of the spectral bands, and the derived spectral indices related to biomass estimation (Clark et al., 2011; Gibbs et al., 2007). The saturation level is the threshold at which optical sensor can estimate biomass levels irrespective of how dense the vegetation in the area. Low saturation level could lead to inaccurate and under-estimation of biomass in a dense tropical forest. Since most tropical forests are dense and store a large amount of carbon, the problem related to the saturation of these optical satellites leads to an unreliable estimate of carbon stock.

More recently, active airborne sensors (e.g. LiDAR and RADAR technologies) have emerged making it possible to acquire remotely sensed imageries with higher resolution (Clark et al., 2011).

These technologies apply the same techniques in acquiring remotely sensed data. While RADAR employs microwave energy, LiDAR employs light energy to capture data from the earth surface (Clark et al., 2011). The advantage of RADAR sensors over satellite-based optical remote sensors lies in the RADAR signal's ability to penetrate through the clouds. However, both RADAR and optical remote sensors suffer from a common setback of decreasing sensitivity with increasing canopy structural heterogeneity and age (Gibbs et al., 2007; Patenaude et al., 2004). Again, this is a severe limitation in tropical forests with increased aboveground biomass.

Remote sensing employing LiDAR (with laser transmitting sensors) has been proven to be an effective method for mapping forest biomass (Clark et al., 2011). The use of this sensing technology comes at a significant cost in comparison to the optical remote sensing technology but it brings with it a more reliable estimation of biomass and the capability to carry out extended spatial analysis in three dimensions (3D).

1.4. LiDAR remote sensing for forestry studies

LiDAR remote sensing has the capability to measure the three-dimensional vertical structure of the vegetation. LiDAR provides the location (XY) and elevation (Z) data of features on the ground. This capability gives LiDAR a cutting edge advantage over other remote sensing technologies such as the high resolution satellite imageries. LiDAR has a very high use potential in forest resource management because it can be used for the measurement and estimation of several forest characteristics. The forest characteristics and how they are derived from LiDAR are summarised in Table 1.1 (Dubayah & Drake 2000), signifying why LiDAR remote sensing is an important tool for forest studies.

.

Table 1-1: Forest features retrieved from LiDAR data (Dubuyah & Drake, 2000)

Forest Characteristics	**LiDAR Derivation**
Canopy Height	Direct retrieval
Sub-canopy topography	Direct retrieval
Vertical distribution of intercepted surfaces	Direct retrieval
Above-ground biomass	Modelled
Basal area	Modelled
Mean stem diameter	Modelled
Vertical foliar profiles	Modelled
Canopy volume	Modelled
Large tree density	Inferred
Canopy cover, leaf area index	Fusion with other sensors
Life form diversity	Fusion with other sensors

LiDAR has been proven to be effective for mapping forest biophysical properties (Drake et al., 2003; Lim et al., 2003). The application of LiDAR with satellite monitoring technology can provide baseline carbon estimates and improve the monitoring of carbon stocks, losses and forest recovery (Goetz & Dubayah, 2011).

1.5. Extraction Information from LiDAR data

Two approaches are generally used in the extraction of information from LiDAR data (Evans et al., 2009; Hyyppa et al., 2008). This includes (I) the individual tree-based approach and (II) the plot-based approach.

The individual tree based approach involves the identification of individual trees from LiDAR followed by the extraction of individual tree parameter such crown projection area and height, and relating these parameters to the forest parameter of interest for example, aboveground biomass. The plot-based approach relies on the extraction of statistical variables on a plot-by-plot base from the LiDAR data, known as LiDAR metrics, and relating them to the forest parameter in consideration. Both approaches have constantly been used for forest parameter estimation. The use of low density data have been found unsuitable for the identification and extraction of individual tree features (Lim et al., 2003).

1.6. Definition of concepts

1.6.1. Allometric equation

Aboveground biomass is derived from field data using allometric equation. An allometric equation is a regression equation that models biophysical parameters, which cannot be derived directly from variables measured from the field e.g. DBH (Brown et al., 1989). These equations are developed from field data collected by destructively harvesting sparse numbers of trees, of varying sizes, from a particular location and estimating its biomass content, which is empirically related to its DBH (Brown et al., 1989). The allometric equation differs for different species and geographic location. Various allometric equations have been developed in line with various criteria such as regions (tropical, temperate), general species groups (tropical moist wood, temperate hardwood), and specific species in different regions (*Shorea robusta*, *Pinus species*) (Brown, 1997).

Studies have shown that an allometric equation using only DBH to explain aboveground biomass yield a significant regression results (Brown et al., 1989). Though, inclusion of height in the allometric equation can improve the coefficient of determination (r^2), measuring the height of

trees across a large sample plot can be time consuming, very difficult and, most times, impossible to determine the top of trees (Brown, 2002). Therefore, allometric equation based on DBH alone, and stratified by species groups or climate types are mostly used (Basuki et al., 2009; Brown et al., 1989; Nelson et al., 1999). In this study, a specie-specific and a general allometric equation will be utilized to derive aboveground biomass from *Shorea robusta* and other tree species respectively.

1.6.2. LiDAR metrics

LiDAR metrics are statistical parameters derived from raw LiDAR point clouds. They describe the vertical and canopy structure of the forest statistically. Forest canopy, the portion actively involved in the photosynthetic process and the primary source of forest productivity, is closely related to tree and stand structure (Drake et al., 2003; Li et al., 2008). Studies have shown that a strong connection exist between LiDAR-derived metrics and aboveground biomass (Clark et al., 2011; Hall et al., 2005; Næsset, 2002, 2004). This can be attributed to the relationship between canopy structure and woody biomass (Clark et al., 2011; Drake et al., 2003; Li et al., 2008; Means et al., 1999; Næsset, 2002; Nelson et al., 1988). Some examples of LiDAR metrics include minimum height, maximum height, mean height, height standard deviation, height variance, height percentiles, etc. In this study, a number of LiDAR metrics will be derived from LiDAR point clouds to predict the amount of aboveground biomass in the study area.

1.6.3. Geostatistical techniques

Geostatistics is a branch of statistics specifically focused on the analysis of spatial data i.e. data with a location in space (Webster & Oliver, 2001). Geostatistical techniques take advantage of the spatial autocorrelation among sampled variable points in the prediction of regionalized variables over an area of interest (i.e. forest study area in this case). Geostatistics includes statistical techniques for modelling spatial variability of regionalized variables through prediction and simulation (Journel, 1987). An example of geostatistical prediction and simulation tool is kriging and stochastic simulation respectively. This study will only utilize kriging as a prediction tool for modelling aboveground biomass. Prediction of a target variable in a region, (i.e. aboveground biomass in this study) employs both the measured values of the sampled target variable and a model of its spatial structure. The semi-variogram is a tool used to represent and model spatial variation. This is represented below in equation 1.1.

$$\boldsymbol{\gamma(h) = \frac{1}{2N}\sum_{i=1}^{N}[z(x_i) - z(x_i + h)]^2} \qquad \text{.....Equation (1-1)}$$

Where:

x_i = a data location,

h = a lag vector,

$z(x_i)$ = a data value at location x_i,

N = the number of data pairs spaced at a distance (Webster & Oliver, 2001)

Nowadays, the extent of geostatistical prediction has gone beyond spatial prediction from sample points alone. Several variants of geostatistical prediction, which combines other auxiliary data layer is increasingly been applied such as regression kriging, co-kriging, etc. Ordinary kriging and regression kriging will be applied in this study.

(A). Ordinary kriging

It is an interpolation method that is based on a linear model of measured parameters for the study area. This linear model is a weighing function for kriging and it can be represented by a semi-variogram (Hudak et al., 2002). It differs from other interpolation method because the weights are "based on the distance between the measured points, the prediction location and the overall spatial arrangement among the measured points" (Webster & Oliver, 2001). The sample variogram is used to fit an appropriate model e.g. spherical, exponential model, which is then used for predicting un-sampled areas.

(B). Regression kriging

Regression kriging (RK) is a geostatistical technique that combines two approaches; a simple or multiple regression of the primary variable on the secondary variable(s) with ordinary kriging of the regression residual. Hence, the method addresses both the spatial dependence of observations and the relationship between the dependent variable (aboveground biomass) and the ancillary variable(s) (Webster & Oliver, 2001). It is represented in equation (1-2).

$$\hat{Z}(s_o) = \hat{m}(s_o) + \hat{e}(s_o) = \sum_{k=0}^{p} \hat{\beta}_k \,.\, q_k(s_o) + \sum_{i=1}^{n} \lambda_i \,.\, e(s_i) \qquad \text{.....Equation (1-2)}$$

Where:

$\hat{m}(s_o)$ = the fitted drift

$\hat{e}(s_o)$ = the interpolated residual

$\hat{\beta}_k$ = the estimated drift model coefficients ($\hat{\beta}_0$ is the estimated intercept)

λ_i = the kriging weights determined by the spatial dependence structure of the residual

$e(s_i)$ = the residual at location s_i (Hengl et al., 2007)

1.7. Problem statement

There has been advancement in remote sensing technologies toward forest resources management including estimation of above-ground biomass. This advancement has gone beyond 2-dimensional mapping (optical remote sensing) to the world of 3-dimensional mapping (airborne laser scanning). All of these, with the goal of achieving accurate inventory work geared toward improved forest resources management. Additionally, consciousness is given to accurate assessment of carbon fluxes as this interferes with the implementation of the REDD`s Programme. LiDAR technology, a portrait of airborne laser scanning has demonstrated to be an effective and accurate method for mapping forest biophysical properties including aboveground biomass (Clark et al., 2011). Most of the previous researches that have been carried out using LiDAR data were conducted in temperate or boreal forest (Naesset et al., 2004) with conifers as the predominant trees species. LiDAR studies are very limited in a tropical forest, only very few studies have been done. However, this study will demonstrate the application of LiDAR data, using a plot based approach in a tropical forest.

Regression approaches are widely applied in modeling relationships between remote sensing data (e.g. LiDAR data) of forest attributes and corresponding ground measurements (Arroyo et al., 2010; Clark et al., 2011; Dubayah et al., 2010; Gregory, 2009; Kim et al., 2010). In contrast, geostatistical estimation techniques that relate remote sensing measurements to ground measurement have not been extensively used. To the best of our knowledge, no study has compared between aspatial regression method and spatial geostatistical method in mapping and estimating above-ground biomass / carbon stock in a forested area using LiDAR remote sensing.

1.8. Main research objectives

The main objective of this research is to derive various LiDAR metrics from low density LiDAR data, to estimate and map aboveground biomass using a plot-based approach and LiDAR metrics in the tropical forest of Chitwan, Nepal.

1. To assess the relationship between the LiDAR metrics and field aboveground biomass in the study area.
2. To compare the accuracy of estimation between regression analysis and geostatistical analysis (ordinary kriging and regression kriging) in the estimation of aboveground biomass.
3. To map and estimate aboveground biomass using the best approach (with the highest accuracy estimate).

The specific objectives, research questions and hypothesis are summarized in the table below.

Table 1-2: Specific objectives, research questions and the hypothesis

Specific Objectives	Research Questions	Hypothesis
1	**1:** How strong is the relationship between LiDAR metrics and aboveground biomass in the study area?	**H1**: There is a statistically significant relationship between at least one LiDAR metrics and aboveground biomass in the study area?
2	**2:** Which of the methods (regression analysis, ordinary kriging, and regression kriging) gives an improved accuracy estimate? What are the accuracies of the different methods in estimating aboveground biomass?	**H1**: Remote sensed data integrated with geostatistics (Regression kriging) improves the accuracy estimation of aboveground biomass in comparison with the regression analysis and ordinary kriging
3	**3:** What is the amount of aboveground biomass in the study area and how is it distributed using the best approach?	

2. STUDY AREA

The study was carried out within Kayarkhola watershed, located in the Chitwan district of the Central Development Region (CDR) of Nepal (figure 2-1). Chitwan district, which is popularly referred to as the "Heart of the Jungle" is part of the seventy-five districts of Nepal.

2.1. Characteristics of Study Area

2.1.1. Geographic location and extent

Chitwan district lies in the central lowlands zone of Nepal. The district is situated at a distance of 150km towards the west (260°) of Kathmandu, the capital city. It is boarded at the North by Dhading, Gorkha, and Tanahun districts; at the South by Parsa and India; at the East by Makwanpur; and at the west by Nawalparasi (ICIMOD et al., 2010). Kayarkhola watershed of Chitwan district is located geographically between 27° 40' 07" to 27° 46' 37" latitude (north) and 84° 33' 25" to 84° 41' 48" longitude (East) – see figure 2-1.

2.1.2. Climatic Condition

Chitwan district is characterised by varied climatic conditions due to its elevation dynamics. It exhibit both tropical and sub-tropical climate and receives an average annual rainfall of 1510mm/year (Panta et al., 2008). The average temperature ranges from a minimum of 16 ˚C to 19 ˚C and a maximum of 29 ˚C to 32 ˚C (Panta et al., 2008). This region experience monsoon rainfall around July, with a summer period that is usually hot with heavy rainfall. In the winter season, the region is relatively cold and dry.

2.1.3. Topography

The altitude of Kayarkhola watershed region ranges from 245m to 1944m. The area is characterised by a landscape with steeps and slopes, with majority of the area having slopes greater than 30°. This high altitudinal dynamics is portrayed in figure 2-1 showing the DTM.

2.1.4. Social and Demographic Information

Kayarkhola watershed region has a population of about 22,000 representing 3935 households (ICIMOD et al., 2010). The watershed is occupied by socially and ethnically diverse local communities. The dominant ethnic groups in the watershed area are the Chepang and Tamang groups. These groups practise shifting cultivation, a traditional system of farming that involves moving from one land to the other. The people mainly depend on the forest for their livelihood.

Their activities are believed to cause a high depletion of forest resources in Kayarkhola watershed.

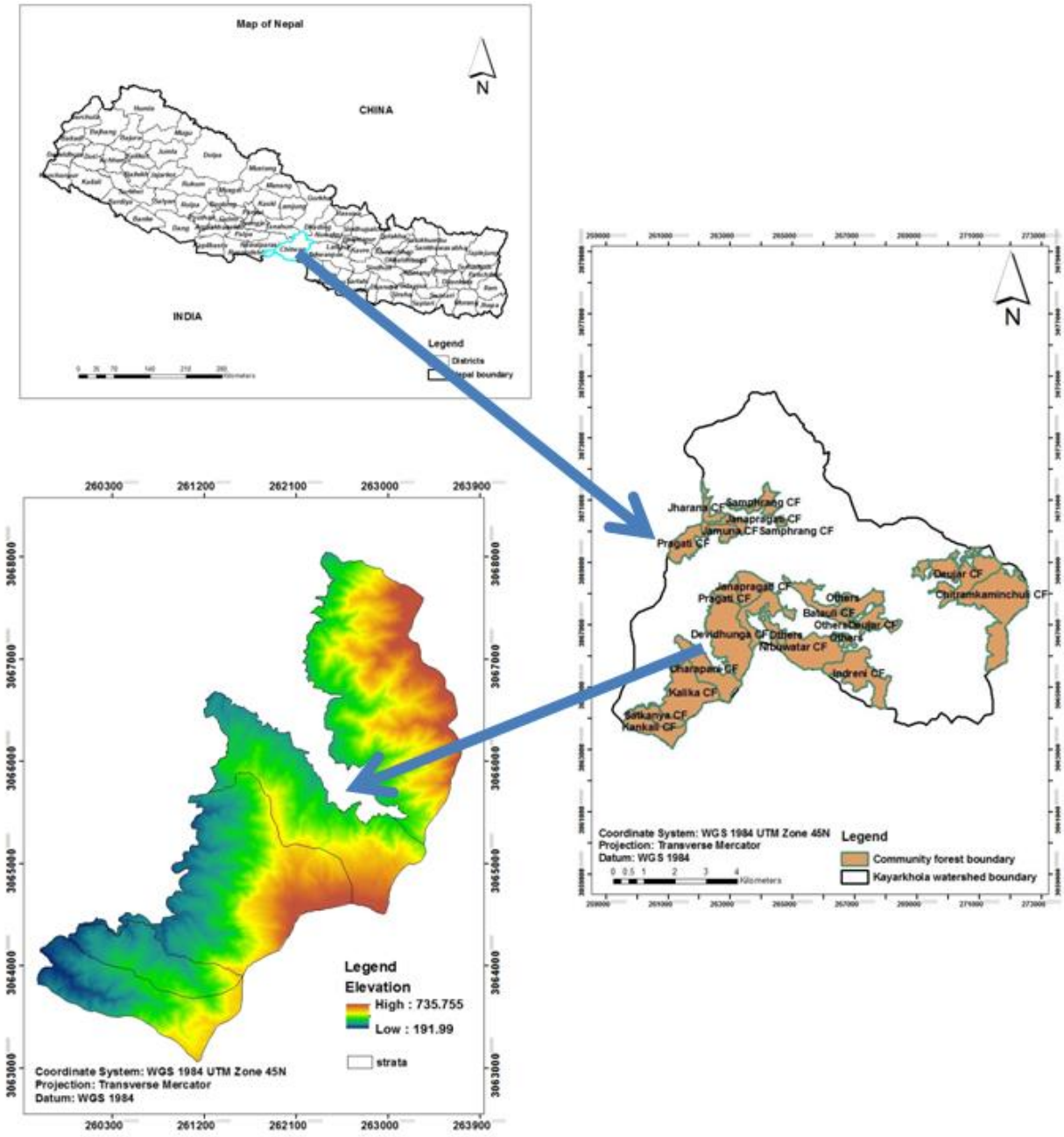


Figure 2-1: Study area showing (a).the map of Nepal (b) Kayerkhola watershed showing the community forests (upper right) (b) 3m resolution DTM of the study area from LiDAR data (lower left)

2.1.5. Vegetation

Forest types in Nepal are diverse due to the differences in climatic and altitudinal characteristics of the region. As a result of this, different species compositions exist according to climatic and altitude ranges.

In Chitwan district, the dominant forest types include *Shorea robusta* forest, mixed hardwood forest and riverine khair-sissoo forest (Panta et al., 2008). The forest type in Kayarkhola watershed can be broadly categorised as tropical broad-leaved forest. The watershed lies in the low hills and inner terai area and is endowed with rich collections of tree species. The dominant tree species found in the watershed is *Shorea robusta*, constituting about 65% of the forest composition. *Shorea robusta* is a woody species with high commercial significance and it is highly sought for because of its numerous exclusive uses. Associate tree species found in the watershed include *Semicarpous anacardium, Cleistocalyx operculatus, Lagerstromia parviflora, Mallotus philliplensis*.

2.1.6. Land cover / Land use

Chitwan district is largely dominated by forest land areas covering about 60% of the total land area of 128,500 hectares. This is primarily because the district hosts two conservation areas namely Chitwan National Park (970 km^2), popularly known as the world heritage site, and part of Parsa Wildlife reserve.

Agricultural lands and urban areas cover about 40% (89,500 hectares) of the area. Kayarkhola watershed has a total area of 8,002 hectares with huge forest areas. The forest areas are classified into dense forest (4119 hectares) and sparse forest (1702 hectares) with the greatest proportion being dense forest (51.48%) – see table 2-1. The watershed comes under the control of the community forest user groups (CFUGs) commissioned to manage the forest resources. There are 16 CFUGs within the watershed area with a total size of 2381.96 hectare. Each community forest has its own management plans and strategies for forest management but it is yet to become fully operational for sustainable management (ICIMOD et al., 2010). The summary of land cover/land use classification in Kayarkhola watershed is given in – see table 2-1.

Table 2-1: Classification of Land cover and land Use in Kayarkhola Watershed

Land cover type	**Area (Ha)**	**Area (%)**
Close to open broadleaved forest (dense)	4119	51.48
Open broadleaved forest (sparse)	1702	21.27
Natural water bodies	31	0.39

Bare soil	30	0.38
Grassland and degraded forest	0	0
Agriculture land and built-up areas	2038	25.47
Unclassified (clouds)	81	1.02
Total area	**8002**	**100**

Source: (ICIMOD et al., 2010) - Land cover analysis report 2010

2.2. Criteria for selection of the study area

2.2.1. REDD+ Pilot Project in Nepal

Kayarkhola watershed of Chitwan district is one of the three watershed areas under the REDD pilot project implementation. Nepal is one of the pioneer countries to implement the REDD pilot project in preparation towards the development of REDD+ strategy for the country. The goal of the REDD project in Nepal is to pilot the feasibility of a payment mechanism in Nepal's community managed forests for conservation, enhancement of forest carbon and REDD. The project covers 104 CFUGs with an area size of over 10,000 hectares (ICIMOD et al., 2010). The objectives of the REDD project are been implemented by a consortium of three institutions (ANSAB, ICIMOD, FECOFUN) with active participation from local communities. The project is funded by the Norwegian Agency for Development Cooperation (NORAD). The ongoing implementation of the REDD pilot project in Nepal formed the main criteria for selecting Kayarkhola watershed as the study area.

2.2.2. Data Availability

Availability of data is another important consideration for the selection of the study site. The on-going Forest Resource Assessment (FRA) project in Nepal (2010-2014) adopted an integrated approach to facilitate the project. This integrated approach entails the combined use of airborne laser scanner data (covering a fraction of the entire project area), aerial imagery data (collected alongside the laser scanning data), and medium resolution satellite data (covering the entire project area) in conjunction with field measurement data to achieve a cost effective and robust estimation of forest resources in Nepal.

Kayarkhola watershed is part of the area covered by the airborne laser scanner data collection mission for the FRA project. Hence, the LiDAR and aerial photo data exploited in this study were sourced from the FRA project. Other additional data such as shapefiles data of the CFUGs and study area boundary were sourced from ICIMOD.

2.2.3. Study Area Accessibility and Time Constraints

The Kayarkhola watershed region is characterised by rugged and undulating terrain with very high elevation attributes (245-1944m). Due to the challenges of terrain, time limitation and budgetary constraints, it was impossible to carry out fieldwork in the entire community forests (CFs) within the Kayarkhola watershed. Therefore, only 5 CFs were sampled out of a total of 16 CFs in the Kayarkhola watershed. The 5 CFs were easy to access with moderate elevation values relative to the rest of the CFs.

3. METHODS, MATERIALS AND DATA

3.1. Research method

Forest parameters (DBH, height) were collected in the field from sampled plots. DBH was converted to aboveground biomass using allometric equations (for *Shorea robusta* and other species). The remote sensing data (LiDAR data) was used to generate the Digital Terrain Model (DTM), Digital Surface Model (DSM) and the Canopy Height Model (CHM). The CHM was validated by comparing its height measurement with field height measurement. Furthermore, predictive variables (LiDAR metrics) were extracted for each sampled plot (plot-based approach) from LiDAR. Subsequently, the predictive variables were related to derived field aboveground biomass using statistical analysis to estimate the total aboveground biomass for the study area. Two statistical analysis methods were applied and compared (regression analysis and regression kriging) to determine the method that provides a more accurate estimation of aboveground biomass. A schematic representation of the research method is shown in figure 3-1. Detailed explanation is provided in the following sub-sections.

3.2. Materials and data

This study made use of the following materials and data to realize its objectives:

1. Remote sensing data - LiDAR data, Orthophoto image and GeoEye-1 image
2. Software tools – GIS software, LiDAR software tools, Statistical tools, Reporting tools
3. Field location tools – iPAQ, Global Positioning System (GPS), Compass
4. Field measurement tools – Diameter tape, Measuring tape, Laser Altimeter, Suunto Clinometer, Spherical densiometer
5. Field data - DBH, Height, Canopy cover

The following sub-sections provide the full description of how the materials and data were utilized for this study.

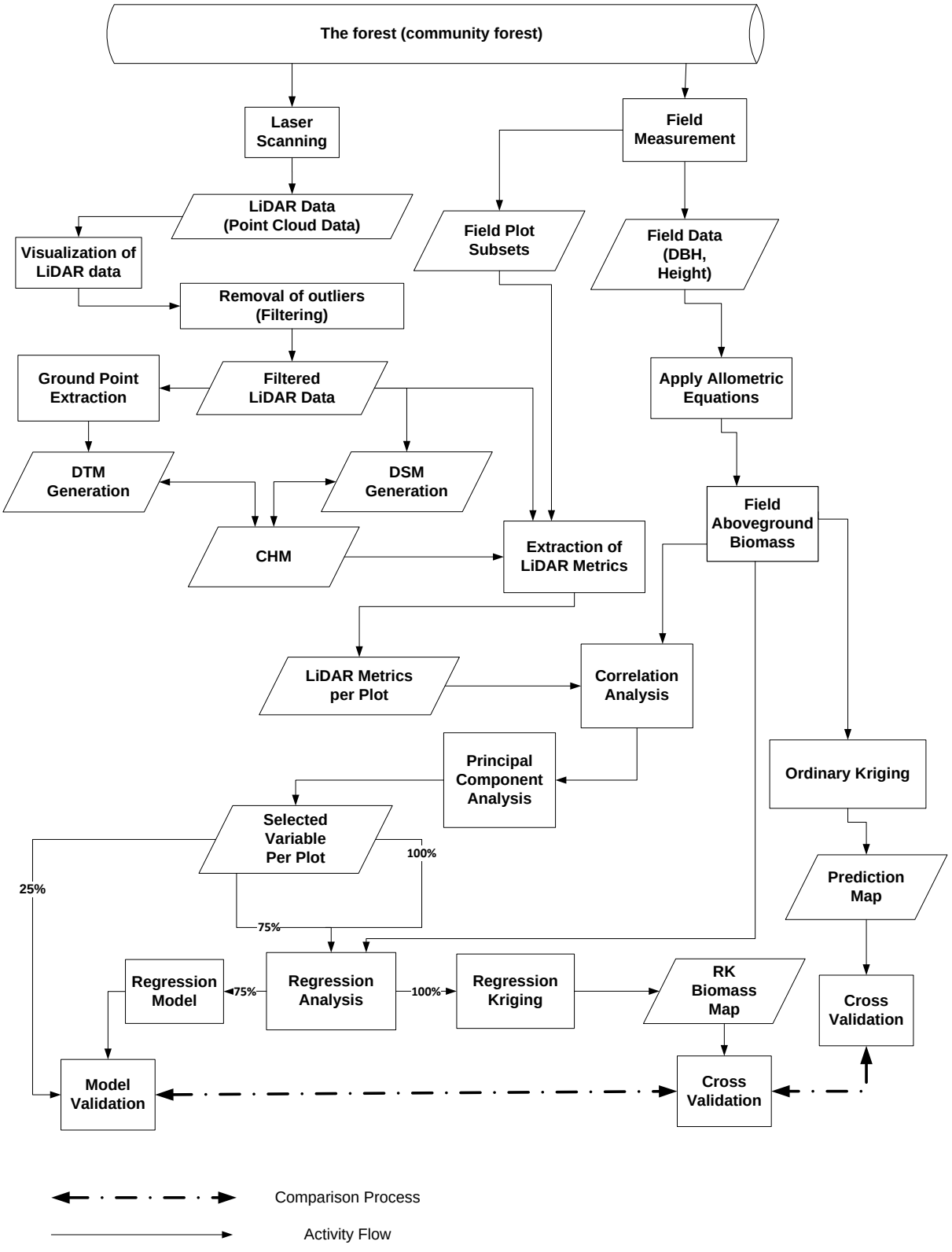


Figure 3-1: Research Method

3.2.2. Remote sensing data

LiDAR data and the Orthophoto image were the main remote sensing data used for the analysis of above-ground biomass in this study.

(A). LiDAR data

The LiDAR data employed in this study was collected by Arbonaut Limited (a Finnish company) under the on-going Forest Resource Assessment (FRA) Project in Nepal (2010-2014) - a joint cooperation between the Government of Nepal and Finland. The FRA project was established to build a National Forest Inventory for Nepal. The flight campaign for acquiring the LiDAR data took place between March and April 2011 with a Leica ALS50-II LiDAR scanner flown from a 9N-AIW helicopter over the study area. The LiDAR data is a point cloud density data with an average return pulse density of 0.8 points/m^2 - see Table 3-1 for additional information.

Table 3-1: Features of the LiDAR data and sensor parameters

Features	Descriptions
Customer	Forest Resource Assessment in Nepal, Ministry of Forests and Soil Conservation
Date Flown	2011-03-16 / 2011-03-28 / 2011-04-01 / 2011-04-02
Times of collection (UTC)	02:45 – 08:20 / 03:46 – 05:00 / 04:01 – 05:45 / 03:31 – 05:30
Date Processed	2011-05-30
Projection	UTM
Datum	WGS84
Files format	ASPRS LAS file
Aerial Platform	Helicopter (9N-AIW)
Flying altitude	2200 m AGL
Flying speed	80 knots
Sensor pulse rate	52.9 khz
Sensor Scan speed	20.4 lines/second
Nominal outgoing pulse density @ground level	Average: 0.8 points/ m^2
Scan FOW half-angle	20 degrees
Swath @ ground level	1601.47 m
Point spacing	max 1.88 m across, max 2.02 m down
Beam footprint @ ground level	50cm

LiDAR data was employed to extract the independent variables (LiDAR metrics) required to carry out regression analysis and estimate the aboveground biomass in the study area. The LiDAR data employed for this study were already post-processed by Arbonaut Limited using

TerraScan Microstation extension (commercial LiDAR software) and ArboLiDAR tools (customized LiDAR processing software by Arbonaut Limited).

The post-processing of the LiDAR data, carried out by Arbonaut Limited, involved classifying and filtering the overlapping points caused by partly overlapping flight route. Three LiDAR point cloud classes were defined after the filtering process, by Arbonaut Limited. The Arbonaut-defined LiDAR point cloud classes are: default (1), ground (2), and error (7).

The classified point data were delivered in LAS format (a binary LiDAR exchange data format). The LAS formatted data contained the X, Y, Elevation, Intensity, Pulse number, Return number, and Nadir angle values (shown on Appendix 2-1) with projection in Universal Transverse Mercator (UTM), Zone 45 North coordinate system and datum, WGS84. Eighteen LAS files tiles covered the study area (see Appendix 2-2).

(B). Orthophoto image

An Orthophoto image is an aerial photograph that has been ortho-rectified. High resolution orthophoto images were simultaneously collected alongside LiDAR data by Arbonaut Limited during the flight campaign between March and April, 2011. The orthophoto image matches well with the LiDAR data because they were acquired at the same time. It has a spatial resolution of 0.45m.The image was delivered with a projection to UTM, Zone 45 North coordinate system, and datum (WGS84).

An initial requirement for displaying and visualizing the LiDAR data within FUSION software - the LiDAR processing software used in this study (see sub-sections 3.2.2A) - is to load an image of the area of interest i.e. the orthophoto image. In this study, the orthophoto image was pre-loaded into FUSION to aid the analysis of LiDAR data in FUSION.

(C). GeoEye-1 image

The GeoEye-1 image used for this study was acquired on November 2rd, 2009. It consists of the panchromatic image at 0.41m and multispectral images (4 bands) at 1.65m. A large scale format map of GeoEye-1 image of each sample plots, showing the location of the plot center and plot area of 500m^2 (circle of radius 12.62m) were prepared, (see Appendix 3). The GeoEye-1 image aided easier and faster navigation to the plots and the identification of some trees in the plots. LiDAR data and the Orthophoto image were not available during the pre-fieldwork stage; hence GeoEye-1 image (0.5m resolution) was utilised.

3.2.3. Software tools

Several software tools (see table 3.3) were utilized for data analysis, processing and presenting the study results. FUSION/LDV is the main software tool utilized in this study. Additional information about FUSION/LDV is provided in sub-section (A) of section 3.2.2.

(A). FUSION/LDV software

FUSION/LDV software is a free LiDAR software (commonly referred to as FUSION software) developed and continually maintained by the "United States Department of Agriculture (USDA), Forest Service, Pacific Northwest Research Station in collaboration with the University of Washington" (McGaughey, 2007). It is called FUSION/LDV because it consist of two main programs, FUSION and LDV (LiDAR data viewer), and a collection of command line programs. FUSION is a comprehensive system designed to handle LiDAR dataset with robust and sophisticated functionalities. FUSION is the main graphical user interface (GUI) for displaying and projecting LiDAR data in 2-dimensional (2D) views, typical of geographic information system (GIS) software tools (McGaughey, 2007). However, LDV enables a 3D visualization, measurement and analysis of LiDAR dataset. The tool also supports command line programming and provides capabilities suitable for processing large datasets within the FUSION package. FUSION can handle various file formats such as shapefiles, images, digital terrain model, and LiDAR point data. Numerous studies involving LiDAR data have been carried out using FUSION/LDV software and published in several journals of remarkable repute (Angelo et al., 2010; Baker et al., 2010; Beets et al., 2011; Gaulton & Malthus, 2010; Goerndt et al., 2011; Kane et al., 2010b; Richardson & Moskal, 2011). The extensive use of FUSION/LDV software indicates the robustness of the tool for LiDAR analysis and justifies its use in this study.

Table 3-2: List of software used for this study

Software	Application
ArcGIS 10	GIS Operations Random sample plot generation and layout. Data visualization and creation of layout for presentation
LAStools	LiDAR data exploration Extraction of ground points from LIDAR point cloud
FUSION/LDV 2.90	LiDAR data visualization and analysis Generation of the DTM, DSM, and CHM LiDAR metrics generation
Quick Terrain Modeller	Filtering of LiDAR data
R Statistical Package (2.14.1),	To perform descriptive statistics e.g. Normality test

Microsoft Excel, SPSS	Statistical analysis e.g. Correction, Regression For Geostatistical analysis - Regression Kriging
Microsoft Office tools 2010 (Word, PowerPoint, Visio)	For report writing, presentation and design of flowcharts and diagrams

3.2.4. Field material and equipment

The following material and equipment, shown in table 3-4, enhanced the data collection and field work process. A detailed explanation of the field work activity is presented in section 3.5 - Fieldwork.

Table 3-3: Field material and equipment used for the study

S/N	Materials and Equipment	Purpose
1	GPS and IPAQ (handheld navigation computer)	Navigation and plot identification
2	Compass	Navigation and direction in the field
3	Stationeries (Pencil, Sharpener, Marker), Clipboard & Datasheet	Recording of field data and marking of measured trees in the field.
4	Diameter Tape (5m)	Measuring DBH of trees in the field
5	Measuring Tape (30m)	Plot layout measurement and crown diameter
6	Laser altimeter	Height measurement
7	Suunto Clinometer	Slope measurement
9	Spherical Densiometer	Canopy cover measurement

3.3. Research stages

The details of the fieldwork and how materials and methods were applied to the fieldwork process is described in subsequent sub-section. The research is divided into three stages;

- Pre-fieldwork
- Fieldwork
- Post-fieldwork

3.4. Pre-fieldwork

The pre-fieldwork stage involved designing of sample plot, and preparation of materials for the fieldwork.

3.4.1. Sampling design

Stratified random sampling approach was adopted in this study. Kayarkhola watershed has varying characteristics in terms of tree density, altitude, slope, stand structure, etc. The land cover types of the watershed range from sparse forest to dense forest (table 2-1). Based on the varying

land cover characteristics in the area, the watershed had already been demarcated into 16 distinct community forests enhancing homogeneity within each stratum. Each community forest represents one stratum.

The study area covers 5 different community forests (figure 3.0) with an area size of 764.67 hectares (table 3-5). Stratified random sampling yields more accurate estimate of forest parameters than random sampling (Husch et al., 2003) in forests with land cover type similar to Kayarkhola watershed. This approach ensures a representative sample is obtained from each stratum and reduces variance by stratifying the area into homogenous blocks. Prior data, available through study done in 2010 by ITC NRM students in the study area, was exploited in calculating the total sample size. According to Husch et al., (2003), preliminary information of the area is required to estimate the sampling unit. A number of sampling units per stratum was allocated in proportion to each stratum area. The sample size for each stratum was estimated using the formula in equation 1 (Husch et al., 2003).

$$n = \frac{t^2 \sum_{j=1}^{M} P_j S^2}{E^2} \qquad \text{(Equation 3-1)}$$

Where:

n = minimum number of samples required

t = value associated with specified probability

M = number of strata in population

P_j = proportion of total forest area in jth stratum = Nj/N

s^2 = variance of X for the area

E = allowable standard error in units of X

Nj = Area size of stratum

N = total size of the area (Husch et al., 2003)

For an area size of 764.67 hectares, a total of 73 plots were estimated to be sampled in the field. However, in the field, forest parameters were collected from 101 plots across the study area, see table 3-5 and figure 2-1. This sampling method was applied for the two mapping approaches (regression and geostatistical analysis).

Table 3-4: Area size and number of sample plots for the Community Forests

Strata	Area	Sample plots
Devidhunga CF	253.86	28
Dharapani CF	147.16	16
Kalika CF	213.77	30
Satkanya CF	58.28	13
Kankali CF	91.6	14
TOTAL	**764.67**	**101**

3.4.2. Pre-field work preparation

Each sample plot center was displayed on the GeoEye-1 image and projected on a larger scale. It was printed out and used for the fieldwork to aid faster navigation to the plots. The data sheets used for recording information on forest parameter collected in the field were also prepared and printed (see appendix 2) in preparation for data collection on the field. Field material and equipment, listed in section 3.1.2, were obtained and checked in advance prior to using them in the field.

In order to prepare the iPAQ and ensure it is ready for use in the field, the iPAQ was calibrated by re-running its backup files, to restore the iPAQ to its default factory setting from the last use. Data for the fieldwork (study area boundary, community forest strata boundary, sample plots centers, GeoEye-1 image of the study area) were loaded into the iPAQ. The GeoEye-1 image was converted into iPAQ file format, Enhanced Compression Wavelet (ECW) format, using ArcGIS 10 software before loading the image into the iPAQ.

3.5. Fieldwork

The purpose of the fieldwork is to measure and collect forest parameters (DBH, Height) from sample plots in the study area. The fieldwork took place in Nepal between September 17th, 2011 and October 19th, 2011. While, the actual field data collection within Kayarkhola watershed, Chitwan District, Nepal took place between September 23th, 2011 and October 13th, 2011. The data was collected in collaboration with two other colleagues (Amado Adalberto and Purity Rima), carrying out similar fieldwork exercise in the area. In addition, the field team received support and assistance from the local community, enabling easy navigation of terrain.

3.5.1. Field data collection

The iPAQ, GPS, Compass, and the printed images showing each of the plot centers aided and eased navigation to the sample plots within the forest. In each plot, the center of the plot was located and determined; in most cases a tree very close to the plot center was selected. The XY

coordinate of the center was marked and stored on the GPS and the IPAQ. From the center of the plots, the compass was used to determine the Northern direction which served as a start direction for tree measurement in all the plots. This ensured uniformity and consistency for the process of measuring of trees in each plot. With the aid of a measuring tape, circular plots of 500m^2 (0.05 hectares) area were defined. This translates to a plot size with a radius of 12.62m on a flat terrains (less than or equals to 5%). Most of the plots in the study area had slope greater than 5%, due to the mountainous and hilly nature of the terrain. Slope was measured using the Suunto clinometer in the direction of maximum slope. A slope correction factor (Appendix 4) was applied to plot measurement with slope greater than 5% to ensure the radius of the plot is properly defined. The slope aspect was also measured for each plot with the compass. The elevation was measured with the GPS and photographs of each plot taken.

Forest parameters such as DBH at 1.3m (using diameter tape), tree height (using LASER altimeter), crown diameter (using measuring tape), and canopy density (using spherical densiometer) were measured concurrently in the field. Trees with DBH less than 10cm have been shown to have little contribution to the total biomass of the forest (Brown et al., 2002). Hence, only trees with DBH greater than 10cm and the associated tree species were recorded. Tree heights were measured at a distance of about 10m from the tree of interest. Due to limited time constraint and the difficulties associated with measuring the height of trees in the field, only height of few trees were measured, in particular height of the trees used to delineate the plot center. The crown diameter was determined by averaging the measurements of the length of the crown taken at two sides of the tree. The canopy density was estimated from the four corners of the plot and the plot center. The average of the measurements were calculated and then expressed as a percentage. Refer to table 3-4 for the list of field equipment and their purpose.

3.6. Post-fieldwork – data analysis

The purpose of this research stage was to carry out extensive data analysis on the field data and LiDAR by using statistical analysis tools. The following subsections describe the sub-stages that make up the post-fieldwork research stage.

3.6.1. LiDAR data verification and visualization

LiDAR data for the study area were displayed in FUSION software GUI providing a 2D view of area. Using FUSION's LiDAR Data Viewer (LDV), LiDAR point cloud data for different parts of the study area was displayed in 3D view to visualize and examine the LiDAR data. The process revealed that the LiDAR data had some outliers. The areas with outliers were represented by extreme elevation values, which do not indicate of the true elevation of the area. The outliers

were visualized in 3D view as sets of floating point clouds in the point cloud distribution. A comparison of the range of elevation values displayed on the histogram bar in figure 3.3 (a section of LiDAR data with no outliers) and figure 3.4 (a section of LiDAR data with outliers) is shown below. Outliers were removed using the filtering process described in section 3.6.3.

In addition, a summary data report was generated with FUSION providing information on the minimum elevation, maximum elevation, total returns, and nominal return density for each LiDAR LAS tile covering the study area (Appendix 5-1). About 22.4 million LiDAR point returns were recorded from the data.

3.6.2. Removal of outliers from the LiDAR data

For this study, the filtering process was carried out to remove the outliers revealed in the study area data during the visualization process, described in section 3.6.1, and ensure that errors are minimized in the subsequent analysis. The presence of outliers in a dataset will affect the quality of the DTM, DSM and the CHM produced. The LiDAR point cloud data for the affected areas, in the study area, was filtered using the "cut tool" in Quick Terrain Modeler software. Figure 3-2 shows an example outcome of the filtering process.

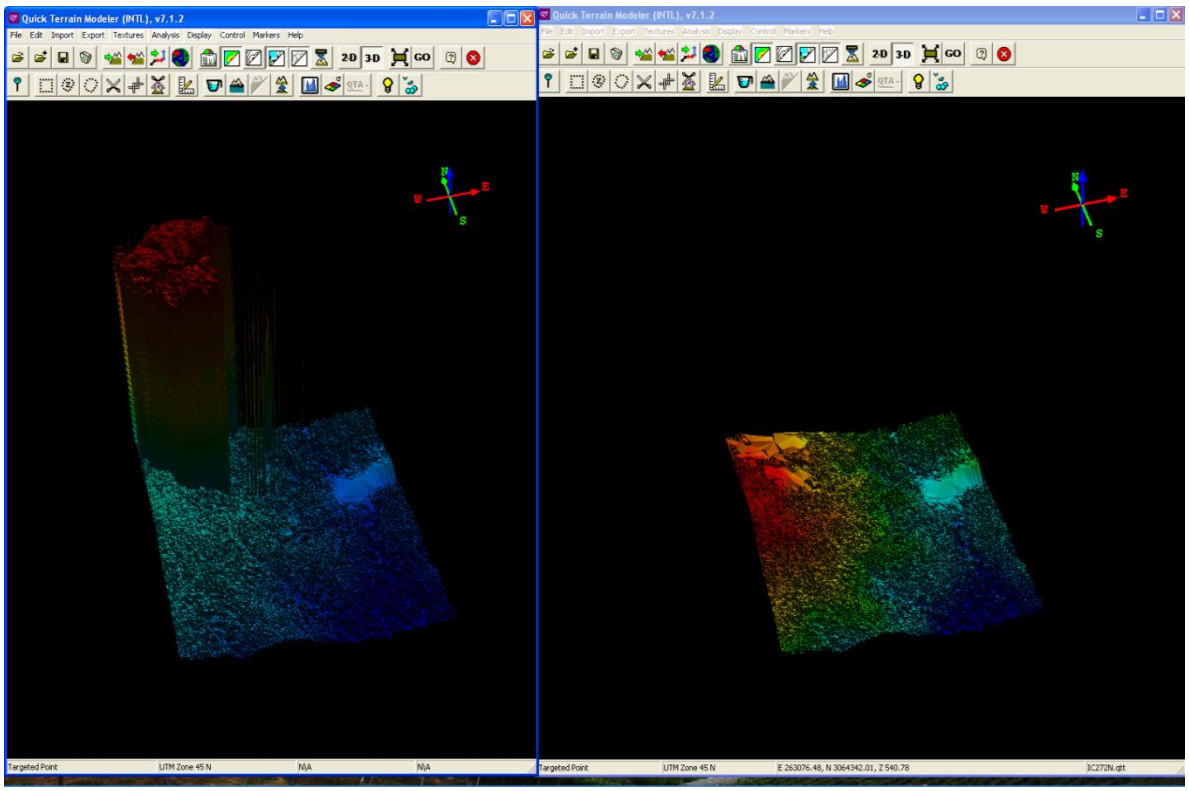


Figure 3-2: Filtering of LiDAR LAS file IC272 in Quick Terrain Modeler (a) On the left side - LiDAR data before filtering (b) On the right side - LiDAR data after filtering

3.6.3. Above-ground biomass calculation using allometric equation

Allometric equation is commonly used in the calculation of aboveground biomass from field data, mostly using DBH. Due to the absence of species-specific allometric equations for most tree species in the study area, aboveground biomass was estimated using the IPCC general broadleaf group-wise allometric equation(see equation 3-4) (IPCC, 2007) while for *Shorea robusta,* species-specific allometric equation developed by Basuki et al.,(2009) was applied (equation 3-2 and 3-3). The IPCC equation is for tropical moist hardwood forest. Tropical moist forests are characterised with receiving an annual rainfall of between 2000mm to 4000mm (IPCC, 2007). The IPCC equation was adopted in this study because the study area falls under the classification of a tropical moist forest and exhibit similar climatic condition. Similarly, the species-specific allometric equation for *Shorea robusta* was developed for East Kalimantan, Indonesia, which experiences an average annual rainfall of 2000m and the temperature range from 21 ℃ to 34 ℃ (Basuki et al., 2009). By virtue of the similarity in the climatic condition with the study area, this species-specific allometric equation for *Shorea robusta* has been adopted in this study. Both allometric equations were developed with DBH alone.

1. For *Shorea robusta*

$$\ln(\text{TAGB}) = -2.193 + 2.371 \times \ln(\text{DBH})$$ ..... Equation 3-2

Where:

TAGB = Total aboveground biomass

DBH = Diameter at breast height

Correction factor (CF) = 1.034 (Basuki et al., 2009)

The allometric model for Shorea robusta was developed from logarithm (log) transformed variables in order to satisfy one of the prerequisites for parametric regression. The process of log transformation tends to equalize the variance over the entire range of biomass values (Sprugel, 1983). Also, transformation introduces a systematic bias in the calculation. To nullify the bias, a correction factor (CF) was applied (Basuki et al., 2009; Sprugel, 1983). The correction factor formula is presented in equation 3-3. TAGB calculations for *Shorea robusta* were multiplied with the correction factor given in equation 3-3.

$$CF = EXP^{(SEE^2/2)}$$ -- Equation 3-3

Where:

CF = Correction Factor

SEE = Standard Error Estimate of the regression (Sprugel, 1983).

2. For other species

$$Y = e^{[-2.289 + 2.649 \times \ln(DBH) - 0.021 \times \{\ln(DBH)\}^2]}$$ ……..... Equation 3-4

Where:

Y = Total above-ground biomass

DBH = Diameter at breast height (IPCC, 2007)

Total aboveground biomass (TAGB) was estimated for each tree and then aggregated per plot. TAGB per plot was converted from Kg/m^2 to Mg/ha.

3.6.4. Generating the Digital Terrain Model (DTM)

The DTM serves as a ground reference for the generation of the canopy height model (CHM). The following steps were employed for generating the digital terrain model (DTM):

- Classify the LiDAR point clouds into ground (terrain) points and vegetation (non-terrain) points

- Generate DTM using the extracted ground points

The ground points of the LiDAR point cloud were extracted using LAStools according to the initial classification that was performed by the vendor (Arbonaut Limited). With the extracted ground points, DTM of 1m, 3m, and 5m were generated in FUSION software with the "GridSurfaceCreate" program. Elevation value per cell was computed by taking the mean elevation of all points within the cell. For cells with empty elevation values, interpolation was done using neighbouring cells. The syntax/command for the DTM computation in FUSION (McGaughey, 2007) was processed using the command line prompt (CMD) and herewith given in section 3.6.5.

```
GridSurfaceCreate output_file cellsize xy_unit z_unit coordinate_system zone horizontal_datum vertical_datum datafile
```

Where:

output_file = name of the DTM file generated by the process

cellsize = model resolution

xy_unit = LiDAR data xy unit (m)

z_unit = LiDAR data elevation unit (m)

coordinate_system = LiDAR data coordinate system - UTM

zone = 45

horizontal_datum = LiDAR data horizontal datum – WGS84 (World Geodetic System of 1984)

vertical_datum = LiDAR data vertical datum - WGS84

datafile = extracted ground points of the study area

3.6.5. Generating the Digital Surface Model (DSM)

The digital surface model generated in this study represents the vegetation elevation in an area. The DSM is computed using the highest elevation value in each grid cell. DSM was generated in FUSION LiDAR point cloud along with the "CanopyModel" program. The DSM was generated for 1m, 3m, and 5m pixel sizes respectively. The syntax/command for the DSM computation in FUSION was processed using the command line prompt and is given in section 3.6.6 as:

```
CanopyModel output_file cellsize xy_unit z_unit coordinate_system zone horizontal_datum vertical_datum LiDAR_datafile
```

Where:

output_file = name of the DTM file generated by the process

cellsize = model resolution

xy_unit = LiDAR data xy unit (m)

z_unit = LiDAR data elevation unit (m)

coordinate_system = LiDAR data coordinate system - UTM

zone = 45

horizontal_datum = LiDAR data horizontal datum – WGS84 (World Geodetic System of 1984)

vertical_datum = LiDAR data vertical datum - WGS84

LiDAR_datafile = LiDAR point cloud

3.6.6. Generating the Canopy Height Model (CHM)

The CHM generated in this study represents the vegetation heights of the area. The CHM is also referred to as the normalised DSM (nDSM) because vegetation heights are normalised to the DTM. The difference between the DSM and the DTM models yields the CHM model. CHM was created using DTM as a reference. In FUSION, CHM for 1m, 3m and 5m pixel sizes respectively were generated by running the "CanopyModel" program using the command line prompt syntax stated below in section 3.6.7.

```
CanopyModel /ground:DTM_file output_file cellsize xy_unit z_unit coordinate_system zone horizontal_datum vertical_datum LiDAR_datafile
```

Where:

output_file = name of the DTM file generated by the process

cellsize = model resolution

xy_unit = LiDAR data xy unit (m)

z_unit = LiDAR data elevation unit (m)

coordinate_system = LiDAR data coordinate system - UTM

zone = 45

horizontal_datum = LiDAR data horizontal datum – WGS84 (World Geodetic System of 1984)

vertical_datum = LiDAR data vertical datum - WGS84

/ground:DTM = The syntax for specifying the DTM used in normalising the data

LiDAR_datafile = LiDAR point cloud

3.6.7. Validation of the CHM and optimum sampling grid selection

In this study, the accuracy of the CHMs (pixel 1m, 3m, and 5m) were assessed by comparing the CHM heights with the heights of trees measured in the field. Height values of trees derived from the field were regressed with CHM heights for each of the resolution sizes. The validation result was quantified using their coefficient of determination (r^2) value to ascertain the appropriate pixel resolution to employ for this study. Of the 3 models (pixel size 1m, 3m, and 5m) generated, the model with a highest r^2 was selected as the appropriate pixel size for this study. For this analysis,

only the height value of the trees at the plot centre was correlated with the height value derived from the CHM model. The comparison process validates CHM and determines the most suitable pixel resolution for CHM applied to this study. In this study, the 3m CHM model was employed.

3.6.8. Extraction of the Field Plots

For LiDAR metrics to be extracted for each plot in this study, the circular area of the plots was calculated in Microsoft Excel using the plot center coordinates and the radius for each plot. The slope measurement of individual plot on the field determined the circular dimensions of the plot. The circumference of each plot was derived by estimating the X and Y minimum, and X and Y maximum either by subtracting or adding the radius to the X and Y coordinates, see equation in this section 3.6.9.

X minimum = X (coordinate) - r

Y minimum = Y (coordinate) - r

X maximum = X (coordinate) + r

Y maximum = Y (coordinate) + r ……..(McGaughey, 2007)

With the circular dimensions, LiDAR point cloud for each plot was extracted with the 'ClipData' program in FUSION. The syntax for clipping the LiDAR point clouds with the plot subsets is given below in this section 3.6.9.

```
ClipData /shape:1 /DTM /height input_LiDAR_datafile output_subset_file (MinX MinY MaxX MaxY)
```

Where:

/shape: 1 = Shape of the plots (1 denotes a circular plot)

/DTM = Digital Terrain Model used in normalising the data

/height = Used to convert the elevation values to height above the ground

(MinX MinY MaxX MaxY) = Dimensions of the plot

3.6.9. Extraction of LiDAR metrics

In this study, LiDAR metrics were computed for each sample plot using only first return of LiDAR point clouds. Height limit of 2 m was used to eliminate points on the ground. The specified height limit ensures ground areas consisting of shrubs and understory vegetation were excluded from the computation (Lim & Treitz, 2004). It also ensures that heights above the height limit represented canopy areas. LiDAR metrics was extracted from FUSION using the 'CloudMetrics' program. The 'CloudMetrics' syntax for is given below in this section 3.6.10.

Three sets of LiDAR metrics were generated for each field plot. The LiDAR metrics include (I) metrics derived from using elevation values from LiDAR (elevation metrics), (II) intensity values from LiDAR (intensity metrics) and (III) canopy cover computations. The canopy cover computations were calculated as percentages of different return ratios above various specified height limit.

DTM was used together with LiDAR returns for the computation of the metrics. Using the digital terrain model, the elevation values were normalized to represent vegetative height values above the ground for the area. The LiDAR metrics were correlated with field aboveground biomass.

```
CloudMetrics /id /above:2 /first_return input_file output_file
```

Where:

/id = provides an identifier for each plot (system generated)

/above:2 = cover estimate with height limit 2m

/first_return = instructs the program to use only the first return value of the LiDAR data

input_file = LiDAR point return values per plot from extraction of field plot process

output_file = file name containing the LiDAR metrics extracted CSV (comma-separated values) file.

3.6.10. Descriptive statistics of the field data / exploratory data analysis

In this study, both dependent and independent variables were explored beforehand to ensure the data satisfies the regression analysis assumptions and meet the geostatistical requirements. Diagnostic graphics of the the variables were explored to check their distribution. Similarly, a non-graphical statistical test (shapiro-wilk test) was performed for both variables to confirm normality assumptions. Based on the normality test, variables that revealed non-normality were log transformed prior to being used for the regression analysis.

Scatterplots of AGB and each of the LiDAR metrics were observed to visualise the association and the strength of the relationship between TAGB and LiDAR metrics. This process is required because variables with weak or no association are unsuitable for modelling.

3.6.11. Correlation Analysis

In this study, correlation analysis (also known as Pearson's product) between each LiDAR metrics and field aboveground biomass was performed using SPSS software. The result of the correlation

matrix showed the inter-relationship between LiDAR metrics and AGB, and intra-relationship within LiDAR metrics. Only the variables (LiDAR metrics) exhibiting considerable relationship with aboveground biomass were considered for further analysis i.e. regression analysis.

3.6.12. Principle Component Analysis (PCA)

The result of the correlation matrix from the correlation analysis process revealed high correlation between the LiDAR metrics. When variables are highly correlated with each other, it results to data redundancy in the model. Data redundancy occurs when the inclusion of extra variables into the model adds no additional information to it. In this study, principal component analysis (a data reduction method) was performed using the 'princomp function' in the R statistical package to filter correlated independent variable and identify uncorrelated independent variables. The uncorrelated independent variables are also referred to as "Main Components".

3.6.13. Regression Analysis

Regression analysis is commonly used for biomass estimation studies. Regression analysis models quantify the relationship between dependent variable and independent variable(s). In this study, the independent variables are the LiDAR metrics selected from the PCA analysis and the dependent variables are the field TAGB.

In this study, the linear model of each of the LiDAR metrics and TAGB were developed using the 'lm function' in R. Furthermore, multiple regression analysis of various combinations of the variables was established. Subsequently, backward stepwise regression was done to generate the best model (with the best combination of variables) for estimating TAGB.

3.6.14. Extraction of the covariate for regression kriging

In this study, the selected independent variables from PCA were used for the kriging process. First, the value of the explanatory variable was extracted for the whole study. This was done using the "GridMetrics" program in FUSION. A grid size of $100m^2$ was used across the study to compute the metrics per each grid cell. The output of the "GridMetrics" process was a CSV file containing the coordinates of the grid cell and metrics value for each grid cell. This was converted into an ASCII file with the "CSV2GRID" program in FUSION. The raster ASCII file was imported into ARCGIS 10 and the area of study was masked out using the study area boundary shapefiles. The ASCII file of the explanatory variable was imported into R software for the regression kriging analysis. The computation of the regression residuals of the model was done in R followed by ordinary kriging of the regression residuals (Hengl et al., 2007)

3.6.15. Model calibration and validation

Model validation involves assessing the performance of the model in predicting the target variable (TAGB). Model validation assesses the accuracy between LiDAR metric predicted model of TAGB with field calculated TAGB.

In this study, the biomass predicted with the model is compared with field estimated biomass. Mean Error (ME) and Root Mean Square Error (RMSE) are methods used to ascertain the level of precision and the amount of error in the models from the three approaches (regression analysis, ordinary kriging, and regression kriging). The formula for ME and RMSE is presented in equation 3-5 and equation 3-6).

For the regression analysis in this study, the dataset was randomly split into 2 parts. One part for was used for calibrating and the other for model validation. In order to build a robust model, more samples are usually required. Also, outliers were removed from the data. Hence the model was developed with 75% of the dataset (training dataset) and the remaining 25% dataset (test data) was used for validating the model.

Cross validation was used to assess the accuracy of the results from ordinary kriging and regression kriging. Cross validation was applied for these techniques because the dataset was too small to be split into two. In this study, the entire dataset was used for both model development and validation. In addition, a variance map that gives the uncertainty estimate of the prediction model was produced.

$$RMSE = \sqrt{\sum_{i=1}^{n}(Y_i - \widetilde{Y_i})^2 /n} \quad \text{.....Equation 3-5}$$

$$ME = \sum_{i=1}^{n}(Y_i - \widetilde{Y_i})^2 /n \quad \text{.....Equation 3-6}$$

Where,

RMSE is root mean square error of the model

Y_i is the measured/calculated value of carbon

$\widetilde{Y_i}$ is the predicted carbon value by the model

n is the number of samples

4. RESULTS

4.1. Descriptive statistics of field data

Kayarkhola watershed in Chitwan, Nepal embodies a wide range of tree species. About 65 tree species were recorded in the field and *Shorea robusta* was identified as the most dominant tree species across Kayarkhola watershed. It accounts for 60.2% of the total tree species recorded in the field as shown in figure 4-1. The distribution of tree species, including *Shorea robusta,* and other species in each community forest stratum is represented in figure 4-2. Across the strata, *Shorea robusta* had a greater percentage than other species except in Devidhunga where other species had a slightly greater percentage.

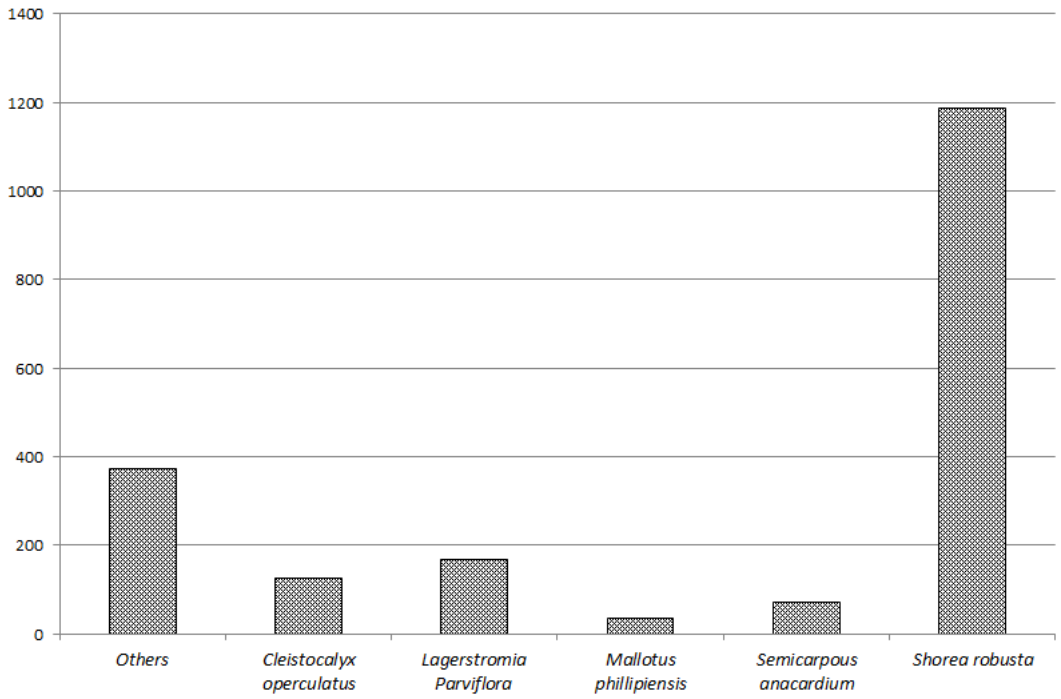


Figure 4-1: Frequency distribution of species in Kayarkhola watershed

Forest parameter data from a total of 1927 trees were obtained from 101 plots within the five constituent community forests in the study area. The distribution of trees - community forest is shown in table 4-1.

Table 4-1: Total number of trees measured

Community forest	Number of trees
Devidhunga	402
Dharapani	230
Kalika	683
Kankali	288

Satkanya	324
Total number of trees	1927

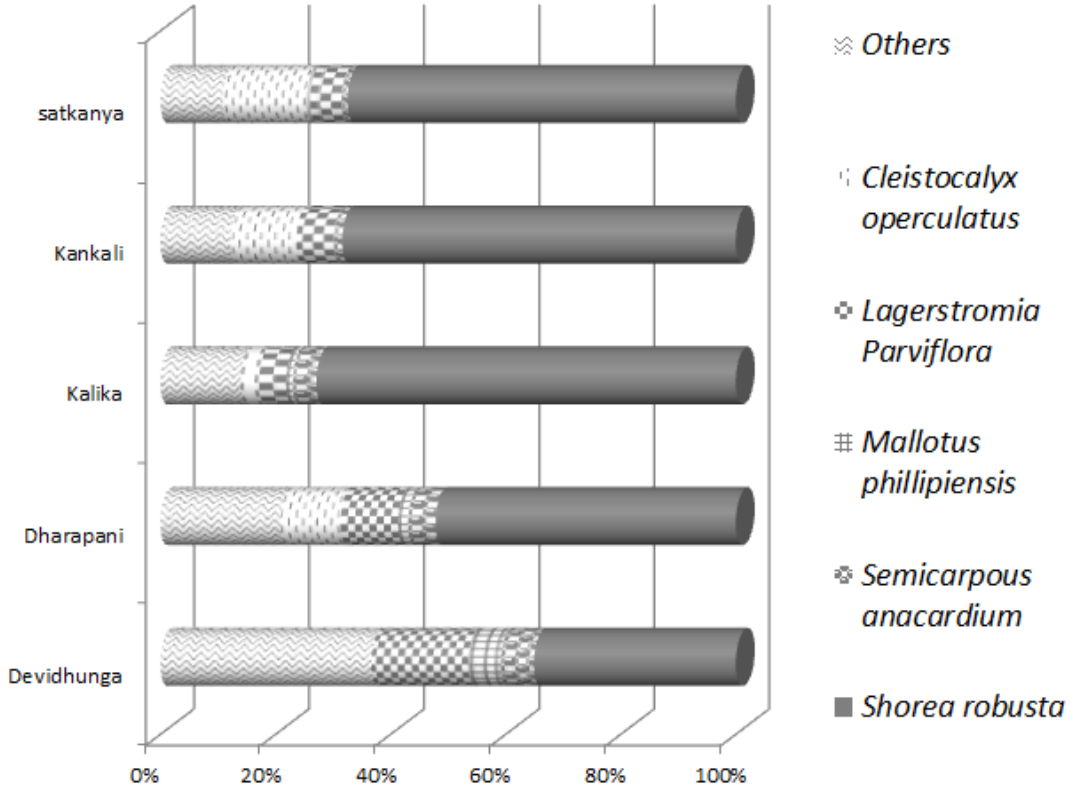


Figure 4-2: Species distribution per stratum in Kayarkhola watershed

4.1.1. Field aboveground biomass calculation using allometric equation

Aboveground biomass (AGB) was calculated for each tree in the sampled plots using their DBH measurement and allometric equations as given in section 3.6.4.Estimates of sampled plot AGB in each stratum (community forest) were summed up and then mean AGB calculated per stratum. Mean AGB estimates in Kg/500m^2 was then converted to mean AGB estimates in Megagrams per hectares (Mg/ha). Subsequently, total aboveground biomass (TAGB) for each community forest was estimated using the area size of each community forest. Table 4-2 presents the results of TAGB calculations.

Table 4-2: Total Aboveground biomass across the study area using field data

Community forest(strata)	No of plots	AGB in sampled plots (kg/500m^2)	Mean AGB in sampled plots (kg/500m^2)	mean AGB per ha(Kg/ha)	mean AGB per ha (Mg/ha)	Ln (Mean AGB per ha (Mg/ha))	Area of community forest (CF)	Ln (TAGB(Mg))
Devidhunga	28	272761.58	9741.49	194829.70	194.83	5.27	253.86	1338.38
Dharapani	16	102703.53	6418.97	128379.41	128.38	4.85	147.16	714.46
Kalika	30	221811.50	7393.72	147874.33	147.87	5.00	213.77	1068.07
Satkanya	14	61385.63	4384.69	87693.75	87.69	4.47	58.28	260.74
Kankali	13	97486.44	7498.96	149979.14	149.98	5.01	91.6	458.96
Total	**101**				**708.76**		**764.67**	**3840.61**

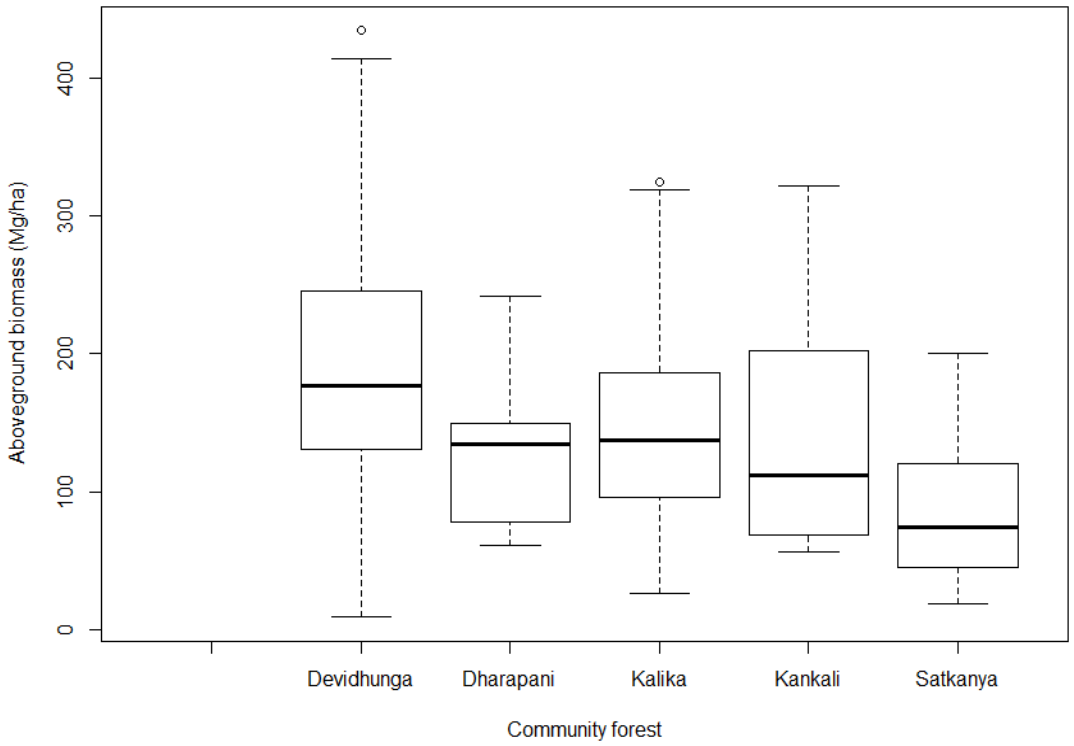


Figure 4-3: Distribution of above-ground biomass across the strata

The distribution of field TAGB across the strata is represented using the box plot shown in figure 4-3. In general, Devidhunga has the largest range and amount of aboveground biomass while Satkanya has the least amount of aboveground biomass.

4.1.2. Homogeneity of variance of above-ground biomass across the strata

Bartlett's test was applied to assess the variation in AGB among the strata and the results are shown on table 4-3. The result shows a P-value greater than 0.05, which indicates that there are no significant differences in the AGB variance among the strata. It also indicates homogeneity of variance among the strata.

Table 4-3: Bartlett's test

Barlett's K-squared	Degree of freedom	P-value
2.3653	4	0.6689

An analysis of variance (ANOVA) test was also performed to assess whether or not the mean AGB among the strata are the same. The result, shown on table 4-4, shows p-value greater than 0.05 meaning there is no significant differences in the mean AGB among the strata. Furthermore,

the ANOVA result shows that the stratification based on community forest user groups an effective criterion for establishing homogenous units of forests in the study.

Table 4-4: ANOVA result

	Degree of freedom	Sum square	Mean square	F value	P(>F)
Strata	4	31323	7831	1.074	0.374
Residuals	96	700200	7294		

4.2. Exploratory data analysis

The shapiro-wilk test was done for all the independent variables and the dependent variable. P-value of less than 0.05 shows that the residuals of the variables are not normally distributed while P-value of greater than 0.05 reveals that the residuals of the variables are approximately normally distributed. Variables with P-value less than 0.05 were log-transformed before it was used for modelling. Also, field TAGB was visualised using histogram graph and a Q-Q plot before and after transformation in figure 4-4.

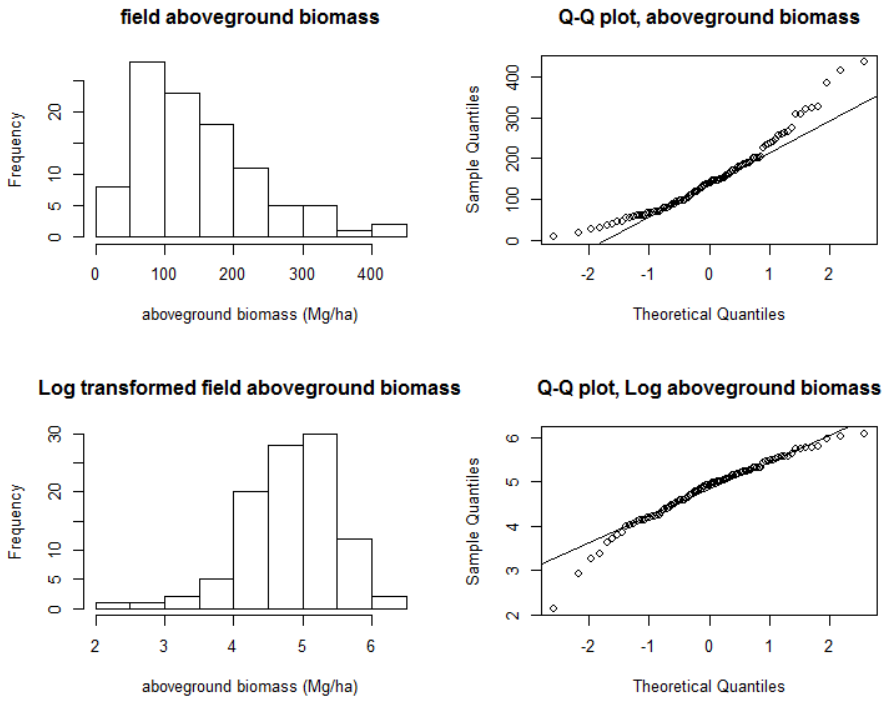


Figure 4-4: Field AGB distribution before and after logarithm transformation

4.3. The models (DTM, DSM, CHM)

The 3-dimensional view of the extracted ground points from LiDAR data used in generating the DTM is visualised in figure 4-5a and the corresponding LiDAR point cloud for the same area is

visualised in figure 4-5b. The ground points show only the laser points on the terrain while the LiDAR point cloud shows the total return for the area.

The DTM of the study is shown figure 2-1. The 2-dimensional view as displayed in ArcGIS 10 and the 3-dimensional view as displayed in FUSION are shown in figure 4-6. From the 2-D view, the lighter areas represent high height values while the darker areas denote low height values. From the 3-D view, the CHM displays as the grey surface and the green triangulated network.

4.4. Pixel resolution and validation of the canopy height model

The relationship between tree heights from field measurement with corresponding heights extracted from CHM for the different pixel sizes (1m, 3m, and 5m) respectively was assessed and is shown in a scatter plot (figure 4-7). A summary statistics of height measurements from the models and the field is shown in table 4-6. The correlation and regression analysis result between height measurement from the field and LiDAR measurement is given in table 4-7. The CHM at 3m resolution has a higher coefficient of determination (r^2) with a value of 0.54. CHM at 1m resolution has an r^2 of 0.26 and CHM at 5 m resolution has an r^2 of 0.47. The result revealed that CHM at resolution 3m had a better relationship to field height when compared with resolution 1m and 5m and therefore was selected as the pixel resolution for the models in this study.

Table 4-5: Summary statistics of height measurements from the models and the field

Summary	1m resolution	3m resolution	5m resolution	Field measurement
Minimum	1.21	4.21	4.55	3
Mean	12.95	20.17	22.53	21.17
Maximum	32.01	37.26	37.83	35
Standard Deviation	8.072	7.31	6.88	6.84

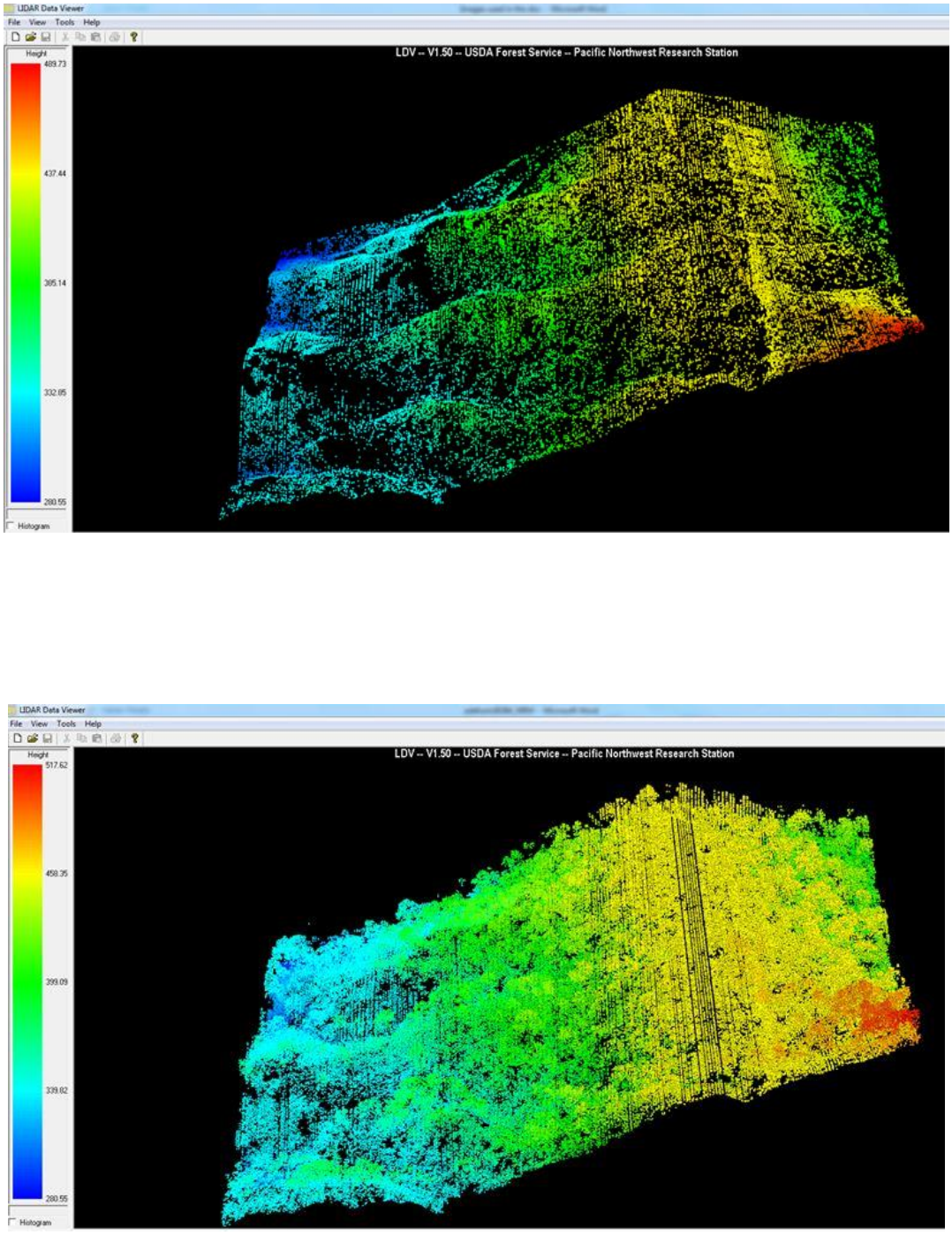


Figure 4-5: 3D view of (a) Extracted ground points from LiDAR (upper image) and (b). LiDAR point cloud of the same area in the study in FUSION software (lower image)

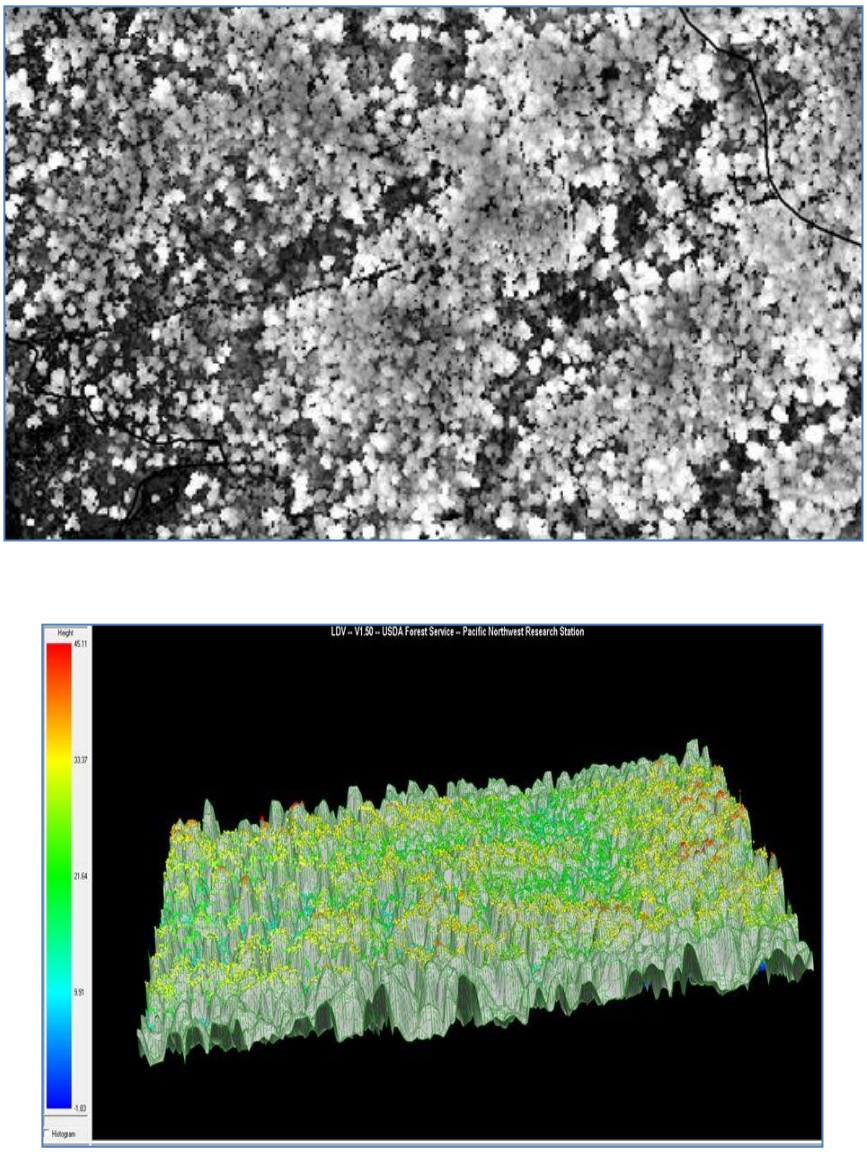


Figure 4-6: The CHM represented in (a).2-D in ArcGIS 10 (upper image) and in (b). 3D in FUSION software (lower image)

Table 4-6: Correlation and regression analysis between LiDAR height and field height for the pixel sizes 1m, 3m, and 5m

CHM resolution	**Correlation coefficient (r)**	**r^2**	**Adjusted r^2**
1m	0.51	0.27	0.26
3m	**0.74**	**0.55**	**0.54**
5m	0.68	0.47	0.47

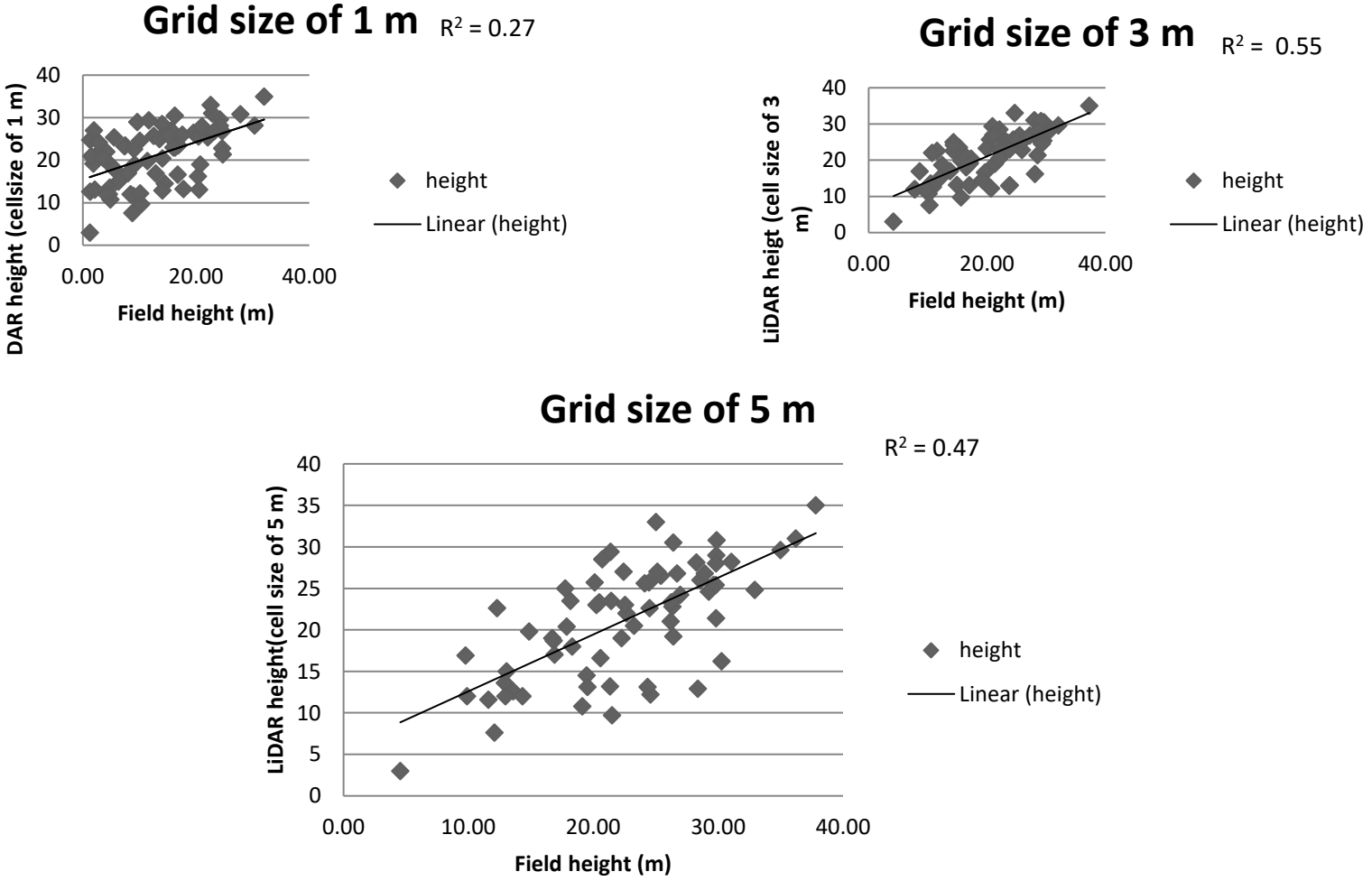


Figure 4-7: Relationship between LiDAR height from CHM and tree height measurement from the field for pixel sizes 1m, 3m, and 5m.

4.4.1. Extraction of Field plots from LiDAR

The area size of each sampled plot was masked out in FUSION using plot's XY coordinates and radius. An example of circular plot dimensions is presented in figure 4-8. Each extracted plot contains LiDAR point clouds used in the calculation of LiDAR metrics per plot. Figure 4-8 presents a field plot in FUSION LDV containing both the ground returns and the vegetative returns of LiDAR point clouds.

In figure 4-8, the blue points depict the ground returns while the greenish-orange points represent the vegetative returns symbolizing trees in the plot. In the figure, vegetative returns (points above the ground) should depict the structure of trees in the plot. The structure of trees is not fully pictured and appreciated due to low density nature of the LiDAR data used in this study.

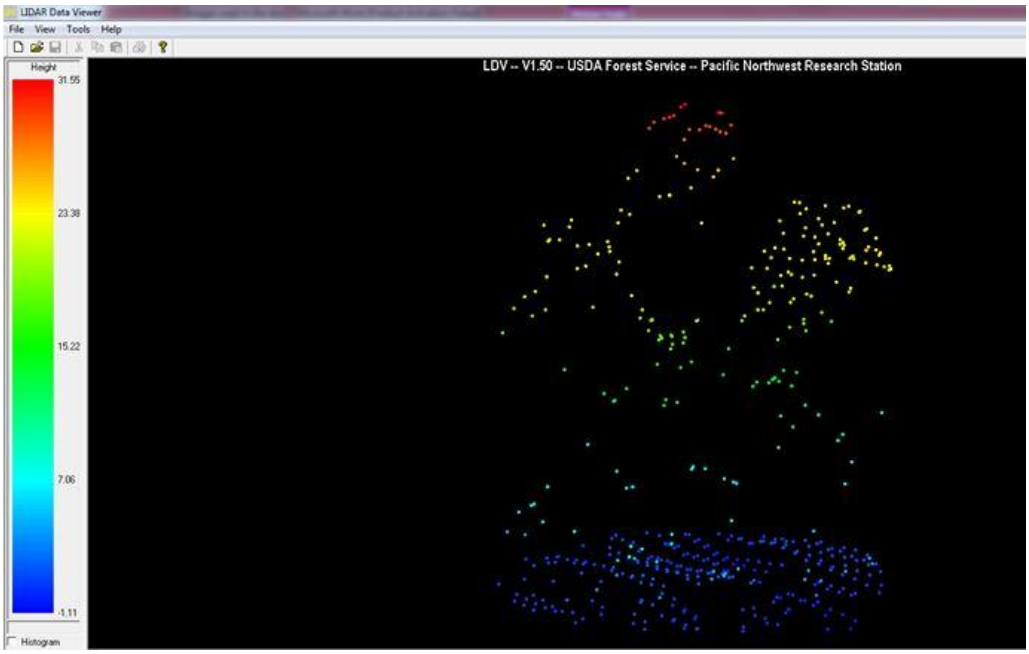


Figure 4-8: An example of a circular plot with trees in the plot displayed in FUSION LDV

4.5. Computation of LiDAR metrics

LiDAR metrics were computed from the 1st returns of the LiDAR data. In total, 81 metrics were computed using FUSION software for each sample field plots. Table 4-7 presents the list of LiDAR metrics generated for this study. Out of 81 metrics, 33 metrics were computed using elevation values (elevation metrics) and intensity values (intensity metrics) respectively and 15 metrics were calculated as ratios of returns above certain height thresholds (canopy cover computations). LiDAR data for this study were available in four return type classifications (Return 1 to return 4)

Table 4-7: LiDAR metrics computed with FUSION software

Metrics from elevation /Intensity values	Metrics of canopy cover computations
Minimum height	% of 1st returns above 2m
Maximum height	% of all returns above 2m
Mean height	% ratio of all returns above 2m to total 1st return
Height mode	1st returns above 2m
Height Standard deviation	All returns above 2m
Height variance	% of first returns above mean height
Height Coefficient of variation	% of first returns above mode height
Height Interquartile distance	% of all returns above mean height
Height skewness	% of all returns above mode height
Height kurtosis	% ratio of all return above mean height to total 1st returns
Height average absolute deviation (AAD)	% ratio of all return above mode height to total 1st returns
Height L-moments (L1, L2, L3, L4)	1st returns above mean height
Height L-moment coefficient of variation	1st returns above mode height
Height L-moment skewness	All returns above mean height
Height L-moment kurtosis	All returns above mode height
Height percentiles (1st, 5th, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, 99th)	

4.6. Correlation analysis of field TAGB and LiDAR metrics

The result of Pearson's product correlation between field above-ground biomass (AGB) and LiDAR metrics yielded the correlation matrix. The Correlation matrix revealed the strength of the relationship between field AGB and each of the LiDAR metrics, and amongst the LiDAR metrics. The strength of relationship between LiDAR metrics and field TAGB was moderately high. One-third of the LiDAR metrics generated had moderate relationship with field TAGB while two-third had a weak relationship with field TAGB. Correlation coefficient (r) for all LiDAR metrics ranged from $0.3 \leq r \leq 0.55$. Intensity metrics (metrics calculated from intensity values of the LiDAR data) showed very weak relationship with field TAGB. Their correlation coefficient (r) value ranged from $0.008 \leq r \leq 0.206$. In all, LiDAR metrics displayed both positive and negative relationship with field TAGB.

So, LiDAR metrics that had poor relationship with field TAGB, including all intensity metrics were excluded from further analysis (i.e. principal component analysis). In summary, a total of 28 variables showed a considerable relationship with TAGB and were used in the PCA.

4.7. Principal component analysis

The output of the correlation matrix revealed high correlation amongst some LiDAR metrics. These suggest multi-collinearity among the predictor variables and were tested with principal component analysis (PCA) to extract significant uncorrelated variables to be used for the regression analysis. Figure 4-9 is a scree plot that gives an indication of the number of principal components (PC) to be considered from the total for the study. The scree plot shows the proportion of variance in the principal components.

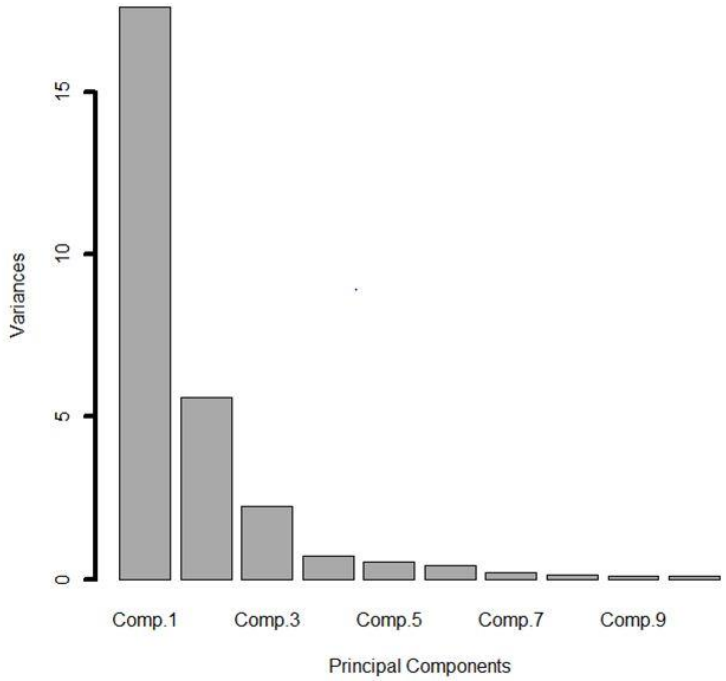


Figure 4-9: Proportion of variance in PCs

Table 4-9 displays the proportion of variance of the first four principal components carrying the bulk of the variance.

Table 4-8: Breakdown of the proportion of variance of the 1st four components

	Comp.1	**Comp.2**	**Comp.3**	**Comp.4**
Variable	Height mean	Height AAD	Canopy cover	Aramean_tfrp
Standard deviation	4.197	2.368	1.496	0.855
Proportion of variance	0.629	0.20	0.080	0.026
Cumulative proportion	0.629	0.829	0.909	0.935

*Where height AAD= Height average & absolute deviation; Canopy cover=% 1st returns above 3metres; Aramean_tfrp= (all returns above mean) / (total first returns) * 100*

The rule of thumb is to retain the PC's up to the point on the graph where the plot curves. This is because most of the variation seemed to be explained by these first sets of variable. However,

caution must be taken when determining which of the component to eliminate from a principal component regression (Quinn & Keough, 2002). Merely discarding the remaining components beyond the plot curve (that has much lesser amount of variance input to the total variance in the predictor variables) can be biased (Quinn & Keough, 2002) leading to the loss of important significant variable. In view of this, the significance of the relationship of each component was tested with TAGB (see Appendix 7).

The test revealed that 7 variables (height maximum, height average & absolute deviation, height L-moment (L2), Height L-moment skewness, 40th, 80th, and 95th height percentile) were significant predictors of TAGB. Height maximum demonstrated to be the only variable that was highly significant (0.001 confidence interval).

4.8. Regression modelling

The seven variables selected from the PCA process were used as explanatory variables in the regression analysis. Multiple linear regression of the seven variables and log-transformed aboveground yielding adjusted r^2 values of 0.56. However, none of the explanatory variables were significant. From the linear model of each of the independent variables, model of 95th height percentile was the best having an r^2 of 0.54, as shown on table 4-10. In addition, multiple linear regressions of two or more variables were explored.

From most of the combinations, 95th height percentile was the left as the only significant variable. This implies that 95th height percentile on its own can predict TAGB in the study. The inclusion of other variables in the model had no significant and additional effect in the model. The best regression model was selected on the basis of the homoscedascity of the model residuals, highest coefficient of determination (R^2), lowest Akaike Information Criterion (AIC), reduced standard error estimate, and validation statistics (RMSE) estimates. A summary of the AIC estimate and the RMSE for each model is presented in table 4-11. The lower the AIC and the Root Mean Square Error (RMSE), the better the model. Validation of each model was done with the validation dataset and the RMSE was computed. Based from the result, model of 95th height percentile was selected and used for prediction of TAGB using the different methods.

The residual from the regression analysis of TAGB and 95th height model is presented in figure 4-10. The residual gives an indication of the quality of the regression model. The residual result shows that the data fulfils the assumptions of regression analysis. From the Q-Q plot, shown in figure 4-10, it is possible to conclude that the data approximate a normal distribution

Table 4-9: Relationship between field aboveground biomass and LiDAR derived metrics

Dependent variable	Intercept	ln(hmax)	(E)AAD	el2	elskew	ln(ep40)	ln(ep80)	ln(ep95)	Adjusted R^2	Residual Standard Error	P(α = 0.05)	AIC
TAGB	-0.7516	1.6918	-	-	-	-	-	-	0.4721	0.5032	4.27×10^{-11} ***	105.0235
	(0.7153)	(0.2152)	-	-	-	-	-	-	-	-	-	-
TAGB	3.3072	-	-	1.1032	-	-	-	-	0.3771	0.5466	1.21×10^{-8} ***	116.4348
	(0.2468)	-	-	(0.1699)	-	-	-	-	-	-	-	-
TAGB	4.27088	-	-	-	-	0.30668	-	-	0.2562	0.5973	5.45×10^{-6} ***	128.677
	(0.13785)	-	-	-	-	(0.06206)	-	-	-	-	-	-
TAGB	1.8899	-	-	-	-	-	1.026	-	0.4355	0.5204	4.18×10^{-10} ***	109.6436
	(0.4099)	-	-	-	-	-	(0.1403)	-	-	-	-	-
TAGB	*0.365*	-	-	-	-	-	-	*1.4273*	*0.5411*	*0.4692*	*3.68×10^{-13}* ***	*95.36214*
	(0.5013)	-	-	-	-	-	-	*(0.1584)*	-	-	-	-
TAGB	3.96785	-	0.13231	-	-	-	-	-	0.2252	0.6097	2.26×10^{-5} ***	131.4921
	(0.20744)	-	(0.02903)	-	-	-	-	-	-	-	-	-
TAGB	4.91151	-	-	-	-1.47477	-	-	-	0.1225	0.6488	0.00186 **	140.0805
	(0.08023)	-	-	-	(0.45519)	-	-	-	-	-	-	-
TAGB	-0.09759	0.50129	-	-	-	0.0288	-0.01749	1.04496	0.5333	0.4732	-	99.35537
	(0.77424)	(0.41697)	-	-	-	(0.07151)	(0.34371)	(0.52852)	-	-	-	-
TAGB	-0.636981	0.480385	-	-0.317632	-	-0.003833	0.150182	1.245745	0.5302	0.4747	0.0394 * (ep95)	100.7215
	(1.050601)	(0.419234)	-	(0.416548)	-	(0.083533)	(0.408977)	(0.592032)	-	-	-	-
TAGB	-0.62489	0.74382	-0.14057	0.82625	-1.00853	-0.12382	-0.0248	0.99958	0.5614	0.4587	-	97.75665

Dependent variable	Intercept	ln(hmax)	(E)AAD	el2	elskew	ln(ep40)	ln(ep80)	ln(ep95)	Adjusted R^2	Residual Standard Error	P(α = 0.05)	AIC
	(1.01558)	(0.42085)	(0.08005)	(0.70255)	(0.58310)	(0.09500)	(0.50480)	(0.69111)	-	-	-	-
TAGB	0.52878	-	-	-	-	0.03424	-	1.35455	0.5361	0.4718	-	97.07046
	(0.59158)	-	-	-	-	(0.06476)	-	(0.21048)	-	-	-	-
TAGB	-0.2468	0.5214	-	-	-	-	-	1.0726	0.5462	0.4666	0.000964 *** (ep95)	95.54701
	(0.6791)	(0.3931)	-	-	-	-	-	(0.3104)	-	-	-	-
TAGB	0.6324	-	-	-	-0.4507	-	-	1.348	0.5454	0.467	3.16×10^{-11} *** (ep95)	95.6689
	(0.5408)	-	-	-	(0.3520)	-	-	(0.1694)	-	-	-	-
TAGB	-0.15438	1.45657	-	-	-	0.09596	-	-	0.4815	0.4988	6.98×10^{-7} *** (emax)	104.7472
	(0.81469)	(0.26546)	-	-	-	(0.06449)	-	-	-	-	-	-
TAGB	0.4883	-	-	-	-	0.04508	-0.13409	1.48403	0.5301	0.4748	0.000254 *** (ep95)	98.89634
	(0.60368)	-	-	-	-	(0.07045)	(0.33087)	(0.38335)	-	-	-	-

*The symbol *** means 'highly significant' at 95% confidence interval; ln(ep95) = log(95th height percentile); ln(ep40) = log(40th Height percentile); ln(hmax) =log(height maximum); elskew = height L-moment skewness; eaad = height average absolute deviation; el2 = Height L-moment (L2); ln(ep80) =log(80th Height percentile);The standard errors of the coefficient are given in paranthesis*

Table 4-10: Summary of the estimate of selection criteria for the best model

	Model	AIC	RMSE
1	ln(hmax)	105.02	0.83
2	el2	116.43	0.79
3	ln(ep40)	128.68	1.13
4	ln(ep80)	109.64	0.7
5	**ln(ep95)**	**95.36**	**0.68**
6	(E)AAD	131.49	0.82
7	elskew	140.08	1
8	ln(hmax)+ln(ep40)+ln(ep80)+ln(ep95)	99.36	0.73
9	ln(hmax)+el2+ln(ep40)+ln(ep80)+ln(ep95)	100.72	0.7
10	ln(hmax)+(E)AAD+elskew+(ep40)+ln(ep80)+ln(ep95)	97.76	0.75
11	(ep40)+n(ep95)	97.07	0.69
12	ln(hmax)+ln(ep95)	95.54	0.73
13	elskew+ln(ep95)	95.67	0.7
14	ln(hmax)+ln(ep40)	104.75	0.85
15	(ep40)+ln(ep80)+ln(ep95)	98.9	0.7

Note: *ln(ep95) = log(95th height percentile); ln(ep40) = log(40th Height percentile); ln(hmax) =log(height maximum); elskew = height L-moment skewness; eaad = height average absolute deviation; el2 = Height L-moment (L2); ln(ep80) =log(80th Height percentile);.*

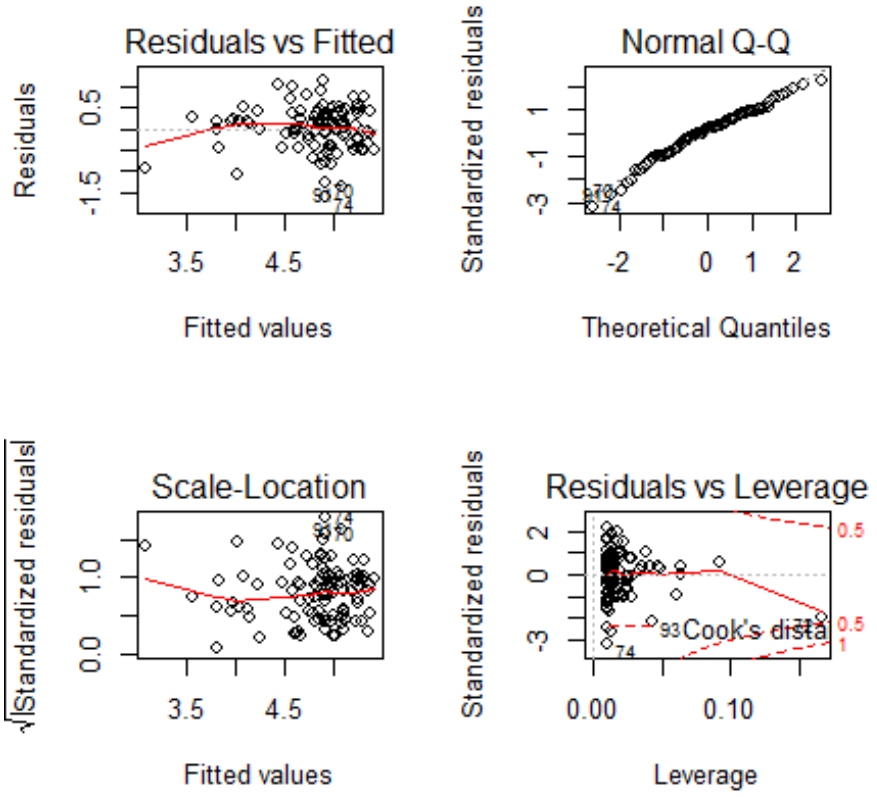


Figure 4-10: Regression model residuals for the 95th height percentile

4.9. Ordinary kriging (Modelling of TAGB from field sample points)

The variogram model was built from field data. A number of plausible models were tested to ascertain the best model relative to its parameter. Spherical model exhibited superior characteristics and was used in this study. The variogram model (spherical) is shown in figure 4-10 and the variogram parameters (nugget, sill and range) are in table 4-12. Subsequently, kriging interpolation was performed, and prediction map of biomass (shown in figure 4-12) and variance map (shown in figure 4-12) were produced. The residuals for ordinary kriging are presented on a histogram shown in figure 4-14. The variance map shows the uncertainty level of the prediction.

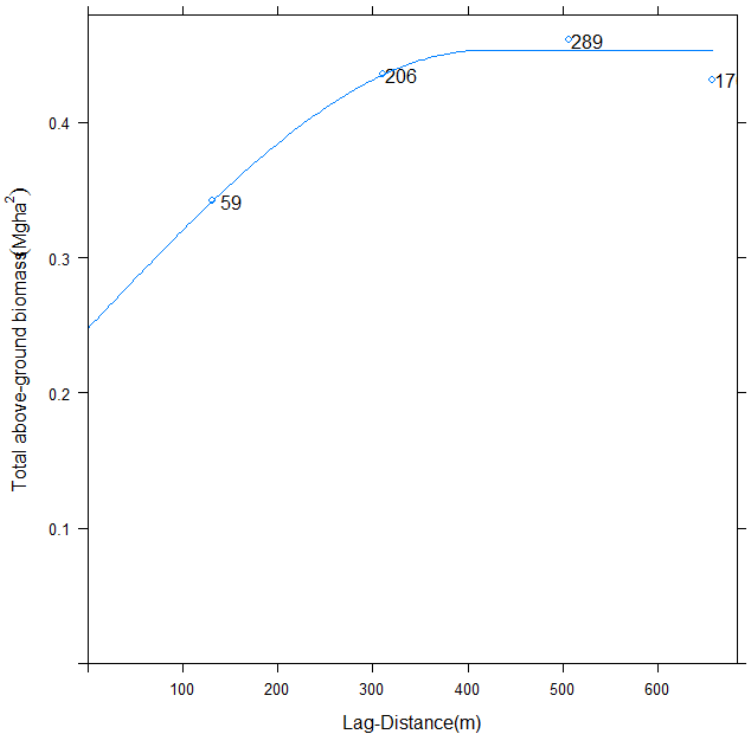


Figure 4-11: Variogram model (spherical) for ordinary kriging

Ordinary kriging predictions ranged from 3.84 log (Mg/ha) to 5.43 log (Mg/ha), with the highest predictions recorded in the northern part of the study area, as shown in figure 4-13. Ordinary kriging predictions were unbiased since the mean error was almost zero, as shown on table 4-13. Variance for the interpolations increased with increasing distance from the sampling locations, as shown in figure 4-13. From the variance map, the variance increased with increased distance from the sampling point while the variance was low around the points.

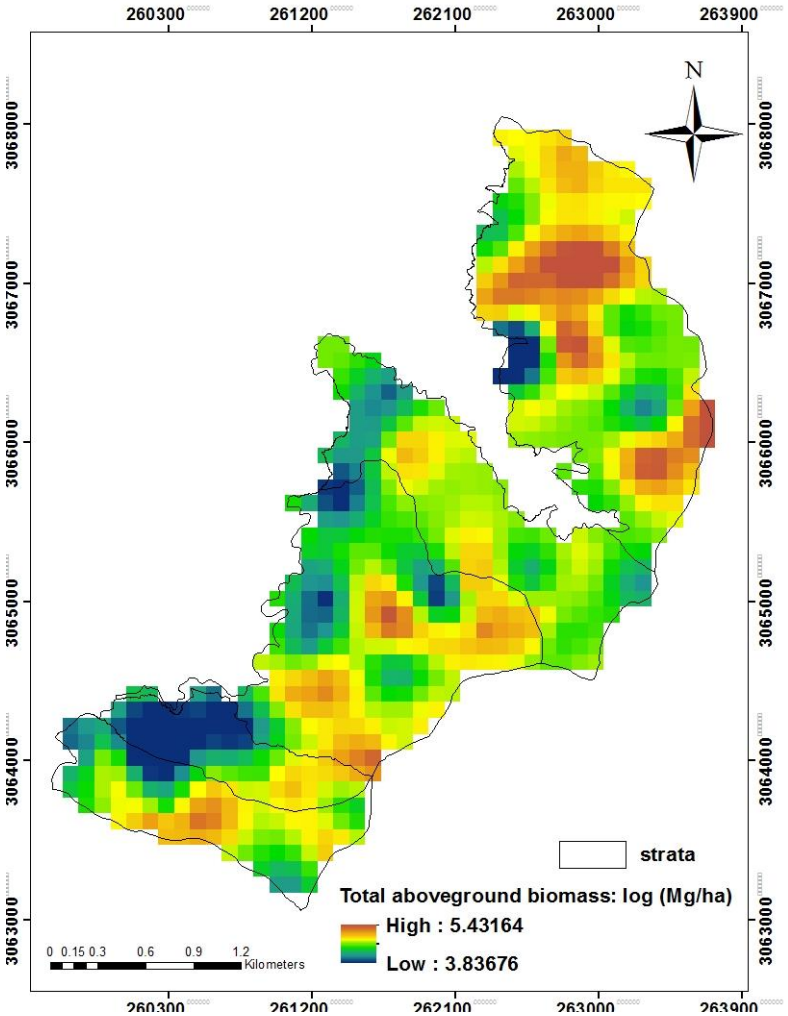


Figure 4-12: Ordinary kriging predictions for the study area

Table 4-11: Model parameter for ordinary kriging and residual variogram for log (95th height percentiles)

Method	Model	Partial sill	Range	Nugget
Ordinary variogram	Spherical	0.45	416	0.25
Residual variogram (ln(95th height percentiles)	Spherical	0.21	307.2	0.12

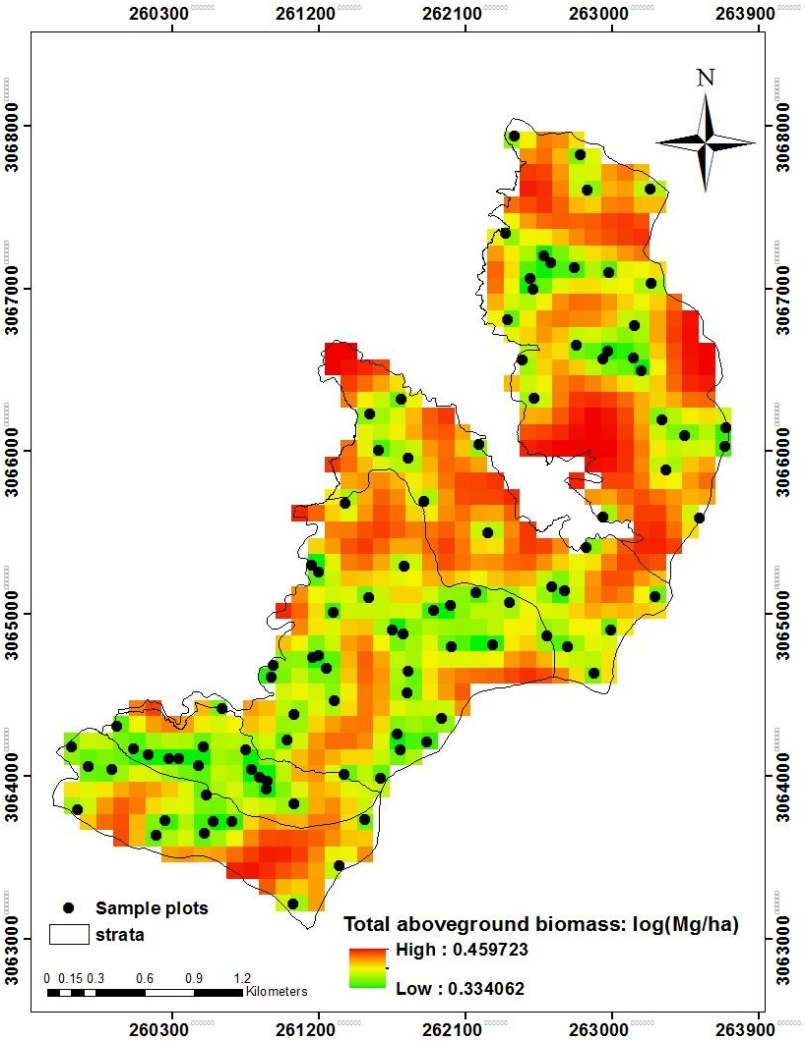


Figure 4-13: Ordinary kriging variances for the study area

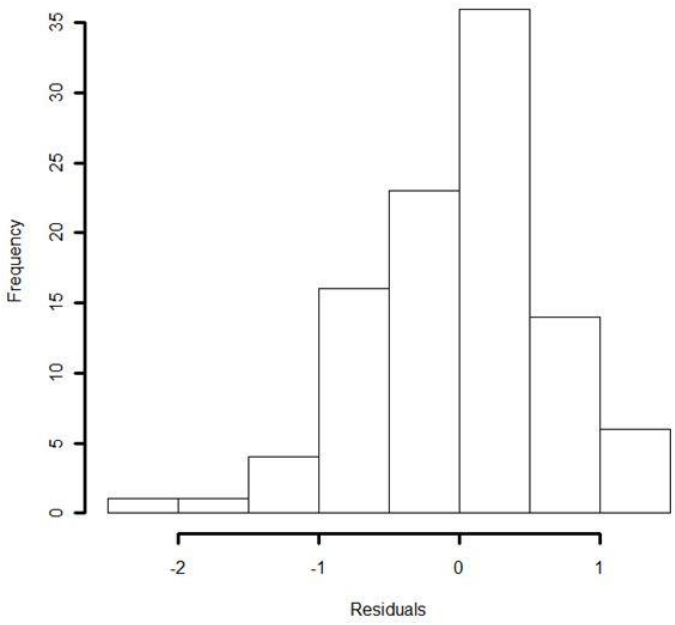


Figure 4-14: Ordinary kriging residuals

4.10. Regression kriging

The variogram analysis of the residual is presented in figure 4-16. From the set of potential models, spherical model was selected because it provided and demonstrated the best fit. The model parameter is displayed in figure 4-12. Prediction map modelled using the residual variogram is shown in figure 4-16, the variance map is shown in figure 4-17, and a histogram of the regression kriging residuals is shown in Figure 4-18, which indicates normality is satisfied

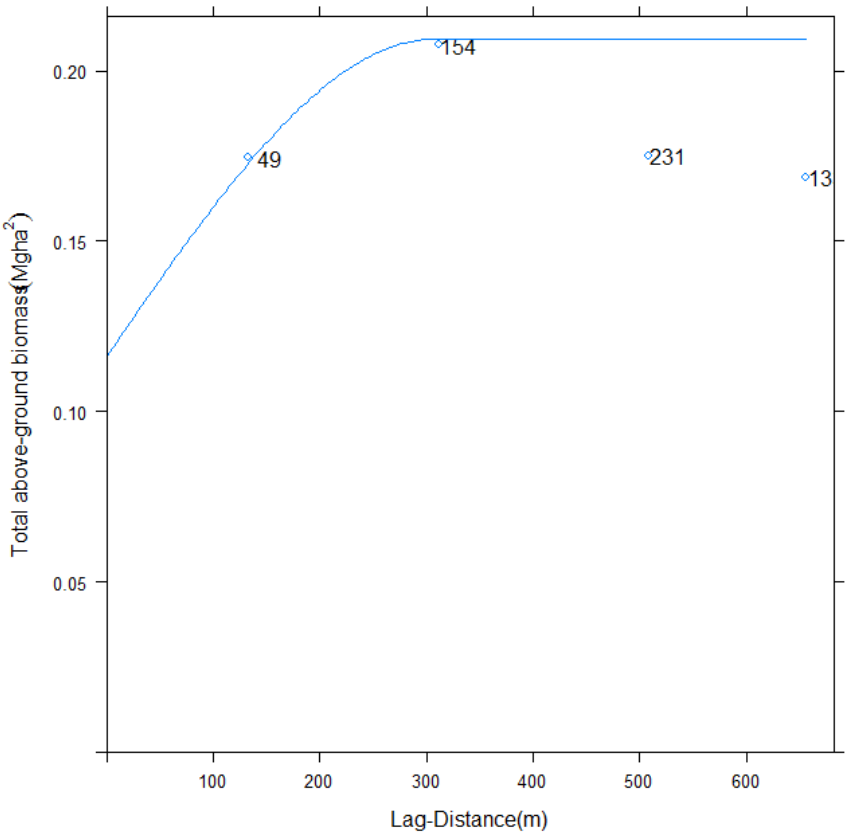


Figure 4-15: Residual variogram

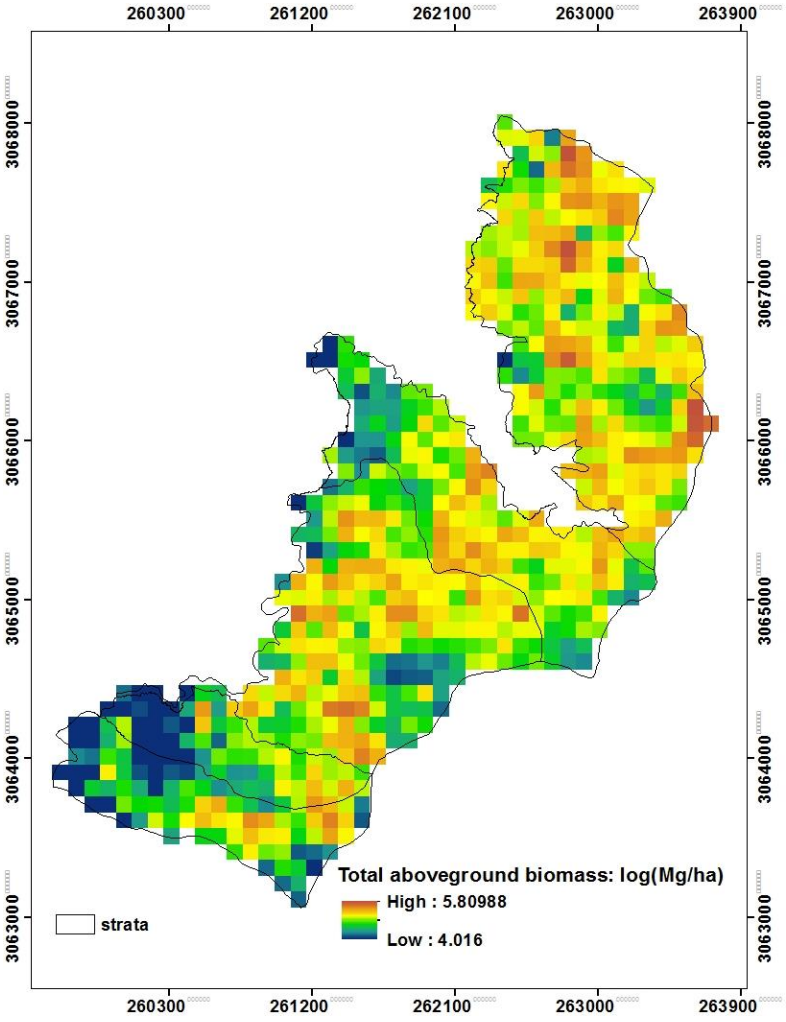


Figure 4-16: Regression kriging predictions

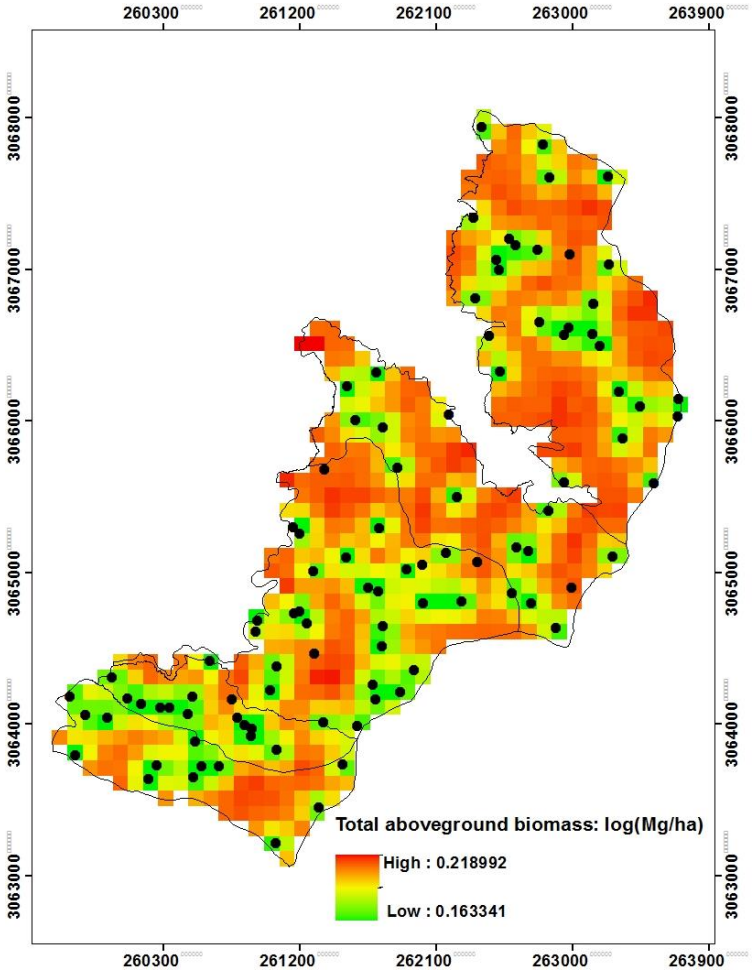


Figure 4-17: Regression kriging variances

Predictions for regression kriging ranged from 4.01 log (Mg/ha) to 5.81 log (Mg/ha) total above ground biomass as shown in figure 4-16. In comparison to ordinary kriging, the kriging variances were greatly reduced with the highest variance range being 0.21 log (Mg/ha). The regression kriging estimates were more accurate and unbiased, with mean error of 0.00023. The residuals of the model are approximately normally distributed, an assumption for meeting the requirements of regression.

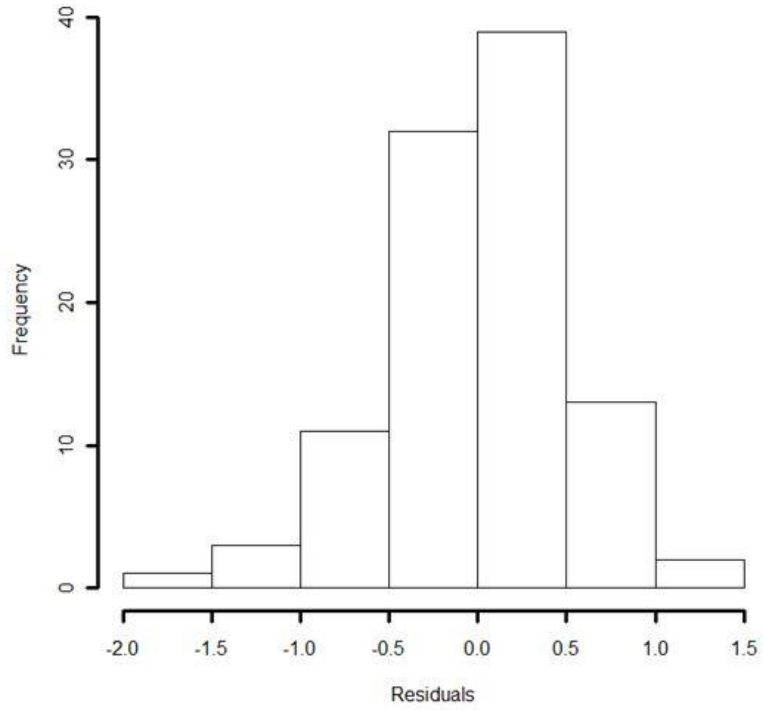


Figure 4-18: Regression kriging residuals

4.11. Validation of the models

With the 25% validation dataset for the regression analysis, above-ground biomass (AGB) calculated from field data (observed AGB) was compared with above-ground biomass predicted by the regression model (predicted AGB). Validation of the linear regression model between log transformed AGB and log transformed 95th height percentile yielded an R^2 of 0.53. The mean error and the root mean square error results are presented in table 4-13. The observed and predicted AGB were graphically visualized in a scatter plot in figure 4-19

Table 4-12: Validation statistics for the methods

Validation statistics	Regression	Ordinary kriging	Regression kriging
Mean Error (ME)	0.03	0.0025	0.00023
Root Mean Square Error (RMSE)	0.68	0.42	0.20

Regression kriging yielded a higher accuracy and more precise estimates of TAGB than ordinary kriging. The mean error and the root mean square error of regression kriging were the lowest in comparison to ordinary kriging and regression analysis. The summary of the statistics for the three methods is presented in table 4-13.

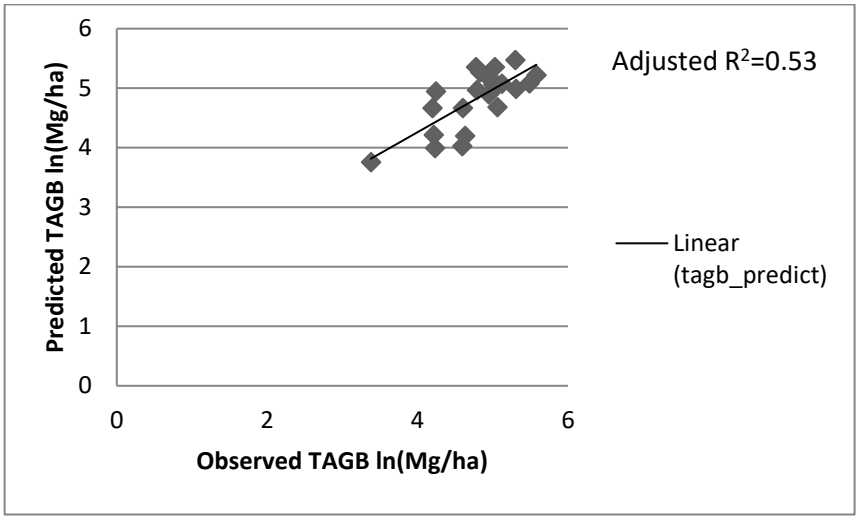


Figure 4-19: Model validation between observed and predicted biomass

4.12. Biomass map of the study area

The total above-ground biomass (TAGB) predicted from regression analysis, ordinary kriging and regression kriging is presented in table 4-16. In this study, regression kriging showed a more accurate estimate of TAGB as compared with regression analysis and gives a more reliable conclusion on the amount of TAGB in the study. The biomass map of the study as predicted by regression kriging is presented in figure 4-17. It shows the distribution of above-ground biomass across the strata. From the map, Devidhunga community forest showed distribution of high amounts of TAGB compared to other community forest. High amounts of biomass showed a trend of increase from the southern part of the study upwards toward the northern region.

Table 4-13: Total above ground biomass estimates with the methods

	Mean (LN (Mg/ha))	**Area (ha)**	**Total TAGB (LN (Mg))**
Ordinary kriging	4.83	764.67	3693.36
Regression kriging	5.19	764.67	3968.64
Regression analysis	5.21	764.67	3983.93

5. DISCUSSION

5.1. Extraction of LiDAR metrics

The forest structure is known to represent the physical and vertical arrangement of trees and the woody vegetation. The possibilities of LiDAR data include acquiring 3-dimensional assessment and measurement of forest structures. LiDAR measurement is characterized by multifaceted arrangements of different returns built together to portray vertical and canopy structure of trees, and the forest as a whole. With these multi-layered arrangements of LiDAR returns over tree and forest canopies, computation of several statistical variables (LiDAR metrics) describing the forest canopy structure becomes feasible (Li et al., 2008). Extraction of all potential metrics from LiDAR is most germane. This allows the investigation of all possible metrics and the determination of the most suitable explanatory variables that expounds the forest parameter(s) of interest. Some studies have focused and tested the use of specific LiDAR variables to estimate forest parameters (Lim., 2003)while others have generated several metrics to estimate forest parameters (Hall et al., 2005). The aim of this study was to select potential predictors from a large set of predictors and estimate TAGB in a tropical mountainous ecosystem, Kayarkhola watershed.

In this study, 48 LiDAR metrics were computed using elevation values, including several measures of canopy density calculations. This approach aligns with previous studies, for instance Hall et al., (2005), Næsset & Gobakken (2005), and Næsset (2002, 2004) extracted 39, 54, 44, and 46 LiDAR metrics (from elevation values and measures of canopy density) respectively. These large sets of metrics are mostly highly inter-correlated and are typically subjected to multi-collinearity diagnostics to select non-collinear variables for use in predicting the target forest parameter.

5.2. Relationship between LiDAR metrics and above-ground biomass

Intensity metrics are metrics computed using the intensity value of airborne light detection and ranging (LiDAR) data. Airborne laser scanning systems measures not only the distance from the sensor to the target but also the reflectance energy (intensity) from target objects on the earth surface, providing 3D information of the earth. Mostly, and as employed in this study, LiDAR have been widely explored with focus mainly on its additional Z-plane information, the distinctive element from other remote sensing devices.

In this study, it was observed that intensity metrics (metrics from intensity values of LiDAR) showed weak correlation ($0.008 \leq r \leq 0.206$) with TAGB. This is because intensity data of the

LiDAR data used in this study contains the initial raw reflectance values acquired by the sensor. It is known that LiDAR intensity data are affected by a host of factors during acquisition such as scanning angle, humidity, weather patterns, background reflections, topography (Goerndt et al., 2010; Yan et al., 2012). These factors account for noise in the data and imply the return signals associated with each return are not true representation of the area of study. Studies have shown that for intensity data to be beneficial for predicting forest attributes, the reflectance values of each return should be geometrically calibrated and undergo radiometric correction processes (Yan et al., 2012). The LiDAR dataset employed in this study is known to have not been calibrated and corrected as appropriate and may partly account for the weak correlation between intensity metrics and TAGB.

In the study by Goerndt et al.,(2010), normalized intensity data was employed to estimate forest attributes like basal area, tree density, lorey height, etc. The study deduced that the use of normalized intensity metrics improved the prediction of some forest attribute like lorey height. Intensity metrics generated in this study was therefore not useful for this study. Hence, intensity metrics were excluded from further regression analysis. On the other hand, nowadays, research has sprung up to explore the viability of un-calibrated raw LiDAR reflectance measurement. It has proven beneficial for tree species differentiation and land cover classification. For example, Kim et al., (2009) analysed LiDAR intensity data acquired in different seasons (leaf on and leaf off period) to classify tree species. Likewise, Ørka et al.,(2009) explored intensity and structural metrics to discriminate between coniferous and deciduous tree species.

5.3. Principal component analysis (PCA)

PCA was appropriate for these dataset because there were a large number of predictors; most of the predictors (LiDAR metrics) were highly correlated with each other. From the analysis, the first 3 component (mean height, height average &absolute deviation and canopy cover) explained the greater proportion of the variance. The 1st component (mean height) carried the bulk of the variance as shown in figure 4-9. However, in order to ensure that important variables were not lost, the strength of the relationship of each component with the response, was tested (Quinn & Keough, 2002). Seven variables (maximum height, height average absolute deviation, height L-moment (L2), Height L-moment skewness, 40th, 80th, and 95th height percentile) were found to be significant and independent variables from the component**.** It deduced that the first 3 component, with the greatest proportion of variables, were not significant with the response (TAGB).

PCA is a linear combination of all tested metrics and as such could yield some variables that cannot be biologically deduced to the predicted variable of interest (Li et al., 2008). Examples of such variables are height L-moment (L2) and height L-moment skewness. Some of the nominated variables from this study have been applied for TAGB and stem volume estimate in previous studies (Hyyppa et al., 2008). For example, in the study of Kane et al., (2010a), 95th height percentile was one of the variables selected from a PCA to measure canopy structure.

5.4. Regression modeling of TAGB and LiDAR metrics

Numerous plausible models could be developed from the PCA selected variables. However, fifteen linear models were developed in order to achieve parsimonious models that could be easily interpreted. The PCA result showed maximum height to be the only highly significant (α = 0.001) variable. This is not surprising because height of trees has been found to have close relationship with their biomass. As a result of this relationship, height is been used to boost allometric equations for the estimation of aboveground biomass (Chave et al., 2005).

For individual tree-based approach of estimating forest attributes, height has also been used as an additional and useful predictor in LiDAR based regression models. With this high significance of maximum height, it was expected that the linear model of maximum height and TAGB would be able to predict TAGB adequately with a high R^2 but this was not the case. Height maximum yielded an R^2 of 0.47.

For plot-based approach, maximum height has a propensity to be less stable because of its high sensitivity to outliers and might not be a good predictor for estimating forest biophysical properties (Kane et al., 2010a; Næsset & Gobakken, 2005). In a study done by Lim et al., (2003) in various stands of an hardwood forest using LiDAR metrics, height maximum was determined to be a good predictor in a closed canopy forest. But for the study area, a forest with irregularly-layered canopies and characterized with presence of some dominant trees having heights higher than their surrounding trees, height maximum will not be a suitable predictor. This is because height maximum will not be a measure of the total canopy area of the plot but of the single dominant tree. The forests in the study are characterized by irregularly-layered canopies and due to the sparsely dense nature of the data; height maximum could not be shown to be a good predictor candidate.

The linear model of each LiDAR metrics and TAGB tested as highly significant except the model of height L moment skewness which tested as moderately significant and had a very weak R^2 of 0.12 (table 4-10). Of all these, model of TAGB and 95th height percentile yielded the highest

coefficient of determination (R^2) of 0.54. Multiple linear regression of all the predictor was shown not to be significant, although, it yielded a higher R^2 of 0.56. This means that one or two variables would be sufficient to yield a significant model. Inclusion of additional variables rendered the model insignificant. Other models having two or more predictors either resulted in an insignificant model or had only 95th height percentile as significant. These models of two or more predictors all had R^2 greater than 0.50, except the model of height maximum and 40th height percentile which had an R^2 of 0.48. Generally, the addition of extra variables in the multiple linear regressions did not improve the model's R^2 and yielded an insignificant model or model with only one variable been significant. The outcome of the regression modelling shows that 95th height percentile alone can predict TAGB without the necessity of extra predictors. The best regression model was selected on the basis of the homoscedascity of the model residuals, highest coefficient of determination (R^2), lowest Akaike Information Criterion (AIC), reduced standard error estimate, and validation statistics (RMSE and ME) estimates.

5.5. Relationship between TAGB and LiDAR metrics

LiDAR metrics gives a statistical description of forest canopy structures that can be related with field measurement of forest attributes (Li et al., 2008). In general, LiDAR metrics generated had a moderate relationship with TAGB in this study. The regression model of 95th height percentile and TAGB yielded a R2 of 0.54 in this study. In comparison with the few studies in tropical forests: Drake et al., (2002) estimated tropical forest aboveground biomass with a large footprint waveform scanner (Laser Vegetation Imaging Sensor (LVIS)) in Costa Rica and attained an R2 of 0.93. The study done by Drake et al., (2002) is not comparable to this study because he utilised a waveform LiDAR while this study used a small footprint laser scanner. Waveform laser scanners can measure and retrieve ground heights consistently in forests with highly dense canopies. Capturing forest canopy structure accurately is therefore enhanced with waveform scanner as compared to a small footprint laser scanner where the probability of capturing forest canopy structure is low. In another study carried out by Drake et al., (2003) in a tropical wet forest in Costa Rica and tropical moist forest in Panama, using LVIS, he found LiDAR metrics to be highly correlated with forest parameters such as mean stem diameter, basal area and aboveground biomass with an R^2 range of 0.65-0.92. Also, biomass change quantification with LVIS in Costa Rica and attained an R^2 of 0.65 (Dubayah et al., 2010).

Following series of studies in Costa Rica tropical rainforest landscape that showed impressive results with LVIS large footed laser data, Clark et al., (2011) investigated the use of small footprint laser data in the region. With laser point density of 6 points/m2, he observed a strong relationship between plot level metrics and aboveground biomass. A more recent study was done

by Kronseder et al., (2012) in Central Kalimantan, Indonesia, a region similar to the study of interest. Here, the author also used a full waveform LiDAR data to assess aboveground biomass estimation in a lowland dipterocarp forest in the region. The best model of 45th & 65th height percentiles and standard error of the mean explained 83% of the variation in the forest. From the review done above, only the study of Clark et al., (2011) is comparable to the work done in this study where the study used small footprint laser data. His result shows a stronger relationship with TAGB and this can be attributed to a higher point density.

Unlike in the tropics, numerous studies have been done in temperate and boreal region (Canadian, Norwegian, Swedish, and Finnish forest) using canopy height metrics with plot-based approach and have reported strong relationship with forest variables (Lefsky et al., 1999; Næsset, 2002). A review of laser scanning experience in this region can be found in (Hyyppa et al., 2008; Naesset et al., 2004). A few of these studies are: Lefsky et al., (1999) achieved 90% and 75% of the variability of aboveground biomass and leaf area index respectively in a Douglar-Fir forest. Means et al., (1999) estimated height, basal area, total biomass and leaf biomass with a large-footprint LiDAR device and achieved R^2 of 0.95, 0.96, 0.96, 0.84 respectively. Majority of this LiDAR studies were conducted in temperate regions like Canadian, Norwegian, Swedish, and Finnish forest (Hyyppa et al., 2008; Naesset et al., 2004). So there is no basis for comparison with work done in these regions because of characteristics differences in the forest type for example species type and composition, and climatic condition. Aside differences between tropical forest and temperate forest, other factors are also prevalence that can cause differences in the relationship between LiDAR metrics and TAGB within inter-region (i.e. temperate – temperate or tropical – tropical). Factors include LiDAR data density, laser scanner parameter used in data collection, allometric equation, and canopy structure.

LiDAR data density refers to the number of pulses reflected from the object of interest per unit square metres. In other words, LiDAR point density captures the tree's vertical structure. Figure 4-8 shows an example cross-sectional view of trees in one of the plot in the study. Due to the sparsely dense nature of the LiDAR data in this study (0.8 points/m2), tree structure was not fully captured (figure 3-2), even though this is the foundation for LiDAR data relationship with aboveground biomass. LiDAR density of about 1 point per square metres has been successfully used to estimate forest parameters in the temperate region; no such study has been carried out for tropical forest aside the one carried out by this study.

One distinctive difference between temperate region and tropical region is in their species type and composition. In the study of (Næsset, 2002, 2004), their study comprise mainly of two

coniferous species (Norway spruce and Scots pine) while in a typical tropical forest, like in this study, species composition could be greater than 50. The reflective property from a forest stand with huge diversity of tree species is different from a stand with few species. In Naesset et al.,(2004), it was confirmed that a stand with a mixture of deciduous and coniferous trees may degrade prediction precision and produce biased estimate especially when there is a larger proportion of deciduous trees. Besides tree shape, complex canopy structure of tropical forest coupled with variable tree height within plots could explain the difference in the relationship observed. To buttress this point, Clark et al.,(2011) with a point density of 6 m^{-2} observed an improved estimate of R^2 from 0.90 to 0.96 from a natural forest with complex canopy and tree height variation to a plantation with an even aged structure with less variation.

In line with the point density, with a higher laser flying altitudes, sparsely dense data are acquired and underestimation of tree height is a great possibility. The study of Yu et al., (2004) tested three flight altitude (400 m, 800 m, and 1500 m) to assess their effect on tree height estimates. There was a decrease in the accuracy of tree height as flight altitude increased. In our study, a flight height of 2200 m was used which provides an explanation to the low density data acquired and moderately strong relationship with aboveground biomass due to low frequency of returns per crown (Hyyppa et al., 2008).

Allometric model used in a study can also affect the relationship of LiDAR metrics and aboveground biomass. The study of Drake et al., (2003) assessed the relationship between LiDAR metrics and allometrically estimated aboveground biomass between two tropical region (tropical wet and tropical moist forest). The study showed a significant difference between the two zones of which he attributed this to the different allometric model applied. His study has also answered the question of whether LiDAR metrics relationship can be generalized. The result from his study also shows that LiDAR relationship with aboveground biomass can be different even in similar regions (i.e. tropical to tropical). Differences in the relationship experienced in his study

LiDAR metrics generated in this study can be grouped into 3 classes in order to assess the strength of relationship between LiDAR metrics and TAGB. These include height distribution measures of central tendency (for example, height mean, maximum, standard deviation, the L-moments, skewness and kurtosis), height percentiles and measures of canopy density.

Height distribution measures of central tendency are statistical estimates around the mean. They are known to be highly susceptible outlier values. LiDAR data available for this study was shown

to have outliers (fig 3-2). Outliers have the tendency of displaying an abnormal sequence of values. Though efforts were made to eliminate them (figure 3-8), total removal is realistically not possible. This possibly will explain the reason why their LiDAR metrics relationship with TAGB was somewhat weak.

Height percentiles represent the distribution of laser first return from the low end of the division (0th) to the high end (100th). Studies have shown that height percentiles are strong predictor of forest biophysical properties because of its direct connection to the vertical structure of trees (Goerndt et al., 2010; Kane et al., 2010a; Means et al., 1999; Næsset, 2002, 2011). So, it is not unexpected that most of the height percentiles had a better relationship with TAGB compared to other metrics. 95th height percentile was shown to be a relatively strong predictor of TAGB in this study. Height percentiles within the range of 90 and 95 have found its application in forest stand parameter estimation including above-ground biomass (Drake et al., 2002; Kane et al., 2010a; Lim & Treitz, 2004).

The result of this study is comparable to previous studies that found 95th height percentile as a significant variable either singly or in combination with other metrics. For example, Kane et al., (2010a) found out that inclusion of 95th height percentile in the regression model with two other metrics (rumple and canopy density) improved the model as compared to the combination of other metrics. The 95th height percentile represents the distribution at the high end region of the plot canopy structure area. It is almost akin to maximum height, but 95th height percentile is less sensitive to anomalous high points than maximum height (Næsset, 2002). Basing on this, Næsset & Gobakken, (2008) excluded maximum canopy height region from the computations of measures of canopy density. Rather, the vertical range between the lower canopy limit (2 m above ground) and the 95th percentile of the canopy height was used. Likewise, Kane et al., (2010a) extracted 95th height percentile of first returns in place of maximum height for the assessment of stand structural complexity. On the contrary, in the study of Clark et al., (2011), with a point density of 6 points m^{-2}, the best model for the estimation of a tropical rain forest aboveground biomass was the model of mean height and maximum (R^2 of 0.90). The high density data could have annulled the effect of instability at the maximum height.

Canopy density metrics have found efficacy in the estimation of forest parameter (Maltamo et al., 2006). The rationale is because of their close connection with forest structural characteristics. In this study, canopy density metrics had weak relationship with AGB. It could be attributed of the low data density of the LiDAR data. The LiDAR returns were sparse such that they were not representative of the true canopy structure in the plot area (4-8). This could forms the rationale

for the poor relationship of the canopy density metrics with AGB. Lim & Treitz (2004) explained that forest canopy cover and laser point density may affect the use of canopy density metrics for the characterization of forest structural parameter and nullify already established facts about the relationship between vertical distribution of LiDAR returns and foliage area of trees and then the forest. The poor relationship exhibited by the canopy density metrics could also be attributed to the complex and heterogeneous nature of the forest. Heterogeneous forest like it is the case of the study area (Chitwan forest, Nepal), embodies assortments of tree species of various shapes and height. This variation in tree species type and height could also have accounted for the poor relationship between canopy metrics, representing the canopy are of the forest and AGB. This might not have been the case in a homogenous forest or a plantation that constitutes even aged trees. Metrics that have been found to be applicable for the estimation of forest biophysical properties are peculiar to different geographic locations and climatic conditions. In other words, it implies that metrics applicable in a particular geographic location cannot be generalized to other areas. Also, it might not be found as relevant in another area that share common climatic similarities with the area due to other factors that could constitute to the difference (Drake et al., 2003). This explains the reason why different studies present different metrics as being related to AGB.

5.6. Regression kriging and Regression analysis

The results of regression kriging using 95th height percentile as an explanatory variable gave a standard error of the mean smaller than the regression approach. This is caused by the ability of the kriging algorithm to account for some of the variability in total above ground biomass. Regression kriging uses both concepts of regression analysis combined with auxiliary information in the prediction of the independent variable (i.e., total above ground biomass) (Hengl et al., 2007). Therefore, regression kriging considers the long term variations (trends) as well as local variations, which makes it superior to the simple regression and ordinary kriging. Conversely, regression approach only makes use of the explanatory information without taking the local variations into consideration. Furthermore, high standard error in regression analysis derives from the moderately strong relationship of total above ground biomass with 95th height percentile.

The variogram model used for the predictions of TAGB had high nugget, implying that most of the variations occurring at small scales had not been captured. This originated from the design of the sampling (stratified random sampling) used. Accurate modelling of a target variable is dependent on sampling density and sample size (Webster et al., 2006). Random sampling usually misses most of the small scale variation which is the justification for the large unexplained

variation. However, in spite of the sampling plan used, the variogram model used (spherical model) was appropriate as the variances of kriging were small, an indication of little uncertainty associated with the TAGB predictions (McBratney et al., 1981).

The variogram of residuals used for regression kriging had a reduced nugget than the ordinary variogram. This is because some of the variation that could not be accounted for by the spatial dependencies of the target variable had been taken care of by the covariable and 95th height percentile. This is the reason for the reduced prediction variances in the regression kriging interpolations (figure 4-18). However, the nugget could not drill down much since 95th height percentile had a very weak correlation with TAGB. The 95th height percentile was obtained from the LIDAR metrics with very sparsely distributed tree density, a scenario that could have resulted in the moderate correlations of the LIDAR metrics with total above ground biomass.

5.7. Sources of error and uncertainty in biomass estimation

All the stages involved in biomass estimation forms potential sources of error and uncertainties to the overall result obtained. It ranges from field measurement processes, LiDAR acquisition, allometric computation of field aboveground biomass, etc. According to Phillips et al., (2000), knowledge of error estimation is important to understand the "level of uncertainty in projected carbon fluxes".

GPS error was a possibility in the field. The study made use of the Garmin GPS for navigation and plot location. The precision and accuracy of any handheld GPS equipment depends on the strength of signal it can receive at any point in time. Several factors such as dense canopy cover, weather conditions, and number of visible satellite to the equipment impacts its precision and accuracy. In a forested mountainous region like Nepal, dense canopy cover was inevitable. Some of the days on the field were cloudy and rainy. Most times, the number of satellite the GPS acquired was very minimal. All of these hindered sharp signals reception in the field resulting to positional error. Measurement error in the field was very possible and unavoidable considering the steep nature of the terrain of the study area. This include improper measurement of DBH, possibility of wrong recording of field data both in the field and during data compilation, inability to correctly locate the tree top because of the dense nature of the forest yielding incorrect height measurement, omission of some trees in the plot.

According to Gibbs et al., (2007), the use of LiDAR data for estimating TAGB with different approaches yields results with high level of accuracy and minimal uncertainties. Despite the improved accuracy that can be achieved with the use of LiDAR, it is very sensitive and prone to

errors. This includes error during LiDAR data acquisition, specification of sensor parameter, and LiDAR model processing (DTM, DSM and CHM). Studies have shown that LiDAR captured at a high flying altitude result to low point density and underestimation of tree heights and a high scan angle greater than 10° is susceptible to measurement errors (Evans et al., 2009). LiDAR for this study was acquired at a very high flying altitude (2200 m AGL) and a scan angle of 20 °. Presence of outliers in the data was seen (fig 3-2) and some portion were completely void of points (had no data) constituting errors in the data. All of these culminate to errors and uncertainties and ultimately affected biomass estimation.

The regression model was developed with logarithm transformed variables in order to satisfy regression and geo-statistical assumptions. Although, the process of transformation introduces bias in the calculation, in order to be able to compare the results of the three approaches, biomass estimates were reported in the transformed unit, rather than their initial original unit.

Choice of allometric equation is very critical to biomass estimation. Allometric equations are either species specific (for each specific species) or general (for group of species). The availability of species specific equation is practically impossible for tropical forests due to their rich diversity in tree species. This study however made use of a general allometric model (developed for moist tropical forest) for all species except for Shorea robusta. The problem associated with general models include a generalisation of biomass estimates across species resulting to an under estimation of biomass i.e. biased estimate. They are developed with the assumption that tree species accumulate biomass at the same rate, which is not the case. Besides this, accumulation of errors in allometric models commences from the initial phase of allometric model development which spans from choice and number of tree species used, diameter range, tree measurement, availability and inclusion of wood specific gravity.

Assessing and quantifying the magnitude of the different sources of error is certainly beyond the scope of this study. It is an already established fact that the use of generalised equation rather than species specific equation gives a bias estimate of aboveground biomass but like earlier mention above, this source of error is unavoidable in a tropical forest study. According to Chave et al., (2004), the most important source of error is associated with the choice of allometric equation. To demonstrate the impact of allometric equation to estimating aboveground biomass, he tested six different equations and obtained aboveground biomass range from 215 to 461 Mg ha-1 with a mean of 347 Mg ha-1 and standard error of 77 Mg ha-1. This range of result obtained illustrates the gravity of its impact in a biomass study.

5.8. Limitation of this study

The study was limited to 5 community forest out of a total of 16 due to time constraint and difficult terrain. The inclusion of other community forests would have provided a broader platform to analyse and compare results. Results obtained from this study however might not be typical of the other forests and cannot be generalized for the Kayarkhola watershed.

The allometric equation developed from DBH and height would have been a better choice for TAGB calculation in this study considering the fact that it would be used with a LiDAR-height regression model. Unfortunately, height of all trees could not be measured due to time constraint and again problems associated with tree height measurement in tropical forests. Inclusion of height parameter would have improved the correlation result. Hence the allometric model for the LiDAR study is slightly inappropriate.

6. CONCLUSION AND RECOMMENDATION

6.1. Conclusion

The main objective of this research was to derive various LiDAR metrics from low density LiDAR data and to estimate and map aboveground biomass using a plot-based approach in the tropical forest of Chitwan, Nepal. The study questions were:

1. How strong is the relationship between LiDAR metrics and field aboveground biomass in the study area?

LiDAR metrics generated in this study had a correlation coefficient range of $0.3 \leq r \geq 0.55$. Seven LiDAR metrics selected by the PCA all tested significant when regressed with aboveground biomass. Out of 15 linear regression models that were developed, 95th height percentile alone proved to be able to predict aboveground biomass without the necessity of extra predictors. This is because inclusion of additional predictors did not improve the model. The model of 95th height percentile and aboveground biomass showed a moderate relationship (R^2 = 0.54). Intensity metrics did not prove useful for this study because of their weak relationship with aboveground biomass. Hence, they were excluded from further regression analysis.

To the best of the knowledge of the author, no study has been done in a tropical forest with a density of < 1 point/m^2 using a plot-based approach. Although some studies have already been carried out in temperate and boreal forest with density of about 1 point/m^2, which achieved some good result, results from a temperate forest cannot be compared with those of the tropical forest. Any such comparison is inappropriate because of glaring differences in factors such as species composition, canopy structure and climatic condition, which all have an impact on LiDAR metrics relationship and aboveground biomass. In the study, a moderate relationship between the LiDAR metrics and aboveground biomass was observed in the study area.

2. Which of the methods (regression analysis, ordinary kriging, and regression kriging) gives an improved accuracy estimate? What are the accuracies of the different methods in estimating aboveground biomass?

The performance of the three approaches was assessed using their Root Mean Square Error (RMSE) values, and Mean Error (ME) of (table 4-13). The regression kriging showed an improvement in the accuracy prediction of aboveground biomass with a lowest RMSE of 0.20 and ME of 2.3 $\times 10^{-4}$. Regression analysis resulted in an RMSE of 0.68 and ME of 3×10^{-2}.

Regression kriging showed an improvement in the estimation because of its ability to account for some variations in aboveground biomass. In the study area, regression kriging gives an improved accuracy estimate.

3. What is the amount of aboveground biomass in the study area and how is it distributed using the most accurate approach?

Regression kriging was used to map the spatial distribution of aboveground biomass. The total aboveground estimate for the study was 3968.64Mg with the regression kriging approach. The biomass map revealed that areas with high to moderately high elevation (toward the north eastern part of the study) had higher distribution of biomass. Devidhunga community forest had the highest distribution of aboveground biomass when compared with the other community forests. This is attributed to the high elevation of the named community forest (figure 2-1).

6.2. Recommendation

This study has revealed a moderately high relationship between the best model of 95^{th} height percentile and aboveground biomass using a small footprint laser scanning data. However, TAGB can be estimated in a tropical region using a low density LiDAR data. The review of LiDAR studies carried out in tropical forest using large footprint waveform laser data and another study with a small footprint laser scanning data having a density of 6 points/m^2, showed impressive results compared to our work. Therefore, I recommend the use of either large footprint waveform laser data or a small footprint laser scanning data of density > than 5 points/m^2 using a plot-based approach for a tropical forest study in future study in the same watershed.

LiDAR studies in boreal and temperate region have demonstrated that there is no significant difference in estimation of forest attributes between different point densities using a plot-based approach. Based on literature reviews done under this study, studies of this kind are lacking in tropical forest regions. The result of the temperate regions cannot be generalised for the tropics until the hypothesis is tested. Furthermore, studies assessing the effect of different sensors and flying altitudes on forest canopy metrics and its biophysical properties have not been tested in tropical forests. Such studies are recommended in tropical forest and plantation region.

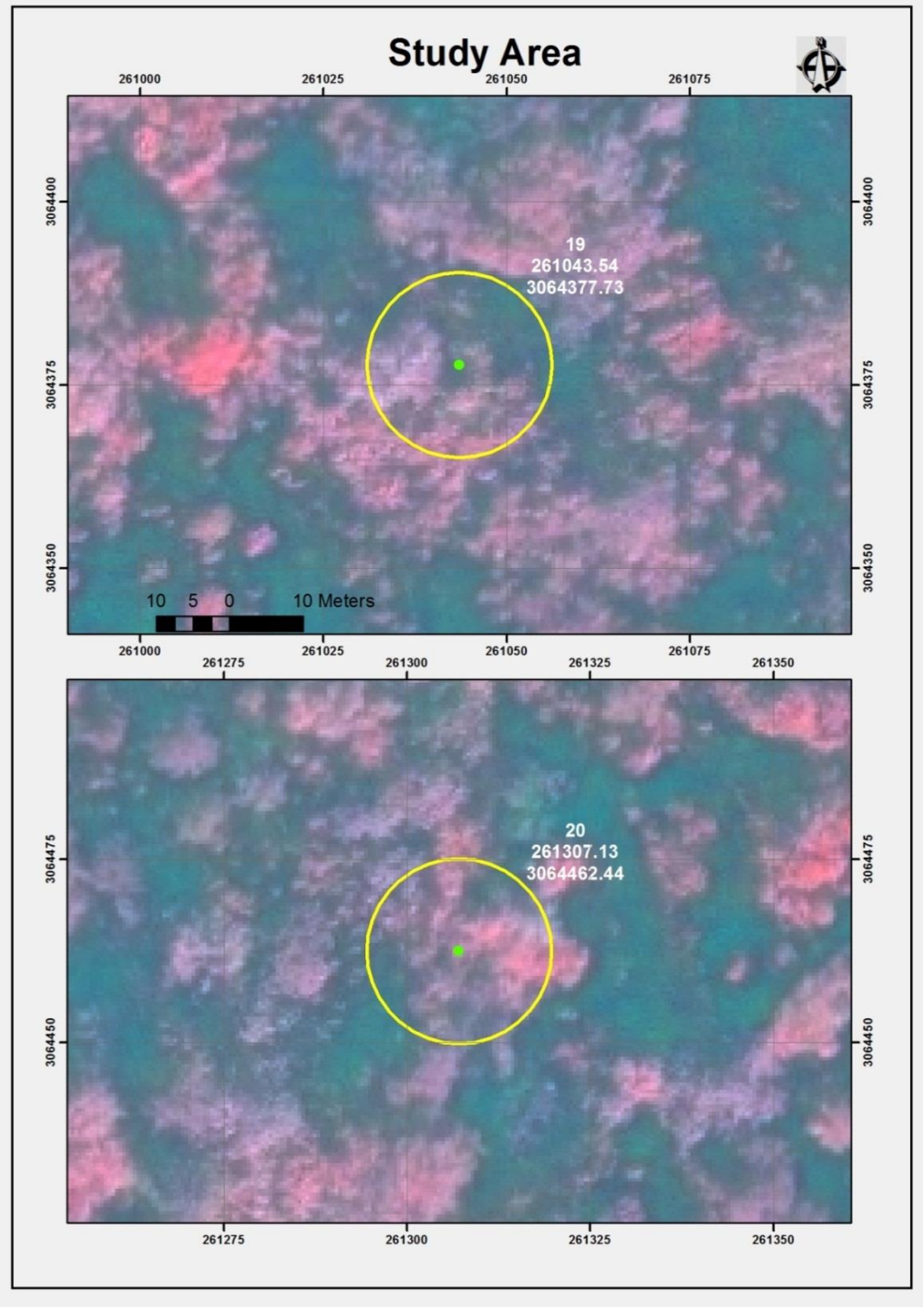
LiDAR remote sensing is known to have an edge over optical remote sensing including high resolution data for the estimation of forest biophysical properties because of its additional 3D characteristics. However, the comparison of these sources of data has rarely been done, especially

with a plot-based approach. It would be interesting to quantify the relative increase in accuracy and precision when LiDAR is employed. This study recommends the comparison between LiDAR data and high resolution data (for example Geo-EYE data) for the estimation of forest biophysical properties while adopting the same approach.

Furthermore, this study recommends the use of differential global positioning system (GPS) for forest navigation and plot identification in a tropical study rather than the use of Garmin GPS. With a differential GPS, more precise information on location would be obtained, positional error lowered and co-registration error of plots on the ground and the image can be considerably reduced.

7. APPENDICES

Appendix 1: Sample plot displayed on GeoEye-1 image showing the plot center and plot area

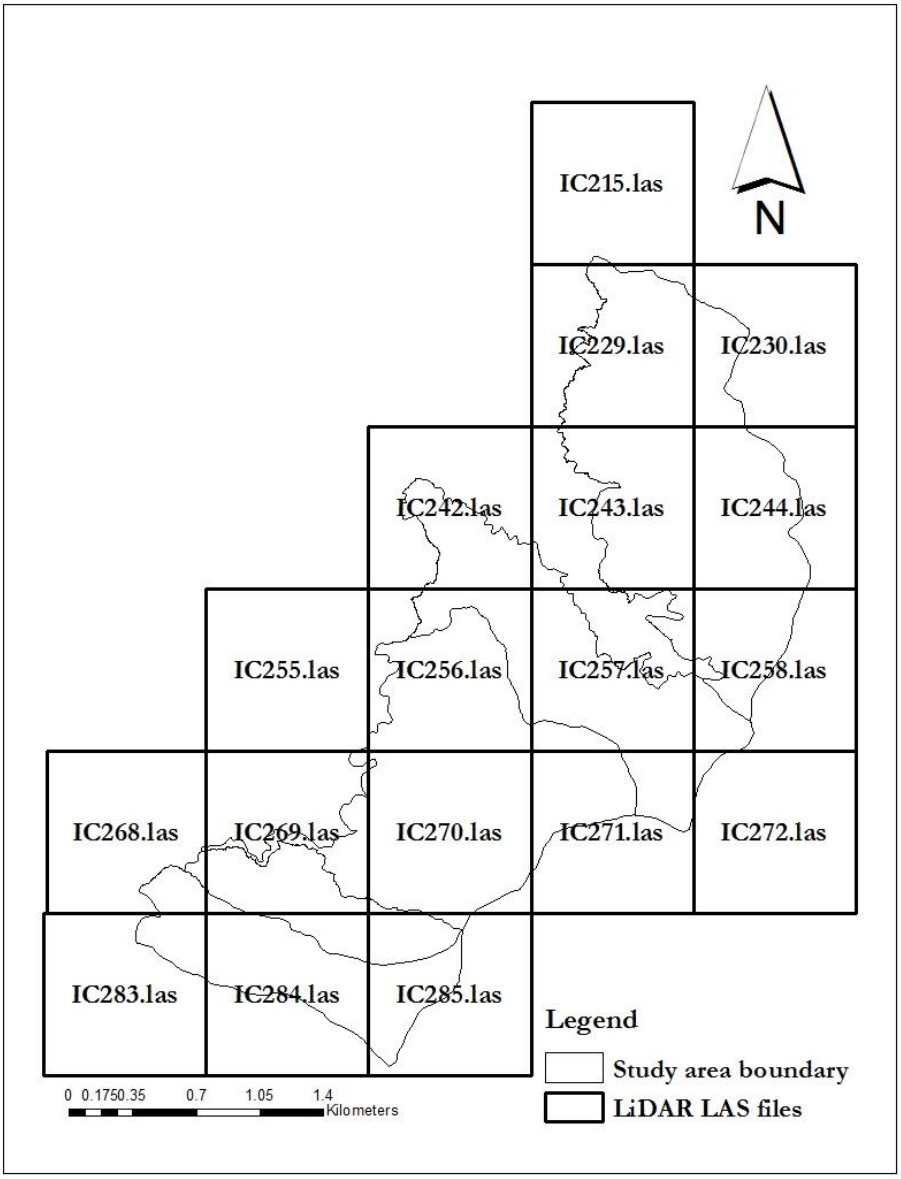


Appendix 2:

A subsection of LAS file tile-IC215 with its attributes

X	Y	Elevation	Intensity	Pulse Number	Return Number	Nadir Angle
262000.5486	3068062.368	243.769989	41	0	1	8
262000.3186	3068061.028	242.470001	44	0	1	8
262003.1586	3068062.508	244.25	26	0	1	8
262090.0786	3068062.388	258.350006	0	0	2	7
262089.9086	3068060.858	253.559998	16	0	3	7
262090.0586	3068061.798	253.610001	23	0	2	7

LiDAR LAS file tiles covering the study area



Appendix 3:

3-A: Data Sheet Template

DATA COLLECTION FORM FOR KAYERKHOLA WATERSHED, CHITWAN, NEPAL

Name of recorder..... **Date**..... **Plot radius size**.....12.62m.....

Sampling Plot No.	Coordinates		Elevation	Slope (%)	Aspect
	X	Y			

Management type:

Strata Name:

Forest type:		Crown cover (%)

Tree No.	Species	DBH (cm)	Height (m)	Crown diam.(m)	Remark	Tree No.	Species	DBH (cm)	Height (m)	Crown diam(m)	Remark
1						16					
2						17					
3						18					
4						19					
5						20					
6						21					
7						22					
8						23					
9						24					
10						25					
11						26					
12						27					
13						28					
14						29					
15						30					

Appendix 4:

4-A: Slope correction table

Plot size 500 m^2

Slope%	Radius(m)	Slope%	Radius(m)	Slope%	Radius(m)
0	12.62				
1	12.62	36	13.01	71	13.97
2	12.62	37	13.03	72	14.00
3	12.62	38	13.05	73	14.04
4	12.62	39	13.07	74	14.07
5	12.62	40	13.09	75	14.10
6	12.63	41	13.12	76	14.14
7	12.63	42	13.14	77	14.17
8	12.64	43	13.16	78	14.21
9	12.64	44	13.19	79	14.24
10	12.65	45	13.21	80	14.28
11	12.65	46	13.24	81	14.31
12	12.66	47	13.26	82	14.35
13	12.67	48	13.29	83	14.38
14	12.68	49	13.31	84	14.42
15	12.69	50	13.34	85	14.45
16	12.70	51	13.37	86	14.49
17	12.71	52	13.39	87	14.52
18	12.72	53	13.42	88	14.56
19	12.73	54	13.45	89	14.60
20	12.74	55	13.48	90	14.63
21	12.75	56	13.51	91	14.67
22	12.77	57	13.53	92	14.71
23	12.78	58	13.56	93	14.74
24	12.79	59	13.59	94	14.78
25	12.81	60	13.62	95	14.82
26	12.82	61	13.65	96	14.85
27	12.84	62	13.68	97	14.89
28	12.86	63	13.72	98	14.93
29	12.87	64	13.75	99	14.97
30	12.89	65	13.78	100	15.00
31	12.91	66	13.81	101	15.04
32	12.93	67	13.84	102	15.08
33	12.95	68	13.87	103	15.12

Appendix 5:

(1) Summary report of the LiDAR LAS tiles in the study area

LiDAR LAS tiles	Min X	Min Y	Min Z	Max X	Max Y	Max Z	Total Returns
IC215.las	262000	3068000	241.34	263000	3068999	2272.32	1,070,831
IC229.las	262000	3067000	241.69	263000	3068000	2322.64	1,457,680
IC230.las	263000	3067000	418.18	263999	3068000	2423.04	2,194,170
IC242.las	261000	3066000	217.75	261999	3067000	2260.48	1,133,956
IC243.las	262000	3066000	249.34	262999	3067000	2287.54	1,294,545
IC244.las	263000	3066000	368.78	264000	3067000	2442.67	1,596,626
IC255.las	260000	3065000	197.15	261000	3066000	2213.18	734,378
IC256.las	261000	3065000	211.83	261999	3066000	2392.96	1,110,014
IC257.las	262000	3065000	272.08	262999	3066000	2476.85	1,156,139
IC258.las	263000	3065000	331.11	264000	3066000	2570.62	1,434,601
IC268.las	259019	3064000	183.87	259999	3064999	2388.55	742,106
IC269.las	260000	3064000	192.84	261000	3064999	2507.22	985,309
IC270.las	261000	3064000	231.79	262000	3065000	2491.26	1,157,715
IC271.las	262000	3064000	363.95	263000	3065000	2410.53	1,747,671
IC272.las	263000	3064000	261.55	263999	3065000	1423.60	1,113,171
IC283.las	259000	3063000	174.77	260000	3063999	2179.69	735,207
IC284.las	260000	3063000	206.58	261000	3064000	2246.43	1,446,579
IC285.las	261000	3063000	280.75	262000	3063999	2386.57	1,346,687
Overall							**22,457,385**

Appendix 6: Result of Correlation Matrix

corelation matrix result-ln transformed

Appendix 7: Result of Correlation Matrix

Variables	Variable Name	Estimate	Standard error	T-value	P(α = 0.05)
(Intercept)	(Intercept)	1.47E+02	6.99E+00	20.973	<0.0000000000000002***
enx1	emax	-1.15E+01	1.69E+00	-6.826	0.00000000231***
enx2	emean	-1.44E+01	8.52E+00	-1.685	0.0964.
enx3	emode	-1.71E+00	5.51E+00	-0.31	0.7573
enx4	estdv	2.22E+01	2.21E+01	1.009	0.3166
enx5	evar	-2.77E+01	1.58E+01	-1.753	0.0838.
enx6	eskew	-1.30E+02	7.08E+01	-1.838	0.0702.
enx7	eaad	-2.39E+02	1.13E+02	-2.117	0.0377*
enx8	el1	9.01E+01	6.42E+01	1.403	0.1649
enx9	el2	2.80E+02	1.39E+02	2.008	0.0484*
enx10	el3	-4.12E+02	2.09E+02	-1.97	0.0526.
enx11	elskew	-3.19E+02	1.50E+02	-2.132	0.0364*
enx12	ep10	1.32E+03	6.73E+02	1.958	0.0541.
enx13	ep20	3.57E+02	2.21E+02	1.612	0.1113
enx14	ep25	-3.06E+03	1.58E+03	-1.939	0.0564.
enx15	ep30	5.17E+02	2.72E+02	1.899	0.0615.
enx16	ep40	-1.38E+03	6.91E+02	-1.997	0.0496*
enx17	ep50	3.23E+03	1.67E+03	1.93	0.0575.
enx18	ep60	-6.89E+02	4.19E+02	-1.642	0.1049
enx19	ep70	-4.77E+02	2.73E+02	-1.747	0.0848.
enx20	ep75	9.57E+03	4.92E+03	1.944	0.0558.
enx21	ep80	-2.07E+03	1.00E+03	-2.072	0.0418*
enx22	ep90	-2.26E+04	1.16E+04	-1.942	0.056.
enx23	ep95	4.79E+03	2.37E+03	2.021	0.047*
enx24	ep99	-2.70E+04	1.39E+04	-1.948	0.0553.
enx25	pfra	7.50E+04	3.83E+04	1.959	0.054.
enx26	para	1.02E+05	5.21E+04	1.95	0.0551.
enx27	ara_tfrp	2.20E+05	1.13E+05	1.959	0.054.
enx28	aramean_tfrp	-1.85E+17	9.49E+16	-1.953	0.0547.

REFERENCES

Angelo, J., Duncan, B., & Weishampel, J. (2010). Using Lidar-Derived Vegetation Profiles to Predict Time since Fire in an Oak Scrub Landscape in East-Central Florida. *Remote Sensing, 2*(2), 514-525.

Angelsen, A., Brockhaus, M., Kanninen, M., Sills, E., Sunderlin, W. D., & Wertz-Kanounnikoff, S. (2009). *Realising REDD+: national strategy and policy options*. Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Arroyo, L. A., Johansen, K., Armston, J., & Phinn, S. (2010). Integration of LiDAR and QuickBird imagery for mapping riparian biophysical parameters and land cover types in Australian tropical savannas. *Forest Ecology and Management, 259*(3), 598-606. doi: 10.1016/j.foreco.2009.11.018

Asner, G. P., Powell, G. V. N., Mascaro, J., Knapp, D. E., Clark, J. K., Jacobson, J., . . . Hughes, R. F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences, 107*(38), 16738-16742. doi: 10.1073/pnas.1004875107

Baker, D. J., Richards, G., Grainger, A., Gonzalez, P., Brown, S., DeFries, R., . . . Stolle, F. (2010). Achieving forest carbon information with higher certainty: A five-part plan. *Environmental Science & Policy, 13*(3), 249-260. doi: DOI: 10.1016/j.envsci.2010.03.004

Basuki, T. M., van Laake, P. E., Skidmore, A. K., & Hussin, Y. A. (2009). Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. *Forest Ecology and Management, 257*(8), 1684-1694. doi: 10.1016/j.foreco.2009.01.027

Beets, P. N., Reutebuch, S., Kimberley, M. O., Oliver, G. R., Pearce, S. H., & McGaughey, R. J. (2011). Leaf Area Index, Biomass Carbon and Growth Rate of Radiata Pine Genetic Types and Relationships with LiDAR. *Forests, 2*(3), 637-659.

Brown, S. (1993). Tropical forests and the global carbon cycle: the need for sustainable land-use patterns. *Agriculture, Ecosystems & Environment, 46*(1-4), 31-44. doi: 10.1016/0167-8809(93)90011-d

Brown, S. (1997). Estimating biomass and biomass change of tropical forests: a primer *FAO Forestry Paper 134*. Rome, Italy.

Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution, 116*(3), 363-372. doi: Doi: 10.1016/s0269-7491(01)00212-3

Brown, S., & Gaston, G. (1995). Use of forest inventories and geographic information systems to estimate biomass density of tropical forests: Application to tropical Africa. *Environmental Monitoring and Assessment, 38*(2), 157-168. doi: 10.1007/bf00546760

Brown, S., Gillespie, A. J. R., & Lugo, A. E. (1989). Biomass estimation methods for tropical forests with applications to forest inventory data. *Forest Science, 35*(4), 881-902.

Canadell, J. G., & Raupach, M. R. (2008). Managing forests for climate change mitigation. *Science, 320*(5882), 1456-1457. doi: 10.1126/science.1155458

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., . . . Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia, 145*(1), 87-99. doi: 10.1007/s00442-005-0100-x

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1443), 409-420. doi: 10.1098/rstb.2003.1425

Clark, M. L., Roberts, D. A., Ewel, J. J., & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2010.08.029

Dhital, N. (2009). Reducing Emissions from Deforestation and Forest Degradation (REDD) in Nepal: Exploring the Possibilities. *Journal of Forest and Livelihood, 8*(1).

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., . . . Prince, S. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment, 79*(2-3), 305-319. doi: 10.1016/s0034-4257(01)00281-4

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography, 12*(2), 147-159. doi: 10.1046/j.1466-822X.2003.00010.x

Dubayah, R. O., & Drake, J. B. (2000). Lidar Remote Sensing for Forestry. *Journal of Forestry, 98*(6), 44-46.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res., 115*, G00E09. doi: 10.1029/2009jg000933

Englhart, S., Keuck, V., & Siegert, F. (2011). Aboveground biomass retrieval in tropical forests - The potential of combined X- and L-band SAR data use. *Remote Sensing of Environment, 115*(5), 1260-1271. doi: 10.1016/j.rse.2011.01.008

Evans, J., Hudak, A., Faux, R., & Smith, A. M. (2009). Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables. *Remote Sensing, 1*(4), 776-794.

FAO. (2010). Global Forest Resource Assessment: Main report: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, Rome 2010.

Fuchs, H., Magdon, P., Kleinn, C., & Flessa, H. (2009). Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment, 113*(3), 518-531. doi: 10.1016/j.rse.2008.07.017

Gaulton, R., & Malthus, T. J. (2010). LiDAR mapping of canopy gaps in continuous cover forests: A comparison of canopy height model and point cloud based techniques. *International Journal of Remote Sensing, 31*(5), 1193-1211. doi: 10.1080/01431160903380565

Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters, 2*(4). doi: 045023
10.1088/1748-9326/2/4/045023

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2010). Relating Forest Attributes with Area- and Tree-Based Light Detection and Ranging Metrics for Western Oregon. *Western Journal of Applied Forestry, 25*(3), 105-111.

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2011). A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Canadian Journal of Forest Research, 41*(6), 1189-1201. doi: 10.1139/x11-033

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management, 2*(3), 231-244. doi: 10.4155/cmt.11.18

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., . . . Sun, M. (2009). Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance Management, 4*, 2.

Gonzalez, P., Asner, G. P., Battles, J. J., Lefsky, M. A., Waring, K. M., & Palace, M. (2010). Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sensing of Environment, 114*(7), 1561-1575. doi: 10.1016/j.rse.2010.02.011

Gregory, P. A. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters, 4*(3), 034009.

Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., & Stoker, J. M. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management, 208*(1–3), 189-209. doi: 10.1016/j.foreco.2004.12.001

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences, 33*(10), 1301-1315. doi: 10.1016/j.cageo.2007.05.001

Holmgren, P., & Thuresson, T. (1998). Satellite remote sensing for forestry planning—A review. *Scandinavian Journal of Forest Research, 13*(1-4), 90-110. doi: 10.1080/02827589809382966

Hudak, A. T., Lefsky, M. A., Cohen, W. B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment, 82*(2-3), 397-416. doi: 10.1016/s0034-4257(02)00056-1

Hunt, C. A. G. (2009). *Carbon sinks and climate change: forests in the fight against global warming*. Chelltenham, Glos., UK: Edward Elgar Publishing Ltd.

Husch Bertram, Beers W. Thomas, & Kershaw John A. (2003). *Forest Mensuration* (4th Edition ed.). Canada: John Wiley & Sons.

Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing, 29*(5), 1339-1366. doi: 10.1080/01431160701736489

ICIMOD, ANSAB, & FECOFUN. (2010). Report on forest carbon stocks in Ludikhola, Kayarkhola, Charnawati Watersheds of Nepal. Kathmandu, Nepal.

IPCC. (2007). IPCC Fourth Assessment Report (AR4). Climate Change 2007. *Mitigation of Climate Change* Retrieved May, 28, 2011, from http://www.ipcc.ch/publications_and_data/ar4/wg3/en/contents.html

Journel, A. G. (1987). GEOSTATISTICS - MODELS AND TOOLS FOR THE EARTH-SCIENCES *Mathematical Geology, 19*(4), 357-359. doi: 10.1007/bf00897846

Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010a). Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research, 40*(4), 761-773. doi: 10.1139/x10-024

Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010b). Comparisons between field and LiDAR-based measures of stand structural complexity. Retrieved from

Kim, S.-R., Kwak, D.-A., oLee, W.-K., Son, Y., Bae, S.-W., Kim, C., & Yoo, S. (2010). Estimation of carbon storage based on individual tree detection in Pinus densiflora stands using a fusion of aerial photography and LiDAR data. *SCIENCE CHINA Life Sciences, 53*(7), 885-897. doi: 10.1007/s11427-010-4017-1

Kim, S., McGaughey, R. J., Andersen, H.-E., & Schreuder, G. (2009). Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment, 113*(8), 1575-1586. doi: 10.1016/j.rse.2009.03.017

Kronseder, K., Ballhorn, U., Böhm, V., & Siegert, F. (2012). Above ground biomass estimation across forest types at different degradation levels in Central Kalimantan using LiDAR data. *International Journal of Applied Earth Observation and Geoinformation, 18*(0), 37-48. doi: 10.1016/j.jag.2012.01.010

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment, 70*(3), 339-361. doi: 10.1016/s0034-4257(99)00052-8

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sensing of Environment, 67*(1), 83-98. doi: 10.1016/s0034-4257(98)00071-6

Li, Y. Z., Andersen, H. E., & McGaughey, R. (2008). A Comparison of Statistical Methods for Estimating Forest Biomass from Light Detection and Ranging Data. *Western Journal of Applied Forestry, 23*(4), 223-231.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. [Article; Proceedings Paper]. *Canadian Journal of Remote Sensing, 29*(5), 658-678.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography, 27*(1), 88-106. doi: 10.1191/0309133303pp360ra

Lim, K. S., & Treitz, P. M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research, 19*(6), 558-570. doi: 10.1080/02827580410019490

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing, 27*(7), 1297-1328. doi: 10.1080/01431160500486732

Lu, D. S. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing, 27*(7), 1297-1328. doi: 10.1080/01431160500486732

Maltamo, M., Eerikäinen, K., Packalén, P., & Hyyppä, J. (2006). Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry, 79*(2), 217-229. doi: 10.1093/forestry/cpl007

McBratney, A. B., Webster, R., & Burgess, T. M. (1981). The Design of Optimal Sampling Schemes for local estimation and mapping of regionalized variables .1. theory and method. *Computers & Geosciences, 7*(4), 331-334. doi: 10.1016/0098-3004(81)90077-7

McGaughey, R. (2007). FUSION/LDV: Software for LiDAR Data analysis and Visualisation *Online PDF Document*: USDA Forest Service, Pacific Northwest Research Station.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., . . . McKee, W. A. (1999). Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment, 67*(3), 298-308. doi: 10.1016/s0034-4257(98)00091-1

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment, 80*(1), 88-99. doi: 10.1016/s0034-4257(01)00290-5

Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research, 19*(2), 164-179. doi: 10.1080/02827580310019257

Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing, 32*(2), 473-501. doi: 10.1080/01431160903474970

Næsset, E., & Gobakken, T. (2005). Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sensing of Environment, 96*(3–4), 453-465. doi: 10.1016/j.rse.2005.04.001

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment, 112*(6), 3079-3090. doi: 10.1016/j.rse.2008.03.004

Naesset, E., Gobakken, T., Holmgren, J., Hyyppa, H., Hyyppa, J., Maltamo, M., . . . Soderman, U. (2004). Laser scanning of forest resources: The Nordic experience. *Scandinavian Journal of Forest Research, 19*(6), 482-499. doi: 10.1080/02827580410019553

Nelson, B. W., Mesquita, R., Pereira, J. L. G., Garcia Aquino de Souza, S., Teixeira Batista, G., & Bovino Couto, L. (1999). Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. *Forest Ecology and Management, 117*(1–3), 149-167. doi: 10.1016/s0378-1127(98)00475-7

Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment, 24*(2), 247-267. doi: 10.1016/0034-4257(88)90028-4

Oli, B. N., & Shrestha, K. (2009). Carbon Status in Forests of Nepal: An Overview. *Journal of Forest and Livelihood, 8*(1).

Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2009). Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment, 113*(6), 1163-1174. doi: 10.1016/j.rse.2009.02.002

Panta, M., Kim, K., & Joshi, C. (2008). Temporal mapping of deforestation and forest degradation in Nepal: Applications to forest conservation. *Forest Ecology and Management, 256*(9), 1587-1595. doi: 10.1016/j.foreco.2008.07.023

Patenaude, G., Hill, R. A., Milne, R., Gaveau, D. L. A., Briggs, B. B. J., & Dawson, T. P. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment, 93*(3), 368-380. doi: 10.1016/j.rse.2004.07.016

Patenaude, G., Milne, R., & Dawson, T. P. (2005). Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science & Policy, 8*(2), 161-178. doi: DOI: 10.1016/j.envsci.2004.12.010

Phillips, D. L., Brown, S. L., Schroeder, P. E., & Birdsey, R. A. (2000). Toward error analysis of large-scale forest carbon budgets. [Article]. *Global Ecology and Biogeography, 9*(4), 305-313. doi: 10.1046/j.1365-2699.2000.00197.x

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologist*. Cambridge: Cambridge University Press.

Richardson, J. J., & Moskal, L. M. (2011). Strengths and limitations of assessing forest density and spatial configuration with aerial LiDAR. *Remote Sensing of Environment, 115*(10), 2640-2651. doi: 10.1016/j.rse.2011.05.020

Rosenqvist, Å., Milne, A., Lucas, R., Imhoff, M., & Dobson, C. (2003). A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy, 6*(5), 441-455. doi: Doi: 10.1016/s1462-9011(03)00070-4

Sales, M. H., Souza, C. M., Kyriakidis, P. C., Roberts, D. A., & Vidal, E. (2007). Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondonia, Brazil. *Ecological Modelling, 205*(1-2), 221-230. doi: 10.1016/j.ecolmodel.2007.02.033

Sprugel, D. G. (1983). Correcting for Bias in Log-Transformed Allometric Equations. *Ecology, 64*(1), 209-210. doi: 10.2307/1937343

Walker, W., Baccini, A., Nepstad, M., Horning, N., Knight, D., Braun, E., & Bausch, A. (2011). Field Guide for Forest Biomass and Carbon Estimation. Version 1.0.: Woods Hole Research Center, Falmouth, Massachusetts, USA.

Webster, R., & Oliver, M. A. (2001). *Geostatistics for Environmental Scientists*. New York: John Wiley and Sons.

Webster, R., Welham, S. J., Potts, J. M., & Oliver, M. A. (2006). Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood. *Computers & Geosciences, 32*(9), 1320-1333. doi: 10.1016/j.cageo.2005.12.002

Yan, W. Y., Shaker, A., Habib, A., & Kersting, A. P. (2012). Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*(0), 35-44. doi: 10.1016/j.isprsjprs.2011.10.005

Yu, X., Hyyppa, J., Hyyppa, H., & Maltamo, M. (2004). *Effects of flight altitude on tree height estimation using airborne laser scanning*. Paper presented at the International Conference NATSCAN 'Laser-Scanners for Forest and Landscape Assessment - Instruments, Processing Methods and Applications', Freiburg, Germany.