Modeling Soil Organic Carbon (SOC) using remotely sensed variables in Chitwan District, Nepal.

SIKDAR MOHAMMAD MARNES RASEL
February, 2013

SUPERVISORS:
Dr. Thomas A. Groen
Dr. Yousif Hossin

Modeling Soil Organic Carbon (SOC) using remotely sensed variables in Chitwan District, Nepal.

SIKDAR MOHAMMAD MARNES RASEL
Enschede, The Netherlands, February, 2013

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.
Specialization: Natural Resources Management

SUPERVISORS:
Dr. Thomas A. Groen
Dr. Youssif Hussin

THESIS ASSESSMENT BOARD:
Chair: Prof. Dr. Kees de Bie
External Examiner : Dr. M. Gerke, ITC-Department of Earth Observation Science, University of Twente, The Netherlands.

ABSTRACT

Now a days the possibility of enhanced carbon storage in soils is of more interest compared to vegetation as it contains more carbon. For this reason, the revised Kyoto protocol includes two new clauses relevant to soil organic carbon sequestration. So, for the countries that have signed the Kyoto protocol, estimation of SOC sequestration is a required strategy. Reliable quantification of carbon held in soil is essential to formulate any kinds of monitoring program. Instead of a traditional laboratory method, estimation of carbon through a model might be an easy alternative. It would save time and remove the tedious task of soil sampling and processing. This study aims therefore to develop a model based on remotely sensed measured variables to estimate SOC in the subtropical forest of Chitwan, Nepal.

To develop a model, six variables were selected, above ground biomass (AGB), elevation, species diversity, litter quality, soil bulk density and soil pH to estimate soil organic carbon. Although soil bulk density and pH cannot be measured through remote sensing technology, they were used to test the robustness of model. Soil organic carbon was analysed through Walkley-Black and Loss on Ignition (LOI) methods. Canopy Height Model (CHM) was developed from LiDAR data by subtracting the Digital Terrain Model (DTM) from the Digital Surface Model (DSM) to estimate the height of the trees. This CHM image was segmented based on an Object Based Image Analysis (OBIA) technique using e Cognition software. Segmented CPA further analysed to develop a model for DBH prediction. With the information of DBH, tree height and wood specific gravity, AGB was calculated. Elevation height was extracted from LiDAR derived DEM. A Worldview -2 high resolution image was classified to extract the information of tree species class. The image was classified into two classes sal (*Shorea robusta*) and non-sal (mixed species). These two classes were further transferred into a litter quality index by using a dummy variables code. A Stepwise regression procedure was followed to select the best fit model.

Results show that there is a positive relationship ($r = 0.79$) between soil organic carbon and above ground biomass ($p<0.001$). Elevation and soil organic carbon is also positively correlated ($r=0.74$). There is no significant relationship between species diversity and soil organic carbon. Based on AIC and p value a regression model with above ground biomass ($p<0.001$) and litter quality ($p=0.07$) was selected to estimate soil organic carbon. ($p=0.07$). Root Mean Square Error (RMSE) for the selected model was 18.14%. Selected variables AGB and litter quality can be measured through remote sensing techniques. Based on AGB (kg/m^2) pixel value and litter quality (0 or 1) pixel value, SOC map was prepared. This model was tested with the field observed SOC value and shows a strong correlation coefficient value ($r=0.82$). Predicted model estimated average 1.77 kg/ m^2 soil organic carbon within 0-10 cm layer in the Chitwan district of Nepal.

Keywords: *Soil Organic Carbon(SOC), Bulk Density (BD), Soil pH, Litter Quality (LQ), Loss on Ignition(LOI) , Walkley – Black(WB) method, Stepwise regression, Biomass, Crown Projection Area(CPA), Diameter at Breast Height(DBH), Species Diversity, Allometric equation.*

ACKNOWLEDGEMENTS

First and foremost I would like to thank the almighty Allah for his blessings, protection and guidance during the whole study period. My special thanks go to NUFFIC which provided me with a fellowship to study and join with people from different countries around the world. I am very much grateful to my employing organization 'Soil Resources Development Institute', Ministry of Agriculture, Bangladesh for providing the opportunity to study in abroad.

I am very much grateful to Dr. Thomas A. Groen, my first supervisor for his continuous support, invaluable suggestions, constructive feedback and comments. Sincere thanks goes to my second supervisor, Dr. Yousif Ali Hussin, for his encouragement in field work, guidance, supervision and feedback.

My heartfelt gratitude goes to Drs. J.B. (Boudewijn) de Smeth Geochemist, Tutor AES-ERE courses, Head GeoScience Laboratory, for his guidance before going to field, support in laboratory work and constructive suggestions. Sincere thanks goes to Dr. Kees de Bie and Dr. Michael Weir for proper and constructive feedback during proposal defence and midterm presentation. Their comments really help me to improve the structure of this work.

I would like to extend deepest thanks to all NRM-GEM classmates, fieldwork colleagues Metadel and Getachew who shared together the tough and cheerful moments. I am very much thankful to Islam Fadel, Sunil Thapa, Gloria, Rory, Nuno, Beza, Sharmin, Arun, Pradeep, Lina, Collins, Amare, Ragab and all of my friends for their support and wishes to do a good work.

Last my everlasting gratitude goes to my parents, my wife and son, my brothers and sisters for their prayers, encouragement and moral support. Your presence has a great value for me. You have always special place in my heart.

Sikdar Mohammad Marnes Rasel
Enschede, the Netherlands
February, 2013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF APPENDICES

LIST OF ACRONYMS

AGB	Above ground biomass
AIC	Akaike Information Criterion
ANSAB	Asia Network for Sustainable Agricultural and Bio-resources Convention on Biological Diversity
BD	Bulk Density
CF	Community Forest
CFGUs	Community Forest User Groups
CHM	Canopy Height Model
CPA	Crown Projection Area
DBH	Diameter at breast height
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
GHGs	Greenhouse gases
GPS	Global positioning system
ICIMOD	International Centre for Integrated Mountain Development
IPCC	Intergovernmental Panel on Climate Change
LiDAR	Light Detection and Ranging
LOI	Loss On Ignition
OBIA	Object Based Image Analysis
REDD+	Reduction carbon emission form deforestation and forest degradation and foster conservation, sustainable management of forest, and enhancement of forest carbon stocks
RMSE	Root Mean Square Error
SOC	Soil Organic Carbon
SIC	Soil Inorganic Carbon
UNFCCC	United Nations Framework Convention on Climate Change
WB	Walkley-Black

1. INTRODUCTION

1.1. Background

The atmospheric concentrations of CO_2 and other greenhouse gases (GHGs) has increased drastically since the industrial revolution(Lal, 2004). According to the records of IPCC. (2001), the concentration of atmospheric CO_2 has increased from 280 ppmv in 1750 to 367 ppmv in 1999 and the current increasing rate is 1.5 ppmv/year or 3.3 Pg C/year (IPCC., 2001). The main greenhouse gases (CH_4, N_2O and CO) and their cumulative pressure in the atmosphere has led to an increase in the average global surface temperature of 0.6 °C since the late 19th century, with a current warming rate of 0.17°C / decade (IPCC., 2001). The global carbon budget for the decade of 1990-2000 included an emission of 6.3±0.4 Pg C from fossil fuel combustion and cement production and an emission of 1.6±0.8 Pg C from land use change (Prentice. (2001);Schimel. et al. (2001)). The mentioned data indicate that land use, soil management and terrestrial ecosystems play an important role in the global C budget. Due to land use, land use change, forestry and other forest activities like biomass burning, fertilization and wetlands restoration, the emission of CH_4 and N_2O is increasing. In a same time, terrestrial ecosystem, in which C - is stored in live biomass, plant litter, organic matter and soil play an important role in the global carbon cycle. There are five main global carbon pools: the oceanic, geologic, pedologic (soil), biotic and the atmospheric pool (Figure. 1). These five C pools are connected with each other and C exchanged from one pool to other through photosynthesis, respiration, decomposition and combustion. Proper monitoring and accurate estimation of these pools help to initiate the mitigation steps of climate change (CC).

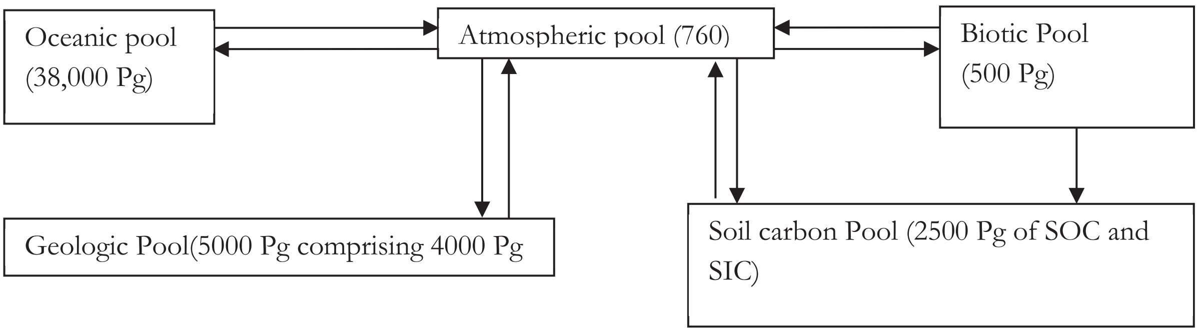


Figure1. Principal global carbon pools, adapted from Lal (2004).

About 2,500 Pg of Carbon (C) is stored in soil, compared to 760 Pg in the atmosphere (CBD., 2009). Globally forest vegetation and soils removed carbon from the atmosphere at a rate of 4.7±1.2 Gt (Giga tones) per year in 2008, compared to carbon emissions from fossil fuels and deforestation of 8.7±0.5 Gt per year and 1.2±0.7 Gt per year respectively (IPCC, 2007). Therefore, biomass C and soil C are considered two important components of carbon storage in forest ecosystem. In forest, biomass and soils contain about 1240 Pg of C (Dixon et al., 1994). Compare to biotic pool soil pool stores more carbon. Soil carbon pool is the combination of soil organic carbon (SOC) and soil inorganic carbon (SIC). Due to the large areas involved at regional or global scale, forest soils play an important role in the global C cycle (Lal, 2005).

Due to much focus on biotic pool and biomass estimation, soil organic pool (SOC) is always ignored or very few works have been done on it. Instead of direct destructive method (cutting and weighing) remote sensing technology are using to improve the monitoring and accurate estimation of tree biomass. But for

SOC pool estimation and monitoring, application of remote sensing is quite difficult in forest due to obstacle made by above ground biomass and other some related issues.

1.2. Overview of Remote sensing technologies for Soil Organic Carbon (SOC) measurement

The composition of soil organic carbon depends on the litter material and it's decomposition rate (Ben-Dor et al., 1997). The biochemical composition that present in the leaf, stems can also be present in the soil. Soil organic matter contains chlorophyll, pectin, oil, starch, lignin and celluloses (Beyer et al., 2001). These biochemical constituents influence the reflectance in the Visible (VIS, 400-700 nm), Near – Infrared(NIR, 700-1400 nm) and Short Wave Infrared (SWIR, 1400-3000 NM) region of the electromagnetic spectrum (Ben-Dor et al., 1997). Based on this principal Bartholomeus et al. (2008) conducted a study on spectral reflectance indices for soil organic carbon quantification. They demonstrated the feasibility of spectral indices derived from laboratory measurements to predict soil organic carbon in different soil types. But the limitation of this method is validation of the model needs a large variance of SOC from sample to sample. Because their training samples were from different soil types, different climatic zones and from different horizon (Bartholomeus et al., 2008). It means the soil samples of same climatic condition and geographic location with small variability of SOC is not suitable to predict SOC. Still this procedure works within laboratory condition and not practiced in the field.

There are some other advance technologies which are not related to image and spectral characteristics but related with reflectance. They are qualitative methods and one is based on nuclear magnetic resonance (NMR) spectroscopy and the other on Diffuse Reflectance Infrared Fourier Transform (DRIFT) spectroscopy (Brian. A. S., 2002). Nuclear Magnetic Resonance (NMR), a tool for the characterization of soil organic matter and works on the principle of "measuring the characteristic energy absorbed and re-emitted or dispersed by atomic nuclei that are placed in a static magnetic field and subjected to an oscillatory magnetic field of known radio-frequency. One specialized form of this technique is cross-polarization magic angle spinning (CPMAS) 13C NMR" (Rumpel et al., 1998). This is capable of distinguishing chemical structures that are characteristic of recently formed organic matter as well as those organic carbon forms derived from the soil's parent material/geology. But the NMR spectroscopy are expensive and time-consuming (Rumpel et al., 2001)."DRIFT spectroscopy is used in conjunction with multivariate data analysis (i.e., partial least squares) to provide a rapid and inexpensive means of differentiating carbon forms in soils and sediments" (Rumpel et al., 2001). By his method, carbon compounds are differentiated by assignment of the main infrared absorption bands to the bands being stretched or deformed at that particular frequency. Here major advantage is both inorganic and organic forms of organic compounds may be identified (Nguyen, 1991). Research is still on-going for these techniques to improve the accuracy.

Very recent study from Nocita et al. (2012), it is proved that visible and near infrared diffuse reflectance spectroscopy has produced promising results to infer soil organic carbon (SOC) content in the laboratory. But, soil spectra measured from the field or with airborne imaging spectrometers still challenging due to uncontrolled variations in surface soil conditions, like soil moisture and roughness and vegetation cover.

As the application of remote sensing for direct measurement of soil organic carbon (SOC) still a process of advance scientific research, an alternative solution of SOC measurement is the use of different proxies or variables; those are related with SOC and can be measured through RS technology. Before selecting those proxies, a clear understanding of soil organic carbon cycle and the fluxes related to this is essential to know.

This SOC is governed by a number of interacting factors including climatic, environmental and anthropogenic factors (Figure 2: Conceptual model of SOC). Among those factors, some are directly related with decomposition and SOC formation. Some factors influence the SOC formation process in an indirect way. Some factors can be measured directly through remote sensing technology and some need alternate methods to measure. Figure 2 depicted the whole story of SOC and the list of influencing factors those can be measured by remote sensing technology.

1.2.1. **Process of SOC formation (Inflow and outflow within a system)**

Soil pool contains more carbon than atmosphere and forests combined. This pool is a result of ecological processes occurring at or near the soil surface, such as litter decomposition, mineral cycling, water cycling, and microbial activities.

To understand the formation process of soil organic carbon, it is essential to know the total influx and outflow within the soil system. The process starts from photosynthesis (inflow) and ends with the soil respiration (outflow), formation of soil organic carbon (SOC) and leaching loss. The whole process of inflow and outflow and a brief description of related factors within this process are presented here.

Step-1: Photosynthesis: The first process through which CO_2 is converted into organic matter is by plants which under the influences of sunlight photosynthesize water (H_2O) and CO_2 into carbohydrate (Figure 2.a). This carbohydrate forms the building blocks of biomass. The chemical composition of biomass varies from species to species. Generally plants biomass consist of about 25% lignin and 75% carbohydrates or sugars (cellulose and hemicellulose)(P Lerouge, 1998).

Step-2: Litter formation: Plants litter forms at or beneath the surface of soil. Leaves, branches, twigs, dead woody parts are the major input of litter. Generally where there is more above ground biomass, there is more chance of leaf falling and litter deposition. The type and richness of species also effects on the size and decomposability of litter. Some litters are decomposed slowly and some decomposed very fast. In other way it can be summarised that, litter amount and the quality of decomposability of litter is determined by the amount and type of biomass above the soil surface (Figure 2b).

Step-3: Respiration and oxidative decomposition: This litter started to decompose with the help temperature, water, microbes and of course the chemical composition of biomass itself. Here lignin % plays a major role to accelerate the decomposition. Under the process of oxidative decomposition organic matter is converted to CO_2 and H_2O (Figure 2.c) over a period of months to years. Higher temperature and precipitation rate influences this oxidative decomposition process. Higher elevation by its cold temperature, frost and water logged situation also influences the decomposition rate. CO_2 comes from the oxidative decomposition in addition with root respiration CO_2 go back to the atmosphere. Most of the soil CO_2 back to the atmosphere by diffusion and by convection under the influence of temperature and water content of soil (Figure 2.d).

Step-4: Dissolving of CO_2: Part of the CO_2 dissolved in the soil solution and removed by drainage (Figure 2.e). How much CO_2 will go back to the atmosphere and how much it will remain in the soil as dissolved form depends on the pore space as well as the bulk density of soil. Higher bulk density means less pore space for CO_2 retention.

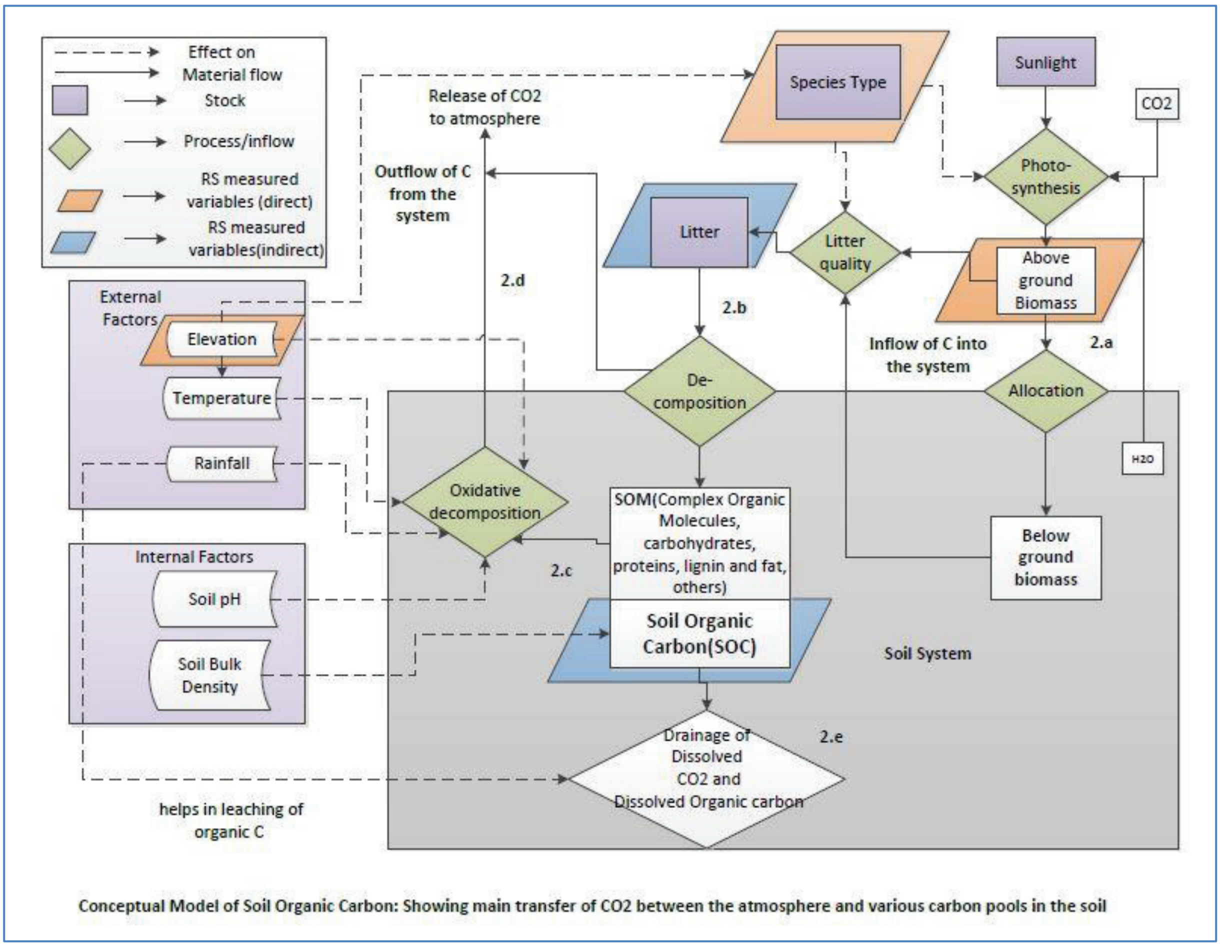


Figure2. Conceptual model of soil organic carbon showing the inflow and outflow of CO_2.

1.2.2. Factors affecting the soil organic carbon pool

The size of the soil organic matter pool depends upon plant growth, litter formation rate, the extent and rate of mineralization of the plant residues entering the soil. This complex process is controlled by several factors including soil type, temperature, and precipitation rate, biochemical composition of the plant residue and the nature and abundance of decomposing organisms. The environmental variables such as: altitude, slope and landscape position can impact on the soil's C stock. This is because of their influence on the soil temperature, soil water and pore space retention (Gulledge & Schimel, 2000). Among those factors, some are very much correlated and mentioned in the steps of SOC formation process (figure 2) are listed below for further discussion.

i) Above ground biomass
ii) Species diversity
iii) Litter quality
iv) Temperature
v) Rainfall
vi) Elevation and topography
vii) Soil pH
viii) Soil bulk density.

Above ground biomass, species type and litter quality

How much litter will be deposited at or under the soil surface depends on the **above ground biomass** and its type. Plant types and amount of biomass significantly affected the distribution of SOC (Esteban. & Robert., 2000). According to their study, the percentage of SOC in the top 20 cm averaged 33%, 42%, and 50% for shrub lands, grasslands, and forests, respectively. They also concluded that globally the relative distribution of SOC with depth had a slightly stronger association with vegetation than with climate. Esteban. and Robert. (2000) suggested that shot/root allocations combined with vertical root distributions, affect the distribution of SOC with depth.

Not only the shoot/root ratio, the amount of soil organic carbon also influenced by the **litter** that deposited from the above biomass and root decomposition. Different chemicals and their amount in leaf, foliage has a relation with litter and can be used as a substitute for litter quality. However, there is no universal **litter quality index** because litter decomposition depends on qualities which differ among species and plant parts. The rate of litter decomposition is associated with the lignin and nitrogen content. So the decomposition of litter turning into soil organic carbon (SOC) is determined by the degradation rate of lignin. During the oxidation process lignin decomposes slowly, much slower than cellulose. This lignin and N concentration in leaf also varies from species to species. Generally deciduous species appear to have foliar litter richer in N than have conifers. Higher N levels in litter, lower decomposition rate and thus a considerable increase in humus accumulation. The deciduous litter shows a decomposition rate slowly compare to conifers.

Deciduous and coniferous is a broad classification even litters decomposition and SOC formation is influenced by the species within deciduous and coniferous group. For example effects of site and tree species on SOC (0–10 cm depth) were examined by Kasel et al. (2011) using soils occurring under four species (*Acacia implexa, Acacia mearnsii, Allocasuarina verticillata and Eucalyptus melliodora*). They found that trees growth had positive effect on SOC. This relationship between SOC and species diversity was tested by Saha et al. (2009) for home gardens (HG), a popular and sustainable agroforestry system in the tropics, in Thrissur district, Kerala, India. They also measured tree density and tree species **diversity (Shannon Index)** of the HG. Results indicated that the soil C stock was directly related to plant diversity of HG. Soil-C storage in case of species-rich home gardens could have relevance and applications in broader ecological contexts (Saha et al., 2009). Tropical forests may be an example where a large amount of CO_2 is sequestered by the photosynthesis process of its diversity of tree species.

Effect of temperature on SOC

Enzymatic activity during decomposition normally increases with temperature, but rapidly falls as the temperature rises above an optimum value. A study of Deqiang et al. (2008) proved that increasing temperature from 15°C to 18 °C significantly increased the amount of CO_2 emissions from the litter. It means the oxidative decomposition increases. The effect of temperature and nutrient concentration on litter quality was investigated by Salah and Scholes (2011). According to their finding temperature affected the N accumulation of the litter. Higher temperatures resulted in more accumulation of N (Salah & Scholes, 2011). Generally trees growing under warmer and wetter climates (higher actual evapotranspiration, AET) tend to shed foliar litter more rich in N than those growing under colder and drier climates. When more N accumulates decomposition rate also increases. Higher decomposition rate means lower SOC formation into the soil. When the temperature is getting colder, it has also an influence on litter decomposition rate. Cold temperature reduces the metabolic activity as well as slows down the process of soil organic matter breaking into CO_2 and H_2O.

Effect of precipitation on SOC

Rainfall affects various soil biological activities because of its influence on soil moisture and temperature. Low moisture level reduces metabolic activity, and as soil moisture levels rise, metabolic activity increases up to an optimum level. Metabolic activities are related with the decomposition of different compounds within leaf and litter. At low moisture level, the decomposition rates of some biochemical compounds are reduced, and some processes are completely suppressed, for example, lignin decomposition (Grizelle., 2001). In litter decomposition, the leaching effect of rainfall increases mass loss at the initial stage of the decay process (Grizelle., 2001).

Effect of Elevation and topography on SOC

Among all the environmental variables those that play the most vital role are slope and elevation. The strong effect of slope and aspect on SOC stock was found in research done of a subalpine forest in the Olympic Mountains of Washington state was (Prichard et al., 2000). They found that soil organic carbon increases with elevation distance up to 1600m. Lal. (2001) found that soil carbon increased with elevation and in their study, they found an almost four fold increase in soil carbon, from 2.1 to 8.0% (mass based) between 600 to 1600m. In high-altitude ecosystems soils play a vital role in the global terrestrial carbon cycle due to their large carbon stock (Post et al., 1982). It happens due to a number of unique factors, e.g. permafrost, cold temperature and water-logging (Hobbie et al., 2000) .

Change of elevation distance has a relation with temperature and precipitation. The higher the elevation height, temperature is going to be colder. In same way in high elevation range, rainfall and precipitation are more active to facilitate the anaerobic condition into the soil system due to it's waterlogging, frosting and others related climatic parameters which already discussed in previous paragraph.

Effect of soil pH and bulk density on SOC

Soil pH has a significant effect on soil organic matter preservation and decomposition, although its precise mode of influence has needed to be fully established. Still there are a lot of conflicting views concerning the relationship between soil pH and soil organic matter. Spain (1990) found a negative correlation between soil organic matter and soil pH in tropical rainforest soils. On the other hand Hardon (1936) observed that in acidic soils organic carbon contents increased.

The most important chemical reaction in SOC systems are humification and mineralization. These processes are responsible for changing the chemical composition of soil organic matter and are of great importance to the terrestrial carbon cycle. From the literature it is proved that soil pH mainly affect these two chemical reaction specially lignin decomposition.

Guggenberger. et al. (1995) observed a significant lignin contribution in strongly acidic soils with a $pH<5$, whereas moderate acidic soils with $pH>5$ showed little evidence of lignin. Motavalli. et al. (1995) also suggested that acidic soil reduces the decomposition rates of freshly added organic materials. From the literature, it is known that soil of the study area is more or less acidic soil. To analyse the relationship between soil pH and the soil organic reservoir in this study, soil sample was analysed in the laboratory. The details result of analysis will be discussed on the chapter 4.

Generally when the bulk density is higher there is less pore space as well as less scope of soil organic matter deposition. Soil organic matter can increase soil aggregate stability, but reduces bulk density and improves moisture retention (Arvidsson, 1998). In this study, it was not possible to deal with all physical parameters of soil due to lack of time and budget. Considering those constraints only bulk density was measured in filed to analyse the relationship between SOC and density of soil.

It is already mentioned that an alternative way of SOC measurement would be the use of proxies those are discussed here. But all of them are not possible to measure through remote sensing technology. Such as for soil pH and soil bulk density, field measurement and laboratory analysis are needed for accurate result. From the literature and finding it is proved that the remote sensing technology can be used to monitor and estimate the following variables those have either a direct or an indirect relation with soil organic carbon. For example

i) Above ground biomass(AGB)

ii) Species class
iii) Elevation

1.3. Overview of remote sensing (RS) technology in relation with elevation, above gound biomass and species classification

1.3.1. Application of remote sensing for elevation delineation

It has already been discussed that some topographic data (slope, aspect and elevation) has a relation with SOC. These kinds of data can be extracted from a Digital Elevation Model (DEM) of a specified area. Light Detection and Ranging (LIDAR) and SAR and stereo-correlation of images are the most widely used sources for constructing a DEM. Lidar has very high vertical accuracy which support it to represent the earth surface with high accuracy (Vaze., 2001). This is the most attractive characteristics of LiDAR. Normally LIDAR allows very accurate and densely sampled elevation points (Woolard. & Colby., 2002).
DEMs provide a valuable tool in soil-landscape modeling. Now days, in case of forest biomass research DEM data are used not only for topographic (elevation and slope) information but also to derive Crown Height Model (CHM) for the accurate height of tree. Digital Surface Model (DSM) and Digital Terrain Model(DTM) can be used to extract the tree parameter information (Drouin et al., 2011) and to develop a CHM.

1.3.2. Application of remote sensing for above ground biomass estimation

Above ground biomass (AGB) contains 47% of carbon which is defined as "all biomass of living vegetation, both woody and herbaceous, above the soil including stems, stumps, branches, bark, seeds and foliage" (IPCC, 2007). The majority of biomass estimations are done for above ground biomass of trees because these generally account for the greatest fraction of total biomass(above and below ground together) in a forest (Malhi, 2010). Many methods have been practiced for forest biomass estimation. The most accurate one is called the destructive method. Here procedure is to harvest the trees, oven dry them and to weigh the dry matter. But it is time consuming and not feasible at all (Hunt., 2009).
The alternate approach is to use satellite images for biomass estimation. There are many remote sensing methods available to estimate carbon stocks but these methods also cannot measure carbon stock directly and thus require additional ground based data collection based on tree parameters. Different remote sensing sensors like optical sensors, microwave sensor (LiDAR, RADAR) are used for biomass estimation. Very High Resolution (VHR) images with spatial resolution of less than 5 m (Lu, 2006) such as aerial photograph, satellite images such as Quick bird, IKONOS, Worldview and Geo-eye images can detect individual tree crowns (Gonzalez et al., 2010). But the problem of cloud cover reduces the accuracy for optical sensor.
This cloud cover problem can be recovered by Light Detection and Ranging (LiDAR) system. Now days it is becoming a promising technique for future forest monitoring. Because it has the ability to assess the forest in 3D structure (Patenaude. et al., 2005) and provide a good data on vertical profiles of vegetation canopies (Belzar et al., 2007). This active remote sensing system operates laser pulses towards the ground to records the elapse time between beam launce and return registration. Records of this return is known as cloud points reflected from tree canopy, trunks, branches, leaves, low vegetation and even reaching to the ground to create a 3 –dimensional profile (Gautam. & Kandel. P.N, 2010). In this way airborne LiDAR offers the unique capability of measuring the three-dimensional vertical structure of vegetation in great detail which in itself is an advantage over high resolution satellite imagery (Song, 2010). But it depends on the point density of LiDAR point clouds. Information of vertical profiles can be provided and differentiated by direct retrieval (through tree canopy height model or CHM) or by integrating with other sensors. In this way traditional biomass estimation can be improved by vertical component (3D) provided by LiDAR separately or fusing with other multispectral sensors (Dubayah. et al., 2000).

The way of feature extracting from image is called image classification. Traditional approach of image classification is pixel based classification. This classification uses the spectral information to extract the features (tree canopy, shade or non-forest area etc.) from image. Instead of pixel based classification Object Based Image Analysis (OBIA) classification is suited for extracting features from LiDAR image. This is a new classification method refers to the partition of an image into discrete non overlapping units called image objects. It uses spectral information and considers homogeneity in terms of texture (mean, variance, contrast), spatial, contextual information to interpret an image (Definiens, 2011).

Lopez Bautista (2012) used this classification technique for same study area to fit a regression model to estimate the carbon stock of a forest. He compared LiDAR with optical images and revealed that LiDAR is more accurate for biomass estimation compared to estimation based on optical images. The advantages of LiDAR is of laser beam penetration into the deep forest and resulting data are without being influenced by clouds and shadows, thus providing more accurate results than any other remote sensing techniques(ARBONAUT. (2010);Naesset. (2009)).

1.3.3. **Application of remote sensing technology for Tree species classification**

Species level tree classification is the next step of biomass estimation. Different allometric equations are used for biomass estimation. This equation differs from species to species. The spectral signature of each individual species is different with others. This spectral signature is related with the band combination of images used for species classification. The recently launched Worldview-2 satellite is said to be a second generation satellite that has a unique combination of various bands (DigialGlobe, 2010). This satellite, partially conceived for applications in precision agriculture and natural resources monitoring, provides spectral information in 8 spectral bands: the 4 most common spectral regions (red, blue, green, near-IR) and 4 new bands (red edge, coastal, yellow and near-infrared 2). The list of band with respective wavelength is presented at the Annex 2. Splitting the NIR region for more information and adding the narrow Red Edge band beyond the Red band, highly increases the sensitivity of this sensor compared to other multispectral sensors in vegetation monitoring. The yellow, Red-edge and two bands of NIR are regarded as important for vegetation study; IR1 band has the great potentiality to identify the vegetation type at species level (DigialGlobe, 2010). Therefore, it is highly recommended by (Baral Jamarkattel, 2011), who carried out her research in Chitwan district, Nepal to use this image to further explain the estimation of carbon stock since she could not achieve a good species classification result due to geo-referencing problems and other artifacts. Karna (2012) also conducted his MSc research in the same study area and used the Worldview – 2 images to classify the major species. He used all bands of a pan sharpened image and classified into major six species. When species classification give a good accuracy result it also helps to estimate the biomass of forest based on species specific allometric equation.

1.3.4. **Application of remote sensing for litter quality assessment**

Species specific effects on litter quality are already discussed in section 1.1 (SOC conceptual model). But the issue is how it can be measured through RS technology. Litter decomposition is related with the lignin and nitrogen concentration of the leaf. Martin. and John. (1997) designed a study to determine whether data from NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) could be used to determine forest canopy chemistry at a spatial resolution of 20 m. They estimated nitrogen and lignin concentrations with their models to predict net ecosystem productivity at Harvard Forest and nitrogen mineralization rates at Blackhawk Island. This is a quantitative measurement of forest canopy chemistry. In another study Katherine et al. (1996) evaluated near infrared reflectance spectroscopy as a method for the determination of nitrogen, lignin, and cellulose concentrations in dry, ground, temperate forest woody foliage. Those all are direct measurement and assessment of lignin through a statistical model.

If quantitative data like lignin %, nitrogen concentration and decomposability rate are absent, alternative solution is data transformation. Here litter quality can be counted as a qualitative data and can be transferred categorical classes of high and low litter quality.

1.4. Justification of the work and selection criterial for study area

As soils sink more carbon than atmosphere and vegetation combined, and can hold it longer, research interest are increasingly looking to soil carbon as an opportunity to mitigate climate change. For regular monitoring, correct algorithms or feasible strategies are essential to estimate soil organic carbon. Remote sensing based technologies for soil organic carbon estimation still are in a process of establishment. So this study proposed to establish a model for soil organic carbon estimation based on other variables those can be measured through remote sensing technology.

The revised Kyoto Protocol includes two new clauses relevant to SOC sequestration: "(1) Countries are allowed to subtract from their industrial C emissions certain increases in C sequestered in "sinks" such as forests and soils; and (2) Countries are allowed to trade emission allowances that can reduce abatement costs. The UNFCCC/ Kyoto Protocol recognize soil C sinks provided that the rate of SOC sequestration and the cumulative magnitude can be verified by standard procedures" (Lal, 2004). So for countries that have signed the Kyoto protocol, estimation of SOC sequestration is a feasible strategy for them.

1.4.1. Selection criteria for study area

The countries that have signed the Kyoto protocol and UNFCCC agreement, monitoring of carbon pools is a major concern for them. Except this, there are other some relevant issuers those helped to select the country Nepal as a suitable area for this study. The issues are as follows:

- **UNFCCC and Kyoto protocol signatory country**

As Nepal is a signatory country to the Kyoto protocol, SOC data is very essential for Carbon trading. The specific location of the study (Kayerkhola watershed) is an area where REDD+ (Reduction carbon emission from deforestation and forest degradation and forest conservation, sustainable management of forest and enhancement of forest carbon stocks) pilot project is implemented.

- **Mountainous and hilly watershed**

Nepal is a mountainous country. The study area is located in Chitwan Province. This study area has a diverse climate due to elevation ranges from 200m to 1200m.

- **Data availability**

LiDAR data was provided by FRA project, Nepal and Worldview -2 image was only available for this study area. In the last two years several studies have been conducted in the same area to estimate the above ground carbon biomass. So above ground biomass and species classification related data are available for this area (Lopez Bautista (2012);Baral Jamarkattel (2011);Karna (2012)).

- **Accessibility**

It is a high mountainous and dense forest area but accessible to do any research work. Because it is a managed forest area managed by Community Forest User Groups (CFUGs). They are the local people who are involved with the management of forest.

But hardly any work has been done on the soil organic carbon stock. Specially in relation with remote sensing parameters. So this area is selected as the area of interest for this study. As soil organic carbon is an important carbon pool, it is also important to research how the SOC correlate with elevation, above ground biomass and tree species diversity. As those variables are directly measured by RS technology, instead of direct laboratory analysis an indirect algorithm may be developed to estimate SOC.

1.5. Research aim

The main aims of this research are to assess the effect of elevation, aboveground biomass and tree species diversity on Soil Organic Carbon (SOC) and to develop a model to estimate SOC stock using airborne LiDAR and high resolution Worldview image -2 measured variables.

1.5.1. Specific Objectives

1. To evaluate the effect of elevation on SOC in Community Forest (CF) of Nepal.
2. To evaluate the effect of above ground biomass (AGB) in Community Forest (CF) of Nepal.
3. To evaluate the effect of tree species diversity on SOC in CF of Nepal.
4. To find out a best fit model of SOC estimation from LiDAR and Worldview-2 image measured variables from the study area.

1.6. Research Questions

1. Is there a positive correlation between elevation and SOC in Community Forest (CF) of Nepal?
2. Is there a positive correlation between above ground biomass and SOC measured by VHR images and airborne LiDAR data?
3. Is there a positive correlation between tree species diversity and soil organic carbon in CF of Nepal?
4. Which regression model best explain the relationship between SOC and all other remotely sensed variables measured by LiDAR data and Worldview-2 image?

1.7. Research Hypothesis

1. There is a strong positive correlation between elevation and Soil Organic Carbon in Chitwan forest, Nepal.
2. There is a strong positive correlation between above ground biomass/carbon stock and soil organic carbon in study area.
3. There is a strong positive correlation between species diversity and soil organic carbon.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Study area

The study be found in Chitwan district of Nepal (figure-3). The area is situated between 27°30'51"N - 27°52'01 N latitude and 83°55'27"E - 84°48'43"E longitude and surrounded by the Makwanpur district in the east and the Nawalparasi in the west. The neighboring districts in the northern part are Dhading, Gorkha and Tanahu while Parsa district and India are located on its southern borders. Chitwan is situated 68 kilometers south east (133°) of the approximate center of Nepal and 82 kilometers west (260°) of the capital Kathmandu. The elevation height varies from 200 m -1100m above sea level. Out of 2218 km^2 of the total district area, Kayerkhola Watershed, the study area is covered by 660. ha of forest including 3 community forest (Shah, 2011). The respective areas occupied by three community forest are as follows: Devidhunga 253 ha, Nibuwatar 329 ha and Janpragati 78 ha.

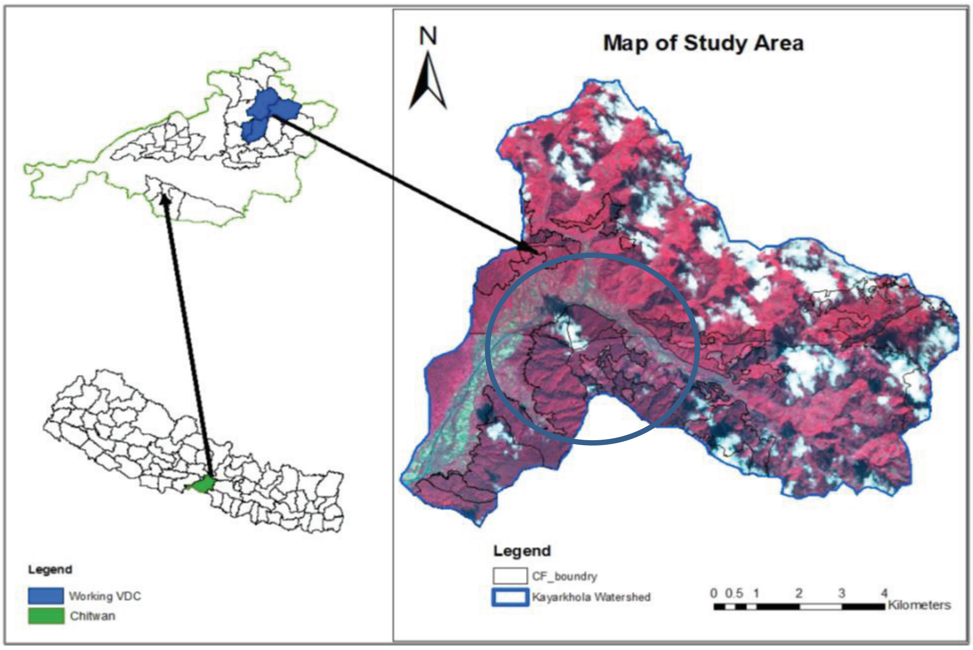


Figure3. Study area

2.1.2. Dataset

For this research and field work, Worldview-2 high resolution satellite imagery (multispectral 2m and panchromatic 0.5m) obtained on 25th October 2010 and small footprint airborne Lidar data (0.5-2 points/m2) obtained in March 2011 were used. Data were already pre-processed. Topographic map was also used in the field during data collection.

2.1.3. Software and tools

To complete this project, analyse the data and write the thesis paper following software (table 1) were used.

Table1. List of software used in the research

Software	Purpose of usage
ArcGIS 2010	GIS analysis
ERDAS Imagine 2011	Image analysis and classification
eCognition	Segmentation and classification
R software	Statistical analysis
Microsoft Excel	
Adobe Acrobat Professional	Thesis Writing
Microsoft Office	
MS Visio	
Endnote X5	

For soil analysis in laboratory following materials were used in laboratory.

Table2. List of soil analysis methods and used chemicals

Method	Chemicals	Other equipment	comments
WB	Potassium dichromate standard solution, con. Sulphuric acid (96%), con. Phosphoric acid(85%), Barium biphenyl sulphonate , 0.16%, Ferrous sulphate solution	Beaker, Burette, safety pipette, illuminated magnetic stirrer, Measuring cylinder 25 ml	-
LOI	-	Furnace or oven, crucible, desiccator, analytical balance, electric drying oven	Electric drying oven heat should be regulated to a constant temperature of 105 C.

2.1.4. **Sampling design**

A stratified random sampling design was adopted. One of the important variables of this study was elevation. When it changes to higher altitude all other climatic variables also change. In field surveys, when subpopulations within an overall population vary, it is convenient to sample each subpopulation independently. Stratification is the statistical sampling approach of dividing members of the population into homogeneous subgroups or strata. After dividing into these strata, simple random sampling was applied within each statum.to improve the representativeness of the total sample as well as to reduce the sampling error.

The whole study area was divided into 5 elevation strata, each stratum covers a 200 m interval. Given the shape of mountains (pyramid shape) to keep the density points equal, the number of sampling points per strata varies. The actual number of sampling points per elevation stratum was determined by using the following formula adopted from the Community Forest Inventory Guideline of Nepal (DoF, 2010)

Equation 1. Determination of sampling plot number

$$Area\ of\ sampling(m^2) = sampling\ intensity(\%) \times Total\ area\ of stratum(m^2)$$

$$No\ of\ plot(n) = \frac{Area\ of\ sampling(m^2)}{Area\ of\ one\ sample\ plot(m^2)}$$

The predicted and collected numbers of plots per stratum are shown in table 3 considering 0.5% sampling intensity and 500 m^2 per plot area.

Table3. Number of sampling plot

Serial No	Elevation Stratum Description	Area (m^2)	Necessity Number of plots	Collected number of plots
1	200 – 400 meter	1266431.5	5	9
2	401 – 600 meter	3141268.25	31	31
3	601 – 800 meter	1888853.75	18	21
4	801 – 1000 meter	308841.5	3	5
5	>1000 meter	357.75	0	0
	Total	6605752.75	57	67

2.2. Methods

The whole methodology of this work are divided into seven segments (Figure 4) to describe it in a logical order, such as-

i) Part-A: Tree and soil parameters related data collection from the study area.
ii) Part-B: Soil sample analysis in the laboratory to extract soil organic carbon data
iii) Part-C: Worldview-2 image classification for species and litter quality related data.
iv) Part-D: Canopy Height Model (CHM) preparation from LiDAR data to extract Canopy Projection Area (CPA) and height information and DEM for elevation information.
v) Part-E: Regression model development to estimate AGB for study area.
vi) Part-F: Stepwise regression to select best fit model to estimate SOC.
vii) Part-G: Soil Organic Carbon (SOC) map preparation.

2.2.1. **Field work for tree and soil parameter data collection (Part-A, Figure-4)**

2.2.1.1. Pre-field work

First of all, a schedule for field data collection was prepared for fieldwork. After that the following work was done:

- A Field data collection sheet was prepared before leaving to field.
- All the necessary field equipment was borrowed from the ITC field equipment section.
- The provided images of Worldview-2 were already pan-sharpened. The pan-sharpened image was converted into ECW format to upload in iPAQ and made ready for navigation.
- For the identification of recognizable trees on the image in the field, Worldview-2 image of every plot with its surrounding areas was printed on paper as well.

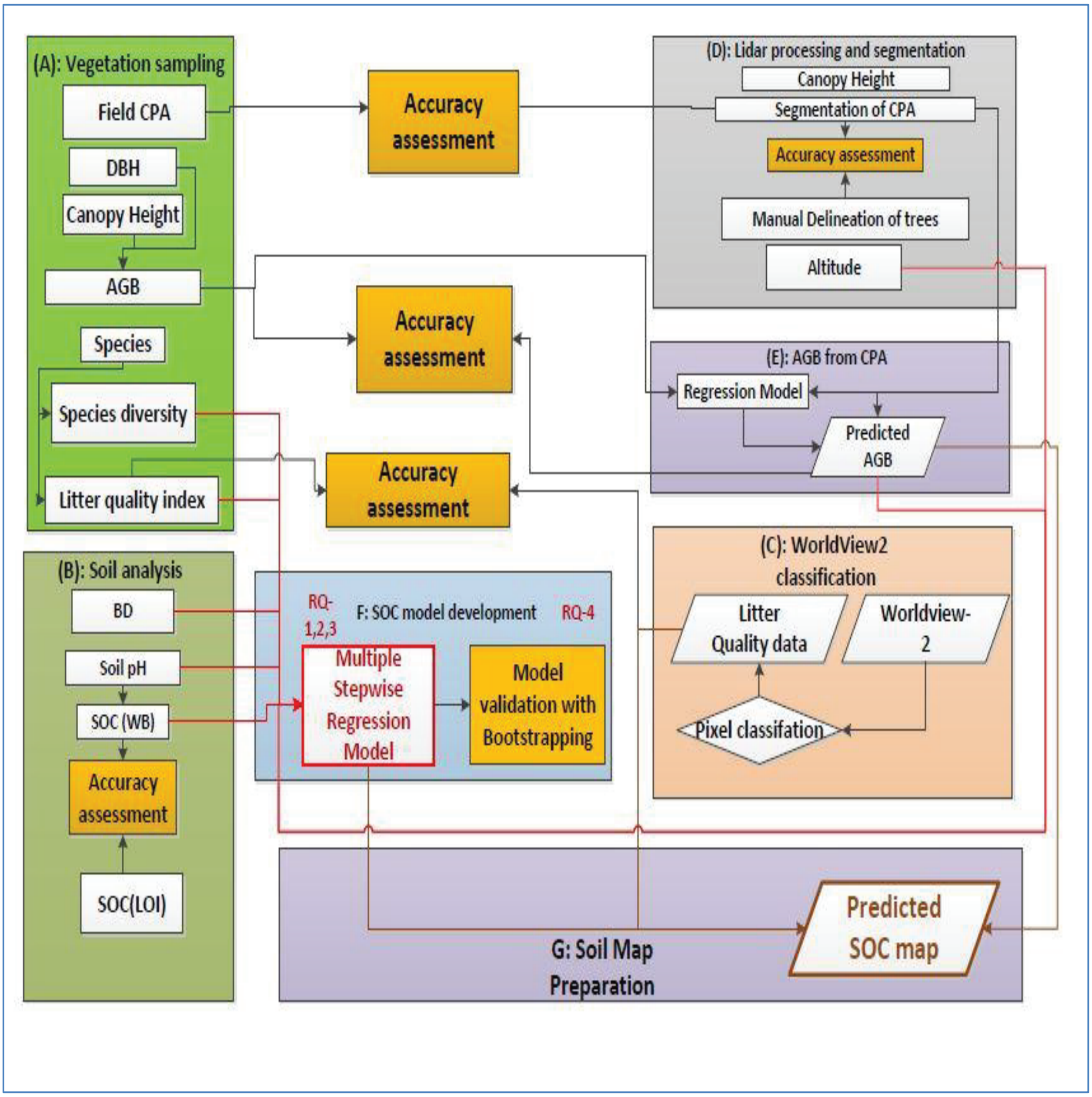


Figure4. Flowchart of research methods and steps.

2.2.1.2. Plot layout for AGB, species identification and soil sample collection

Rectangular and circular plots both are used for forest carbon estimation. In this study circular plots were used as they are relatively easy to establish. As illustrated in Figure- 5, three concentric plots were

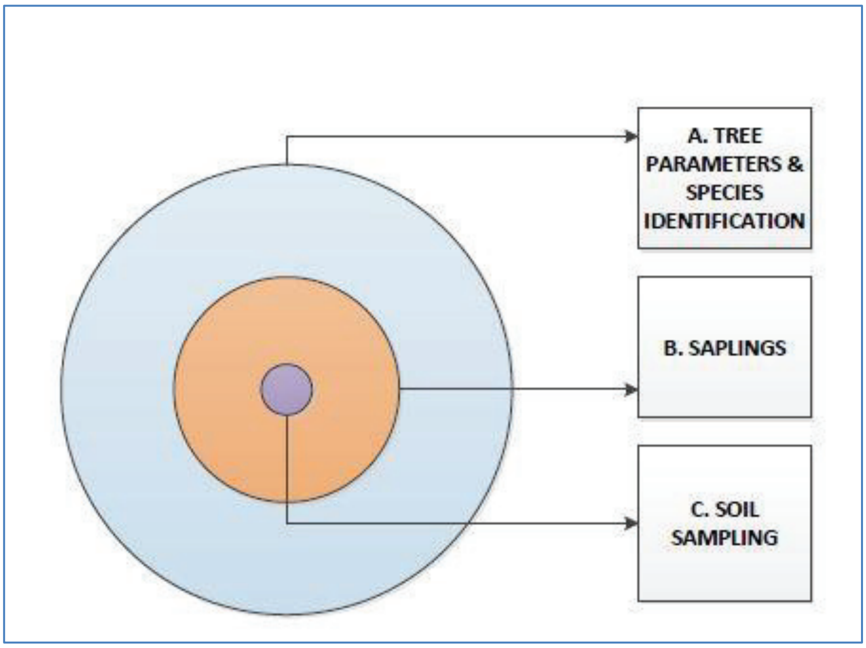


Figure5. Design of the sampling plot

established within each plot for specific purposes: inside of the 12.62 m radius plot (Figure 5, Plot A), a sub plot with a 5.64 m radius (Figure 5, Plot B) is established for saplings and a sub-plot (Plot C) is established for collecting soil sample. All soil sample were collected from the centre of the plot. The main plot was 500 m² (Figure 05, plot A) and second concentric plot was 100 m² (Plot B) respectively.

2.2.1.3. Measuring tree parameters

The XY coordinate of the centre of each plot was located using an iPAQ. Within each main plot, only trees with a DBH of 10 cm or greater were measured because trees with less than 10cm have a small contribution to the total biomass of a forest (Brown, 2002). The following tree parameters: DBH, tree height, crown cover and crown diameter were measured in the sample plot. DBH and height were measured to estimate the biomass of individual tree through allometric equation. Crown diameter was measured to calculate the CPA of tree. For Shannon Diversity Index analysis, all the species within the plot were identified and noted on the sheet. Slope correction has been done in the areas more than 5 ° slope during the measurement of plot radius and crown diameter.

2.2.1.4 Method of Tree Diversity analysis

Tree diversity analysis of each sample plot was calculated using 'Shannon Diversity Index' (Magurran. E, 1988) (Figure 4, Part -B). The formula of Shannon Index:

Equation 2. Shannon Diversity Index

$$H' = -\sum p_i \times ln p_i$$

Where

H' = Shannon Diversity Index

Pi = the proportion of individuals belonging to ith species and

ln = natural log (base 2.718)

The number of tree species with DBH more than 10 cm, DBH less than 10 cm to 5 cm and DBH less than 5 cm was recorded in each plot. In same way, seedlings data were recorded in the tally sheet in order to calculate the proportion of individual species in each plot.

2.2.1.7 Methods of litter quality determination

As the study area is mostly dominated by sal (*Shorea robusta*) forest, species was classified into two groups *Shorea robusta* and other mixed species to make a relationship with other variables in regression analysis. Total number of species within a plot was recorded. In an individual plot where *Shorea robusta* is more than 70%, it was treated as mono species dominant. On the other hand when it was less than 70%, that plot was counted as a mixed species plot. Mono species (>70% sal tree) litter was coded as 1 and mixed species litter was coded as 0. The following rule (Table 4) was used to correlate species % with litter quality index.

Table4. Dummy variables code and their rules.

Class Name	Species richness	Code
Litter quality	More than 70% *Shorea robusta (sal tree)*	1
	Less than 70% sal or the combination of mixed species	0

2.2.1.8 Sampling method for soil bulk density and soil organic carbon

To estimate the bulk density, individual soil sample of 100 cm^3 (from: 0-10cm depth) was collected with the help of standard metal soil sampling core (100 cm^3 volume)(ANSAB. et al., 2010). In similar way, three separate samples were collected from the centre of the plot. After collecting the soil sample with metallic core, it has weighed and air dried. The dry sample with core was weighed and the empty core was weighed as well. Separately empty core was weighed also. Then soil bulk density was calculated by applying the following formula:

Equation 3. Soil bulk density

$$Soil\ bulk\ density\left(\frac{gm}{cm^3}\right) = \frac{air\ dry\ weight\ of\ soil\ with\ metal\ core(gm) - weight\ of\ empty\ core(gm)}{Volume\ of\ soil(cm^3)}$$

The average bulk density value of three samples was treated as the bulk density of this plot.

After measuring the bulk density, a composite soil sample was prepared by mixing the three separate samples of each plot. At first sample were sieved to remove the clods, gravel more than 2 mm dia. Finally samples were carried to ITC soil laboratory for further analysis.

2.2.2. Methods of laboratory analysis for soil organic carbon (Part -B: Figure 4)

Most commonly used methods of SOC determination are: a) Walkley –Black Method and b) Loss On Ignition Method (Brian. A. S., 2002). In this study, both methods were used (Figure: 4, Part- B) to measure the carbon content from forest soil. A brief description of these two methods are presented below

2.2.2.1 Loss On Ignition (LOI) method

Destruction of all organic matter in the soil or sediment through heat is the main principal of the Loss On Ignition (LOI) method to determine the organic matter content. In this method, a sample (known weight) was dried at 105°c to remove the moisture and weighed again. Afterwards it was placed in a ceramic crucible or similar vessel which is then heated to between 550°C - 600°C overnight or for 12 hours

(Nelson. et al., 1996). The heated sample is then cooled in a desiccator and weighed again. The organic matter content is calculated by using the following formula

Equation 4. SOC% determination from LOI method

$$SOC\ \% = \frac{a-c}{b-c} \times 100$$

Where
a = final weight (gm) of crucible and ash
b = weight (gm) of crucible and sample
c = weight (gm) of empty crucible

2.2.2.2 Walkley-Black Method

The Walkley-Black (WB) procedure involves a wet combustion of the organic matter with a mixture of potassium dichromate and sulphuric acid. The residual dichromate is titrated in this method against ferrous sulphate. Finally to compensate the incomplete destruction, an empirical correction factor of 1.3 is applied in the calculation of result.

2.2.2.3 Quality Control Sample:

To test the precision of methods quality control samples were tested also. In this study, two types of 'quality control' measured. One is related to field sample known as 'field duplicate'. Another one is 'laboratory duplicate. During sample collection 10% samples were collected as a field duplicate sample from same plot and same location. In laboratory analysis, laboratory duplicate was prepared for both methods to test the reliability of analysis method. Laboratory duplicate means to divide the one sample into subsamples for laboratory analysis.

2.2.2.4 Accuracy assessment of soil analysis result

To test the accuracy of soil sample analysis result, a regression test was done between the results of LOI methods and WB methods. Based on R^2 value the relationship between these two methods was assessed. To see the precision of analysis, LOI method was compared with 1:1 fitted trend line. RMSE value was also calculated to know the errors rate.

2.2.2.5 Conversion of SOC fraction to total soil organic carbon (kg/m^2)

SOC fraction needs to transfer into Total soil organic carbon (TOC) with the value of sampling depth and bulk density. So it can be measured by a unit like kg/m^2. This transformed value or TOC (kg/m^2) is used for further analysis like stepwise regression with other remotely sensed variables, stepwise regression and model development.
The following formula was used to calculate the mass of carbon in single plot (Centre for Standardization and Environment, Ministry of Forestry, 2011):

Equation 5. Conversion of SOC to TOC

$$C_T = C_F \times D \times V$$

Where, C_T = Total carbon for the layer in metric units.
C_F = Fraction of carbon (Percentage carbon divided by 100)
D = Density (Bulk Density)
V = Volume of the soil layer in cubic meters (area of the plot multiplied by the depth of sampling).

2.2.2.6 Soil sample analysis for soil pH

Though soil pH is not a remotely sensed variable but it was measured using a pH meter. It was included in the regression analysis together with other variables to know the effect of pH on the soil organic carbon content.

2.2.3. **Image classification and litter quality data extraction (Part-C, Figure-4)**

2.2.3.1 Image classification process

The principal idea behind pixel based image classification is that a pixel is assigned to a class based on its feature (Annonymous, 2011). The process and steps are summarized in figure 6. Using the ground data collected from the field, a supervised classification was performed.

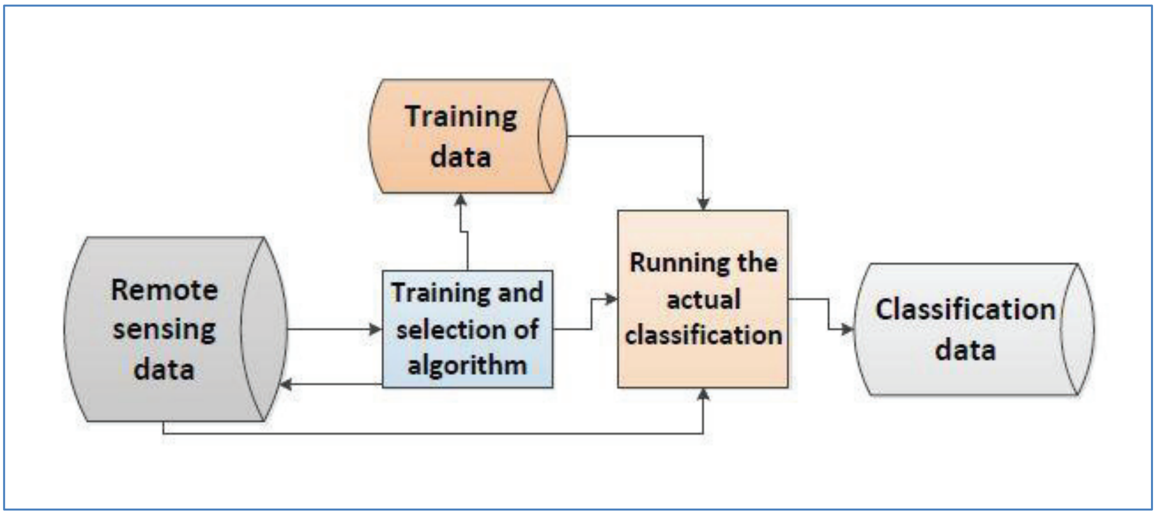


Figure6. Represents the classification process, adapted from(Annonymous, 2011)

Pixel based supervised classification technique used to classify the image. To classify the image in ERDAS, different band combination was tested. Finally 754 band combination was selected to visualize the image for classification. This band combination differentiates the dominant species *Shorea robusta* from other minor species like *Lagerstromia parviflora, schima wallichii, Semicarpus anacardum* etc. There are different types of classification algorithm used in ERDAS like box classifier, minimum distance to mean classifier and maximum likelihood classifier. In this study, classification process was carried out using the worldview -2 image and the maximum likelihood classifier approach. The advantages of maximum likelihood (ML) classifier is that it considers not only the cluster centres but also the shape, size and orientation of the clusters.

2.2.3.2 Accuracy assessment for result validation

Preparation of a confusion matrix or error matrix table is the process of accuracy assessment of image classification. This matrix table shows the overall accuracy percentage of the image. Kappa statistics value is also used for qualitative data. But overall accuracy assessment is the most cited and used measures of mapping accuracy. This is the number of correctly classified pixels and done based on matrix table. Here two parameters users' accuracy and producers accuracy helps to calculate the overall accuracy of the image. Users' accuracy indicates the error of commission (inclusion). On the other hand, producer's accuracy corresponds to the error of omission. Generally overall accuracy is counted as the total accuracy of image classification.

2.2.3.3 Transformation of image classes into dummy variables for litter quality assessment.

A classified image displayed qualitative data. This needs to be adjusted to include it together with other quantitative variables in a regression analysis. Dummy variables are used in a regression analysis for qualitative variables where qualitative variables are transferred into quantitative data through a numeric value 1 and 0. Classes of classified image were named as "Litter quality". Litter quality means either slowly decomposable litter or fast decomposable litter. The classes of image were coded into 1 and 0 values to put in stepwise regression process. The rule of determination these value already discussed in section 2.2.1.7.

2.2.4. **CHM for biomass extraction and DTM for elevation (Part-D, Figure-2)**

To find the exact location of a tree, a canopy height model (CHM) was extracted from LiDAR data. By subtracting the Digital Surface Model (DSM) from Digital Terrain Model (DTM), a CHM was derived to calculate the tree height (Figure7). Popescu. (2003) used the same method to develop a CHM. First two steps i.e DTM and DSM were extracted using LasTools software but the third step i.e CHM preparation and height calculation was done in ArcGIS by raster calculator (table 5). As an output, a CHM with 0.5 m spatial resolution was prepared which contains pixel values of the height of trees.

Table5. Represents the extraction methods of CHM and contributing software list.

Steps	Command	Software	Output
1	blast2dem -i cloud_ points.las -o-sub_ dtm .tif -v -step 0.5 -keep_ class 2	LasTools	DTM
2	lasgrid -i cloud_points.las -o sub_dsm.tif –first_only -highest -step 0.5 -fill 5 -mem 2000	LasTools	DSM
3	Generate Canopy Height Model (CHM) Difference between DTM and DSM	ArcGIS	CHM

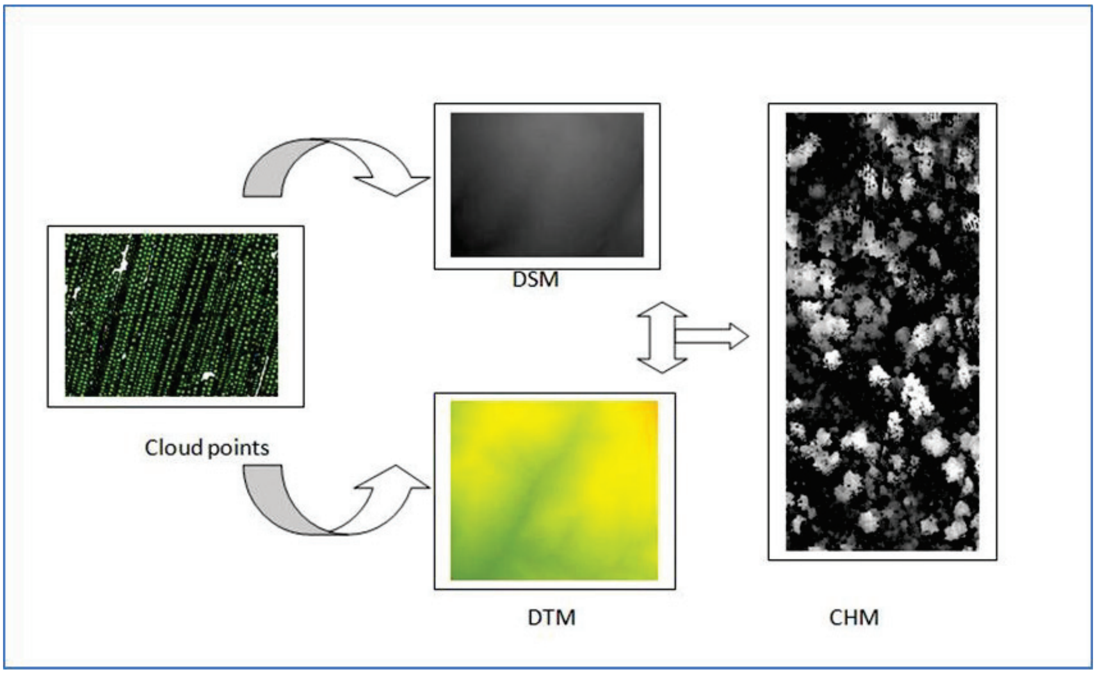


Figure7. Represents the extraction of CHM from LiDAR DTM and DSM

2.2.4.1. Segmentation and above ground biomass calculation

Object Based Image Analysis (OBIA) is a technique of image classification based on rule set. Stepwise rule set helps to refine the object. In this research, image segmentation and accuracy assessment is not a part of objectives but above ground biomass data is needed. For image classification and multi-resolution segmentation a rule set from another study was used. Lopez Bautista (2012) used the same rule set (Figure 8) for LiDAR CHM segmentation in same study area and got a good accuracy compared to segmenting an optical image. eCognition software is used to extract Crown Projection Area (CPA). This CPA was further employed to model above ground biomass.

2.2.4.2 Manual tree crown delineation

The aim of the manual delineation is to validate the segmentation done by e Cognition software. About 204 trees were identified in the image. For this purpose, 99 trees out of identified trees were manually delineated on both the Worldview and CHM images. As LiDAR CHM was used for segmentation, the reference 99 tree polygons from CHM were used to validate the segmented map.

2.2.4.3. Segmentation accuracy

As the CHM was segmented using secondary rule set, the accuracy assessment was very important. There are many methods for segmentation accuracy assessment. Two accuracy measures named one to one (1:1 matching) and distance index (D) value were carried out for accuracy assessment. 1 to 1 matching means the visual interpretation of reference polygon with same tree segmented polygon. Moller et al. (2007)developed the visual assessment technique. Here accuracy is determined by comparing the objects overlap with the reference. According to Zhan. et al. (2005) when manual delineated objects are overlapping by at least 50%, it means that objects take right position, size and shape and can be considered as right match.

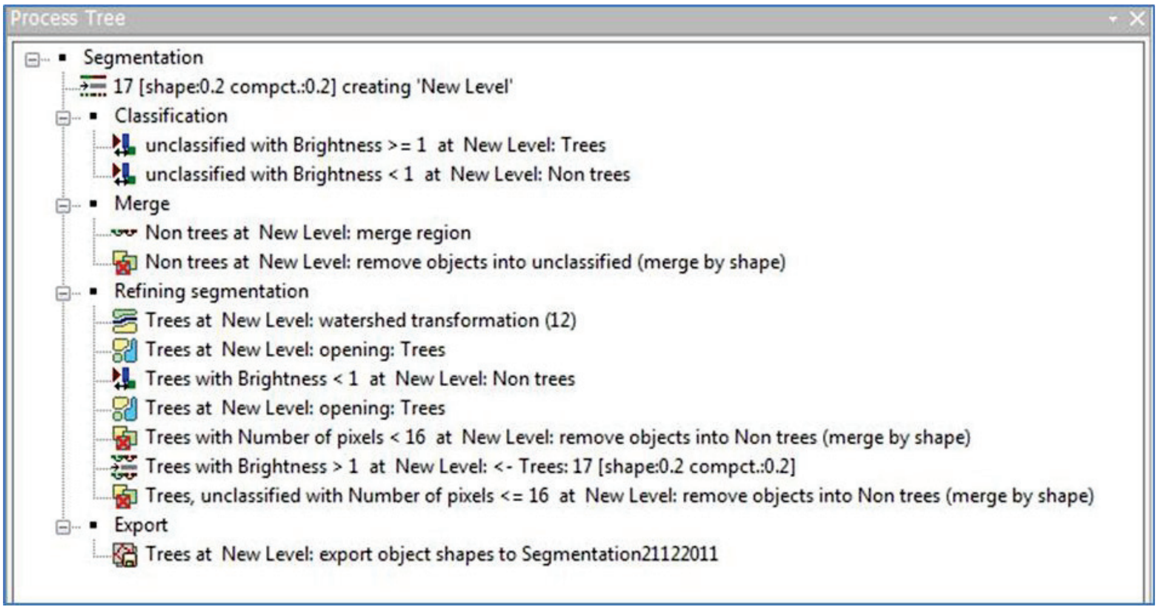


Figure8. Rule set used in LiDAR CHM segmentation, adapted from (Lopez Bautista, 2012)

The other method is geometrical assessment which one is the comparison of segmented objects with training /reference objects in terms of various indices. Clinton et al. (2010) developed this method based on over segmentation (Equation 7) and under segmentation (Equation 8). Afterwards the over and under segmentation calculate the 'Distance' index or D value to indicate the quality or accuracy of segmentation. Standard D (Equation 6) value is 0 means perfect matching. It ranges from 0 to 1 based on the accuracy of segmentation (Clinton et al., 2010).

Equation 6. Measure of closeness or D value determination

$$Dij = \sqrt{\frac{Oversegmentation_{ij}^2 + Undersegmentation_{ij}^2}{2}}$$

Over segmentation and under segmentation values were calculated by using the following formula

Equation 7. Calculation of over segmentation

$$Oversegmentation_{ij} = 1 - \frac{area(x_i \cap y_j)}{Area(X_i)}$$

Equation 8. Calculation of Under segmentation

$$Undersegmentation_{ij} = 1 - \frac{area(x_i \cap y_j)}{Area(Y_j)}$$

Where
x_i= is the training objects or reference polygons, relative to which the segmentation to be judged
y_j= is the set of all segments in the segmentation

2.2.5. **Regression model development for biomass estimation**

Regression implies the cause and effect relationship between dependent and independent variables which means that changes in the independent variable can produce changes in the dependant variable (Husch et al., 2003).
In this study area, site specific allometric equations are not available, so the equation developed by Chave et al. (2005) for tropical moist forest was used to calculate AGB. Same equation was also used by REDD+ pilot project (ANSAB. et al., 2010). The equation is as follows:

Equation 9. Above ground biomass calculation

$$AGB = 0.0509 * \rho D^2 H$$

Where
AGB = above ground biomass (kg)
ρ = wood specific gravity (gm/cm3)
D = tree diameter at breast height (DBH) (cm) and
H = tree height (m).

One important parameter of this allometric equation is wood specific density (kg/m3). Sharma and Pukkala (1990) build up a model to determine this value. In table 6, wood density of different species determined by (Sharma & Pukkala, 1990) is presented.

Table6. Wood density of major tree species and others common species.

Species Name	Wood dnsity (kg/m^3)
Shorea robusta	880
Lagerstroemia parviflora	850
Schima wallichii	689
Miscellaneous	720

According to the allometric equation (Equation 9), DBH is another sole input of biomass estimation. To estimate the above ground biomass for the whole study area a regression model was developed to predict DBH based on Crown Projection Area (CPA). At first, the field derived DBH was plotted against field extracted CPA using different types of linear models to see which best fits the relationship between these two variables. At a same time field derived DBH was compared with LiDAR CHM segmented CPA to predict the DBH. This predicted DBH was used in the mentioned allometric equation to estimate the biomass per tree. Then total plot biomass was calculated to estimate the biomass in kgm^{-2}.

2.2.5.1. Accuracy assessment and model validation for DBH

The reference trees which are easily recognized in the image were used to validate the model of predicted DBH. Out of 99 recognized trees in the image, only 69 trees were correctly matched and used for model development. This data set was divided into 70% as training data and rest 30% as validation data. The model developed from training data set was validated by rest 30% independent data set to calculate the RMSE. It is a tool of validation for predicted model and calculated by the following formula

Equation 10. Root mean square error calculation

$$RMSE = \sqrt{\sum \frac{(DBH_p - DBH_O)^2}{N}}$$

Where
DBHp = Diameter at breast height (predicted from LiDAR data)
DBHo = Diameter at breast height (observed from field data)
RMSE= Root Mean Square Error
N= Number of observations

2.2.6. Accuracy assessment of biomass estimation

Above ground biomass was estimated by using equation-9. Here DBH is predicted based on segmented CPA and height is extracted from LiDAR CHM. As both parameters are estimated from two types of input with two types of uncertainty, a regression test was done between the field observed biomass and predicted biomass. Based on R^2 and RMSE value, the strength of the estimated biomass was judged either it is effective or not to develop a model.

2.2.7. Multicollinearity test for all variables

When two or more predictor variables in a regression models are highly correlated is called multicollinearity. Due to collinearity issue, the coefficient estimates of variables may be changed erratically in response to a few change in the predictive model. But it never reduces the predictive power of the model (Charlotte. H et al., 1991). If there is any collinearity among the variables still model can be reliable but it depends on the degree of collinearity. Because collinearity itself is assessed by the degree of correlation, not only in terms of presence or absence. There are many methods in literature to test the collinearity. Commonly practiced methods are observing the correlation matrix or Variance Influence Factor (VIF). In this study both methods were practiced to judge the collinearity issues.

2.2.8. Stepwise linear regression (Part-F, Figure 4)

Regression quantifies the relationship between dependent and independent variables. A common problem in regression analysis is that of variable selection. Normally a potential large number of independent variables are related with one response variable. But to create a best fit model most important variables are selected based on automated procedure. Stepwise regression is an automatic procedure to select one or more independent variables against one dependant variables. It may be either backward regression or forward regression. Six important variables named above ground biomass (AGB), elevation, species diversity, litter quality, soil bulk density and soil pH were selected as predictors for soil organic carbon. In both cases, finally independent variables are selected based on Akaike Information Criterion (AIC) and p value.The Akaike information criterion (AIC) used in statistics to select the model and it is a measure of the relative goodness of fit of a predictive model. Among the predictive models, model with minimum AIC value is treated as the best fit model.In this study, all variables (AGB, elevation, species diversity, litter quality, soil pH and soil bulk density were used in stepwise regression analysis to select the best variables based on AIC value to estimate SOC.

2.2.8.1. Model validation

A straightforward and fairly popular approach of model validation is to randomly split the training data in two parts: one to develop the model and another to measure its performance. With this split-sample approach, model performance is determined on similar, but independent data. A more sophisticated approach is to use cross-validation, which can be seen as an extension of the split-sample method. With split-half cross-validation, the model is developed on one randomly drawn half and tested on the other and vice versa. The most efficient validation has been claimed to be achieved by computer-intensive resampling techniques such as the bootstrap. Bootstrapping replicates the process of sample generation from an underlying population by drawing samples with replacement from the original data set, of the same size as the original data set. Then the coefficient of regression and confidence interval were calculated for the selected model.

Bootstrap and the jackknife, estimate the variability of a statistic from the variability of that statistic between subsamples rather than from parametric assumptions (Michael. et al., 2011). And both methods are used to validate the predicted model and they produce similar result. It is subsampling process with a random draw from the original sample with replacement. On the other hand, jacknife is a subsampling from the original dataset without replacement. Bootstrap is a computer based way of estimating standard errors, biases and confidence intervals for any statistics. A bootstrap sample is a random sample of size n taken with replacement from x

Equation 11. Bootstrap replication

$$X^* = (X^*1, X^*2, X^*3 \ldots\ldots\ldots\ldots\ldots, X^*n)$$

It was fitted 1000 times a model on 1000 different stochastic realisation of the original dataset which contain 61 observations. After bootstrapping, predicted soil organic carbon was compared with the field observed soil organic carbon.

2.2.9. Preparation of predicted soil organic carbon map (Part-G, Figure 2)

In this case, variables that are be available as spatial data are selected as a predictor for SOC. The spatial values (i.e raster value) of those selected variables (predictors) are included in the selected model to make a map of the spatial distribution of SOC.

3. RESULTS

3.1. Descriptive statistics related to tree parameter:

Total study area was 660.60 ha. In total number on 1228 trees data was collected on 67 plots. The seven most dominant species are presented in figure 9, representing 91% of all trees recorded. There are 13 other species representing the remaining 9% (Figure 9).

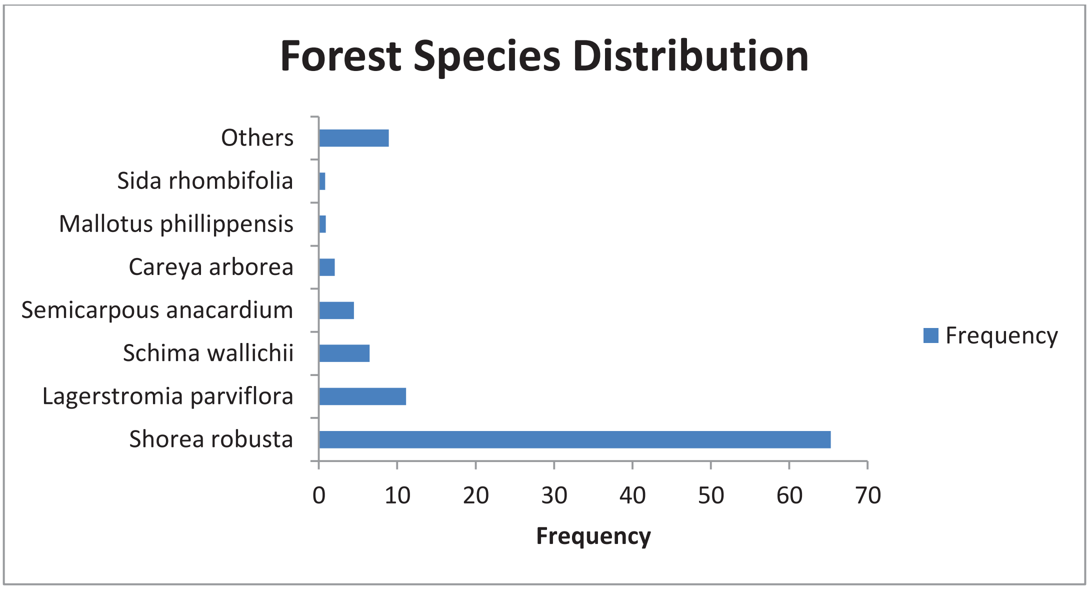


Figure9. Distribution of forest species on study area.

It's clearly observed from this graph that *Shorea robusta* is the dominant species followed by *Lagerstromia parviflora and Schima wallichii* representing 65, 11 and 7 % respectively. The average height and DBH of trees is presented in table 4 showing that *Shorea robusta* and *Semicarpous anacardium* are the tallest tree species followed by *Schima wallichi* and *Lagerstromia parviflora*. In case of DBH, *Semicarpus anacardium* and *Shorea robusta* are the species with highest DBH followed by *Schima wallichi* and *Lagerstromia parviflora* (Table 7). From the value of standard deviation it is clear that the data points tend to be very close to the mean showing a low variability from height values and DBH values.

Table7. DBH and height of the major species observed in the field.

No	Species	Average DBH (cm)	SD (cm)	Average Height (m)	SD(m)
1	*Shorea robusta*	**29.4**	17.39	**13.8**	6.60
2	*Lagerstromia parviflora*	22.2	13.22	10.2	4.5
3	*Schima wallichi*	25.9	16.4	12.4	5.9
4	*Semicarpous anacardium*	**34.4**	20.9	**13.6**	7.3

Similarly Above Ground Biomass (AGB) was calculated per individual tree (using equation 9). The mean and other statistics of the major species biomass (kg/tree) are presented on table 8. It is observed that the four values are the same species that present the highest values in terms of DBH and Height of tree. This is logical because DBH and Height are the major input in the allometric equation (equation 9). Other input, wood specific density differs from species to species. For example wood specific density for *Shorea* *robusta, Lagerstromia parviflora* is 880 Kgm^{-3} and 850 Kgm^{-3} respectively. But most of the cases for minor species this value is 720 kgm^{-3}. Wood specific density of different species is presented in materials section 2.2.5 (see table 6).

Table8. Biomass estimation per species (kg/tree) derived from field data.

Species	Mean (kg)	Maximum(kg)	Minimum(kg)	Standard deviation(kg)
Lagerstromia parviflora	**374.98**	5935.12	16	962.97
Shorea robusta	**1088.11**	11198	13.43	1527.02
Semicarpus anacardium	**1263.44**	7587.05	9.16	1992.45
Schima walichii	**993.23**	11198	21.67	1698.26
Sidha rumbofolia	243.70	1761	10.99	537.28
Carea arborea	612.07	3609	14.65	915.81
Caeseria graveolens	83.64	183	31.66	57.06
For all species (combined)	**932.97**	11198	9.16	1434.30

In most cases standard deviation is very far from mean value. The higher SD value indicates that there is a non-normal distribution with a fat high end –tail. This variability of above ground biomass content is not only within the species but from species to species also.

Descriptive statistics related to tree species diversity

Species diversity is the number of equally abundant species that are represented in a collection of individuals or any population data set. Here the number of species means the number of equally-abundant species. In this study, the maximum diversity value was 1.96 and the minimum diversity value is was 0 (table 9). Individual plots where many species exist, shows higher diversity values. As the study area is dominated by *Shorea robusta*, some sample plots have very low diversity (i.e only *Shorea robusta*) and therefore a diversity value of 0. For diversity large standard deviation indicates that the data points tend to be not to close to the mean showing a high variability for diversity value.

Table9. Represents the statistical description of species diversity value

Variable	Mean	Maximum	Minimum	Standard deviation
Species diversity	0.87	1.96	0	0.49

3.2. Descriptive statistics related to soil parameter:

Three parameters related to soil named bulk density, soil pH., soil organic carbon were measured for each sample. Average value of these soil properties are presented below (Table 10) according to the elevation range. To get an overall view of the status of soil and to collect variables data for stepwise regression soil organic carbon (SOC), soil pH, and soil bulk density results were prepared. The ranges of SOC in high elevation (600– 1000) is higher to other ranges. According to Table 10, results in standard deviation (SD) indicating the values are close to mean and sows a low variability from all parameters.

Table10. Ranges of different soil parameter values at different elevation ranges.

Elevation Distance (m)	Bulk Density (gm./cm³)	Soil pH	Soil Organic Carbon (WB)	Soil Organic Carbon (LOI)
200-400	1.17-1.69	4.70-5.37	0.73 - 3.10	1.35 to 4.09
401-600	1.11-1.66	4.42-5.71	0.62 - 3.15	1.2 - 4.75
601-800	1.14-1.65	4.34 - 5.72	1.78 - 3.14	2.15 -3.95
801-1000	1.19-1.24	3.91-5.58	2.64 - 3.98	3.51 - 4.16
Mean Value	1.37	4.9	2.29	3.01
SD	0.15	0.38	0.71	0.72

3.2.1. **Laboratory analysis: Relationship between Loss on Ignition (LOI) and Walkley-Black (WB) soil carbon analysis methods.**

The comparison of SOC values from both SOC determination techniques are plotted in figure 10, showing represents that there is an overestimation of SOC values derived from LOI methods or underestimation of SOC values derived from WB method. This is clearly visible when line 1 to 1 is fitted falling below the regression line comes from those two datasets.

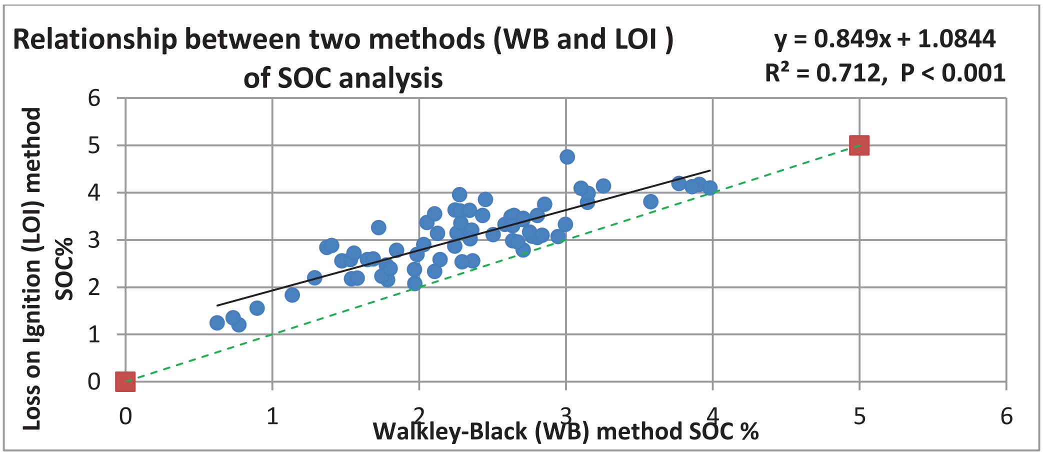


Figure10. Relation between WB and LOI method (regression line) fitted with 1:1 trend line (green).

A regression analysis indicates that there is a strong relationship between these two methods ($P<0.001$, $R^2 = 0.71$). At a same time, field duplicate and laboratory duplicate samples were also tested to know the variation. When Field duplicate samples were compared with each other, R^2 value for Walkley -Black and LOI was 0.32 and 0.66 respectively. In similar way when, laboratory duplicate sample compared with each other R^2 value for Walkley Black and LOI was 0.99 and 0.87 respectively (See annex 6). It means field condition may be varied due to different factors but in laboratory analysis WB provide more precise result compare to LOI method.

As Walkley- Black method is more accurate compare to LOI, the value of this method is transformed into Total Organic Carbon (TOC) by applying the equation 4. This transformed value or TOC (kg/m^2) is used for further analysis, i.e. stepwise regression with other remotely sensed variables and model development.

3.3. Image classification and litter quality data extraction

3.3.1. Image classification

Classification was performed using maximum likelihood classifier technique of ERDAS. Based on pixel, image was classified into three classes' i.e. two classes for species and one class for non-forest area. Tree classes are namely: *Shorea robusta* (sal forest) and all other tree species into one class named 'mixed species'. Another class of image is for non -forest area.

The classification dataset comprised of 70% as training dataset and 30% for validation. These trees are identified and randomly selected from the field. Here 67 trees were used for validation purposes to assess the classification accuracy.

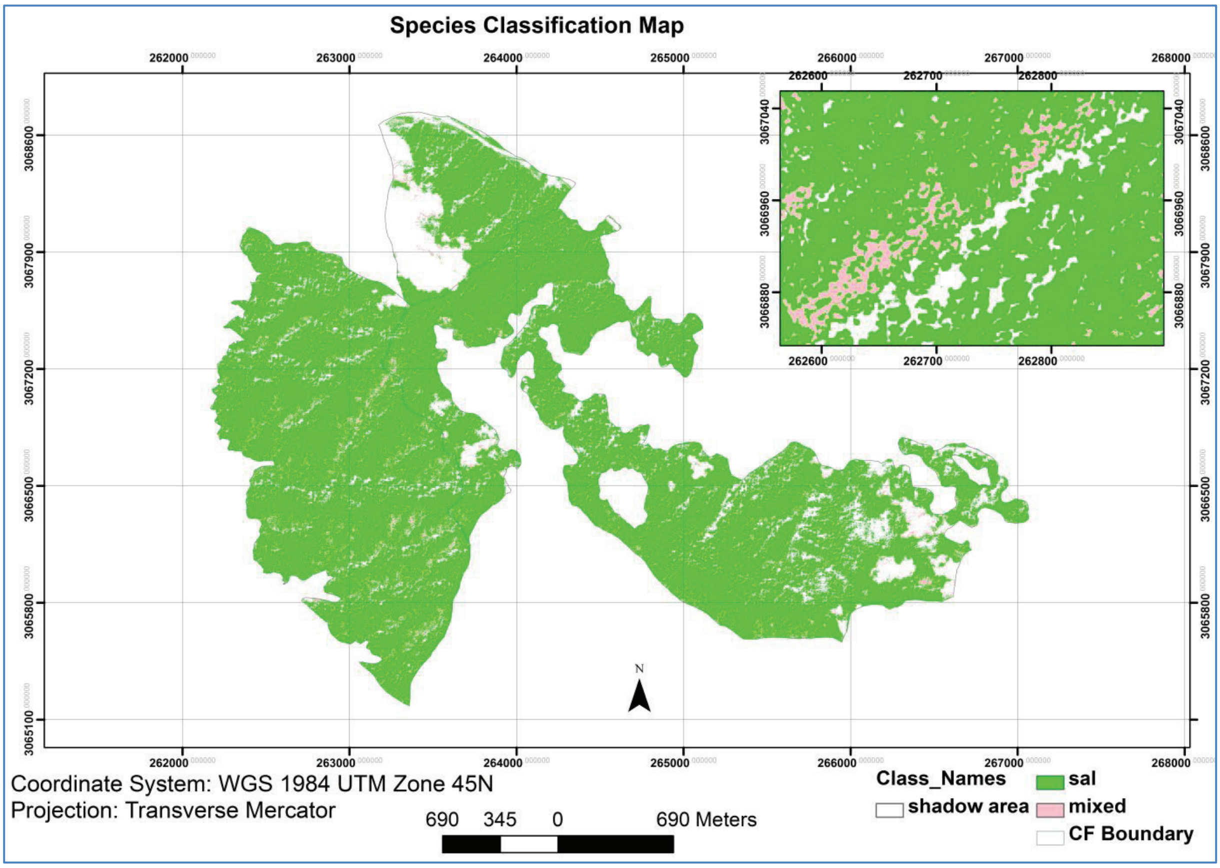


Figure11. Species classification (Litter quality index) map.

From the classification matrix table (Table 11) it is summarized that in case of sal forest class producers accuracy is higher than users' accuracy. On the other hand, in case of mixed species class, users' accuracy is higher than produces accuracy. Generally overall accuracy is counted as the total accuracy of image classification. The overall accuracy was 72.85 %.

Table11. Accuracy report for species classification map

Class name	Reference total	Classification total	Correct total	Producer's accuracy	Users' accuracy
Shorea robusta (sal tree)	37	41	29	78.37%	70.73%
Mixed species (Non sal)	30	26	20	66.66%	76.92%
Non-forest area	3	3	2	66.66%	66.66%
Total	70	70	51		
Overall classification accuracy 72.85%					

3.3.2. **Extraction of litter quality data**

Finally pixels number was counted to calculate the area for respective class. According to this calculation, sal forest (*Shorea robusta)* class occupies 489 ha area from total study area. Mixed species class are found within 37 ha. Rest 130 ha area is classified as non-forest area.

According to the code of litter quality this sal forest class will be treated as fast decomposable and mixed species class will be treated as slowly decomposable litter during stepwise regression or model selection process. This classified map is a raster output and each pixel has a value either 1 or 0. These two indices carry the quality of the litter. If pixel value indicates 1, it means sal forest (mono-species) and highly decomposable litter. Contrary, if pixel shows 0 values, it means mixed species and slowly decomposable litter. When litter are slowly decomposable, it produces more organic carbon in soil. On the other hand, if it is fast decomposed, less amount of organic carbon produced into the soil.

3.4. CHM accuracy and segmentation to extract AGB information

3.4.1. **Relationship between tree heights measured in the field and extracted from the LiDAR CHM.**

The canopy model height model was assessed for accuracy comparing with field extracted height. At 95% confidence level, regression test shows that there is a significant relationship between LiDAR estimated height and field observed height. Based on the field reference tree and LiDAR CHM a regression line was plotted to test the relationship. The scatterplot also shows the position of regression line in relation with1 to 1 trend line (dotted green line). The R^2 value is 0.71 and correlation value is 0.84 (figure 12). The roots mean square error (RMSE %) is 18.75.

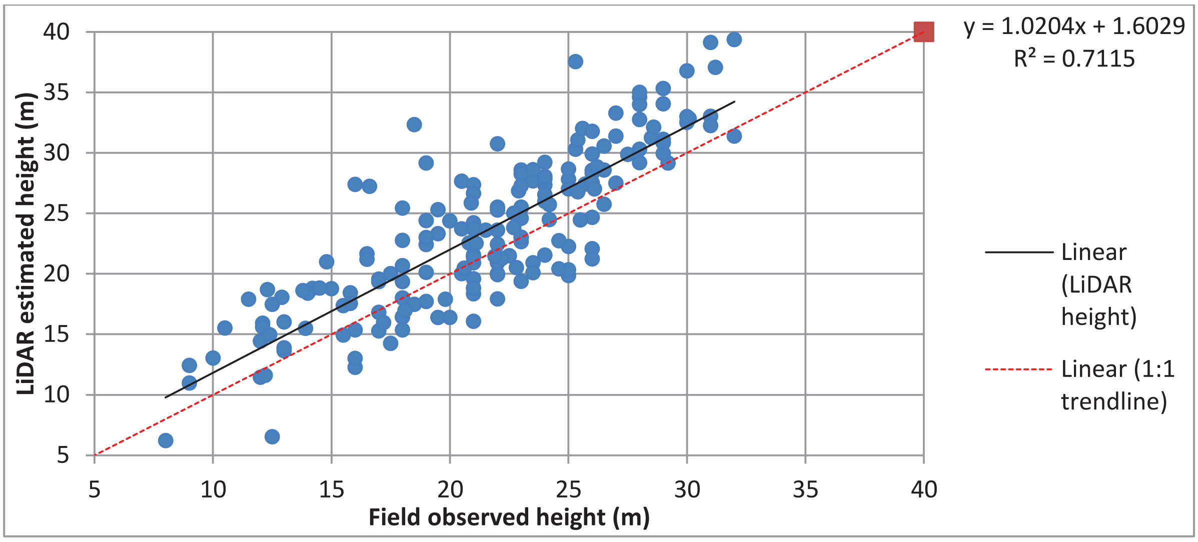


Figure12. Relationship between observed tree height and CHM height fitted with 1:1 trend line (red).

3.4.2. **Image Segmentation to estimate ABG for whole study area**

Segmentation was done based on rule set discussed on section 2.2.4.2. Segmentation accuracy was assessed by matching the manually delineated polygon with the automatic segmented one. Out of 99 references polygon only 69 polygons were correctly segmented. The reference crown and the auto segmented crown were overlapped and those crowns which overlapped at least 50% or more counted as the correctly segmented crowns. Figure 13 represents the scenario when reference polygon (blue line) overlaid with automatic segmented polygon. The 1:1 matching table (table 12) shows how many CPA were correctly segmented.

Table12. Matching of 1:1 relation of the segmented CPA with the reference CPA

Total number of 1:1 match	Total reference CPA	Correctly segmented CPA
69	99	69.69%

Table13. Results of LiDAR CHM segmentation with manual delineated reference polygon

Data source	Value	Residual D value
LiDAR CHM	Scale parameter 17	
Over segmentation	0.38	
Under segmentation	0.21	
D-value	0.30	0.70

Those 69 polygons which were correctly segmented further assessed to calculate the D value (using equation 6, 7 & 8). Table 13 summarized the segmentation accuracy through over and under segmentation value and showed how accurately CPA is segmented by using CHM model. Result from this table indicates that LiDAR CHM over segmented the image as the over segmentation value is higher than under segmentation value. The residual D value shows the accuracy and proves that 70% CPA is correctly segmented through LiDAR CHM image.

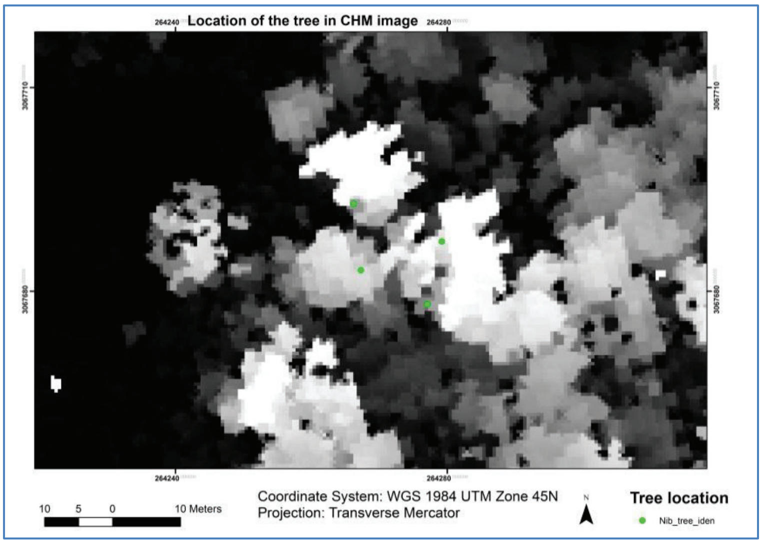


Figure 13. a : Tree location (green dot) in CHM image, black area represents the shadow area.

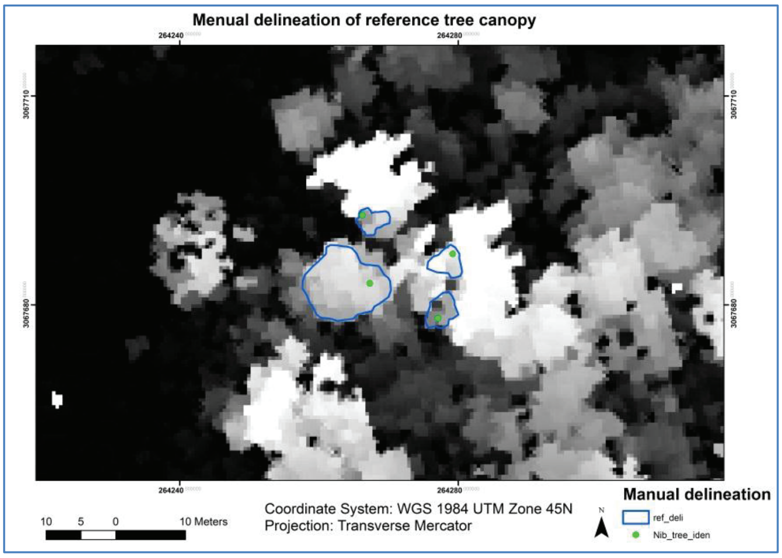


Figure 13.b: Reference tree canopy (blue line polygon) of the located tree.

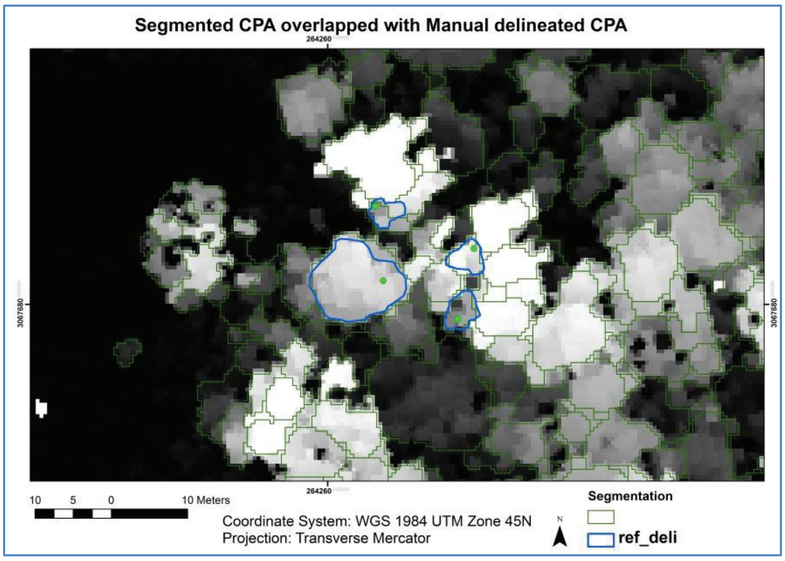


Figure 13.c: Reference polygon (blue polygon) overlapped with segmented polygon (green polygon).

Figure13. Reflects (a) tree location (b) manual delineation (c) Segmented and reference crown.

3.4.3. **Development of regression model for DBH prediction**

Four different types of models (linear, logarithmic, polynomial and power) were performed to compare the relationship between Crown Projection Area (CPA) and the diameter at breast height (DBH). All models were developed to extract the regression and RMSE value (table 14).Based on the lowest RMSE and highest R^2 value, only one model was selected to derive tree biomass from the LiDAR CHM segmented image for the entire study area.

Table14. Represents different kinds of DBH predicted model.

Data sources	Name of the model	Regression model	R2	r	RMSE	RMSE%
Field observed data	Linear	0.5744*(CPA)+20.81	0.71	0.84	8.68	16.23
	Logarithmic	26.87*ln(CPA)-52.78	0.66	0.81	9.33	17.21
	Polynomial	$0.0018*(CPA)^2+0.3746*(CPA)+25.48$	0.71	0.84	8.96	17.59
	Power	$6.19.1*(CPA)^{0.5347}$	0.68	0.82	**8.64**	**16.18**
LiDAR CHM segmented image	Linear	0.39*(CPA)+28.719	0.68	0.82	9.64	18.02
	Logarithmic	22.45*ln(CPA)-36.57	0.64	0.80	11.42	21.45
	Polynomial	-0.0004(CPA)2+0.4548*(CPA)+26.76	0.68	0.82	10.42	19.48
	Power	$9.2928*(CPA)^{0.4257}$	0.66	0.81	**9.17**	**17.14**

When the field observed DBH was tested(Figure 14) with field observed CPA, it was found that both the linear and the polynomial model had the highest R^2, both with a value of 0.71 (table 14). The power model had a R^2 of 0.68 and RMSE of 16.18%. Here power model gave the lowest RMSE. Figure 14 shows the regression line of different models when field observed DBH was compared with field observed CPA.

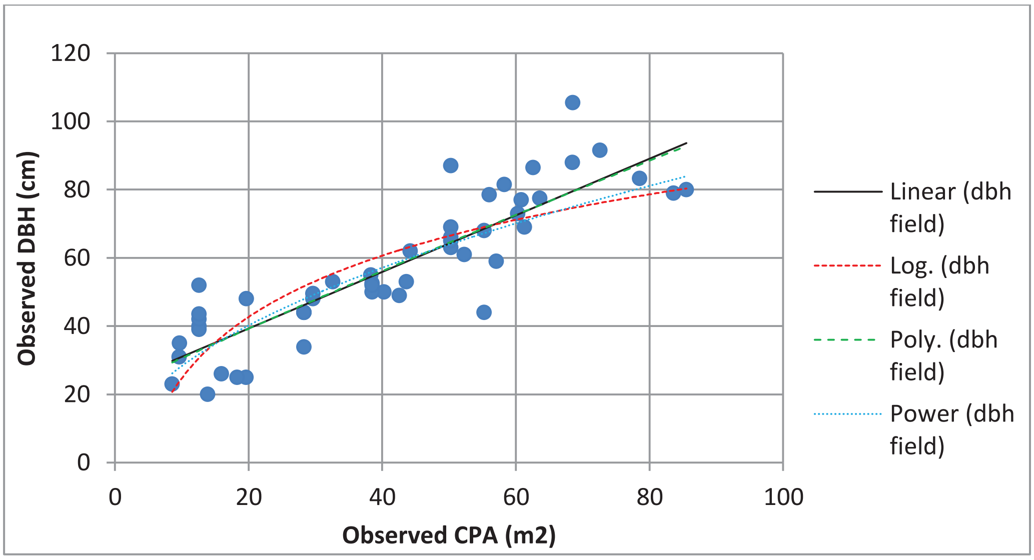


Figure14. Relationship between observed CPA and DBH from the ground data.

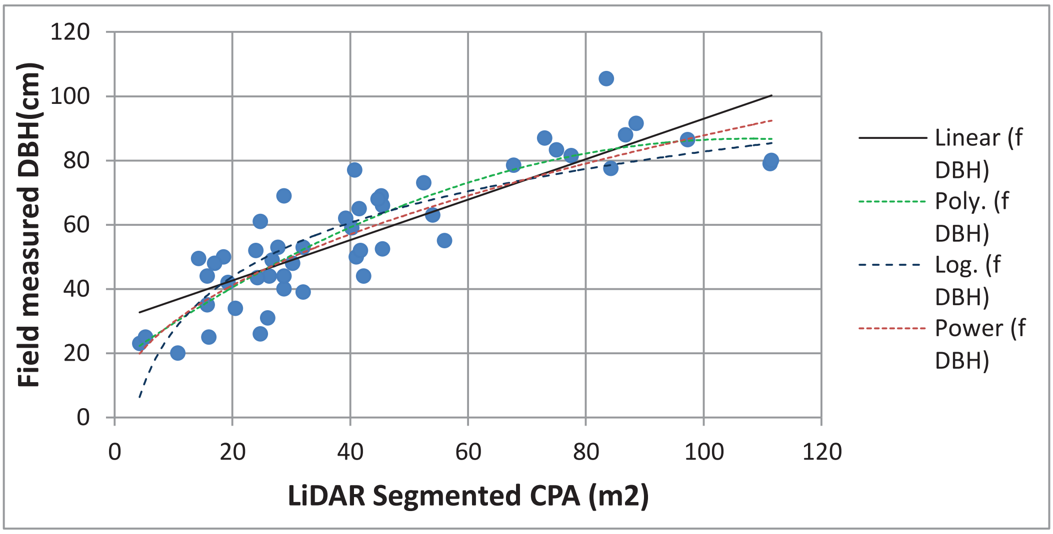


Figure15. Relationship between LiDAR segmented CPA and DBH from the ground data.

Fitting models with LiDAR CHM segmented CPA to predict field observed DBH showed that the power model has also the lowest RMSE, here 17% and an R^2 of 0.66 (Figure 15).

3.4.4. Model Validation

From table 14, it is clear that power model gives the lowest RMSE in both cases for field observation data set and as well as for LiDAR data. So power model was selected as a best model to predict the DBH by using CPA parameter. To test the accuracy of the model two different linear regressions were performed. At first field measured CPA was plotted with segmented CPA. Here validation data set was used to test the relationship. The correlation coefficient indicates (Figure 16) that they are strongly correlated.

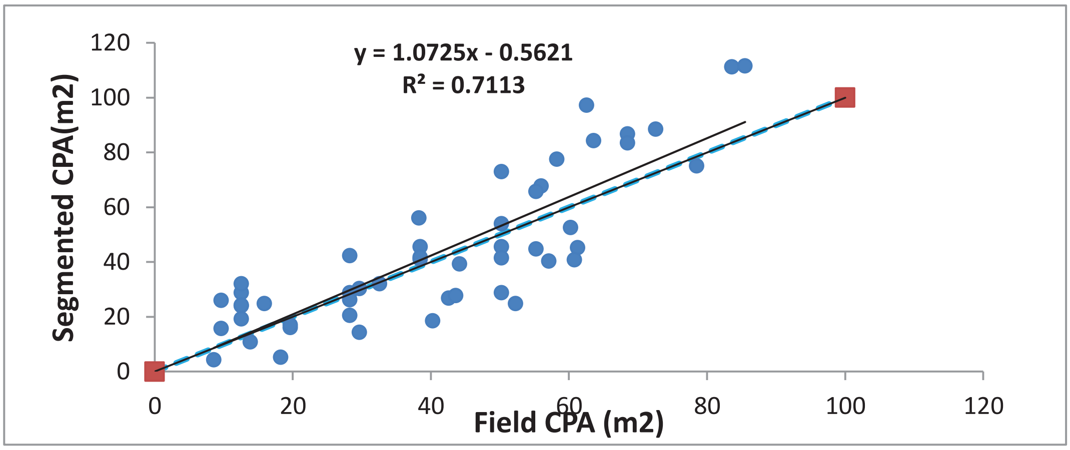


Figure16. Relation between field measured CPA and segmented CPA(m2) fitted with 1:1 trend line.

Same validated dataset was used to know the strength of the power model. Here predicted DBH using power model was plotted against the independent field observed DBH data (Figure 17). The linear regression line shows the R^2 is 0.71 and correlation value is 0.84. At 95% confidence interval, they are strongly correlated (p value<0.001).

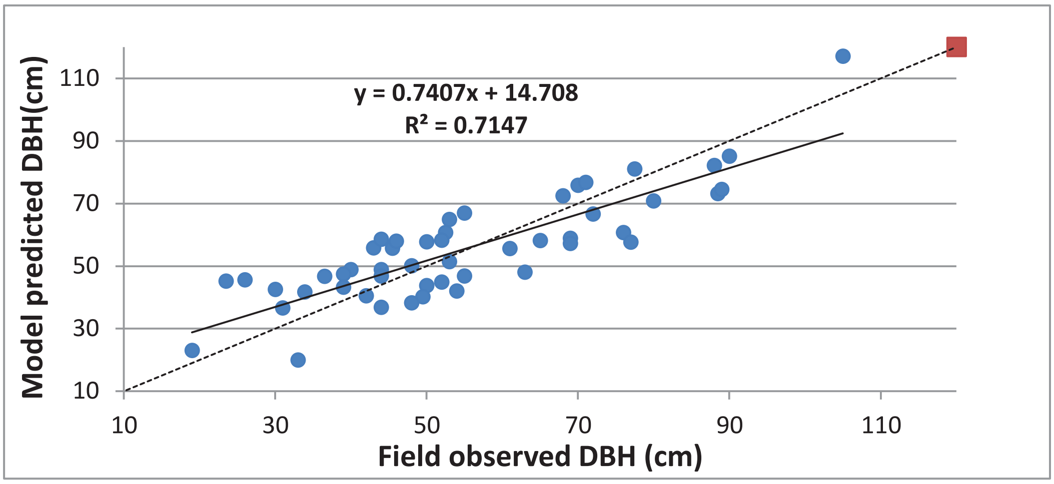


Figure17. Relationship between predicted and field measured DBH in relation with 1:1 trend line.

3.4.5. **Above Ground Biomass (AGB) estimation and map preparation**

To estimate above ground biomass for the study area equation (equation 10) was used. Here sole input DBH was replaced by the CPA developed from power regressions model. So the final equation was

Equation 12. Calculation of Above ground biomass (after adopting power model)

$$AGB = 0.0509 \times \rho \times \{9.2928 \times (CPA)^{0.4257}\}^2 \times H.$$

Where
AGB = above ground biomass (kg)
ρ = wood specific gravity (= 0.88 gm/cm³) and
CPA = Crown Projection Area (CPA) from LiDAR CHM segmented image
H = tree height (m).

The figure 18 shows the estimated amount of above ground biomass for the whole study area. Total 230 Gg was estimated in the whole study area. The mean value of biomass is 592.33 kg/tree. LiDAR derived biomass comes from two types of input with two types of uncertainity. Here DBH was predicted from segmented CPA, one uncertainity and another uncertainity is LiDAR derived tree height. So it is necessary to check to what extent these two uncertainities propagate into the final result. To validate this result a linear regression line was fitted with field observed tree biomass against the biomass estimated from the segmented image from the same reference tree. The result of this regression test is shown figure 19.

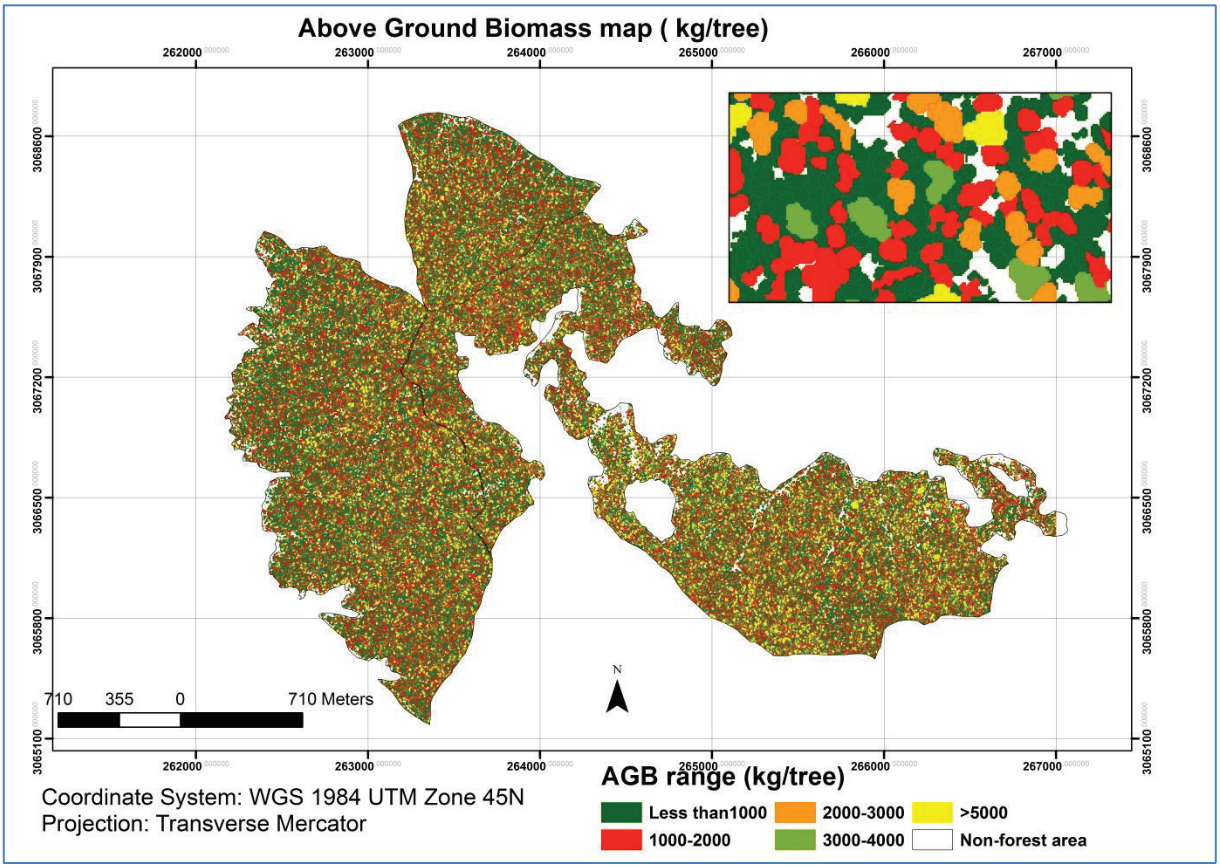


Figure18. Map of above ground biomass (kg/tree)

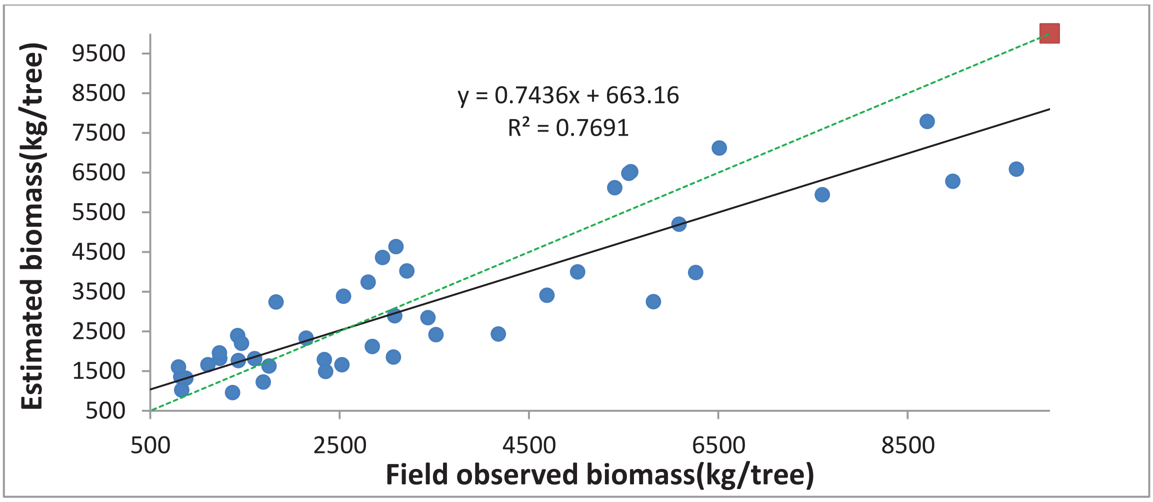


Figure19. Relation between field measured biomass and segmented biomass fitted with 1:1 trendline.

From the figure 19, it is observed that 1:1 fitted trend line crosses the regression line. It means in some cases biomass is overestimated and in some cases it is underestimated. But the $R^2 = 0.76$ represents that there is a strong correlation between these two measurement.

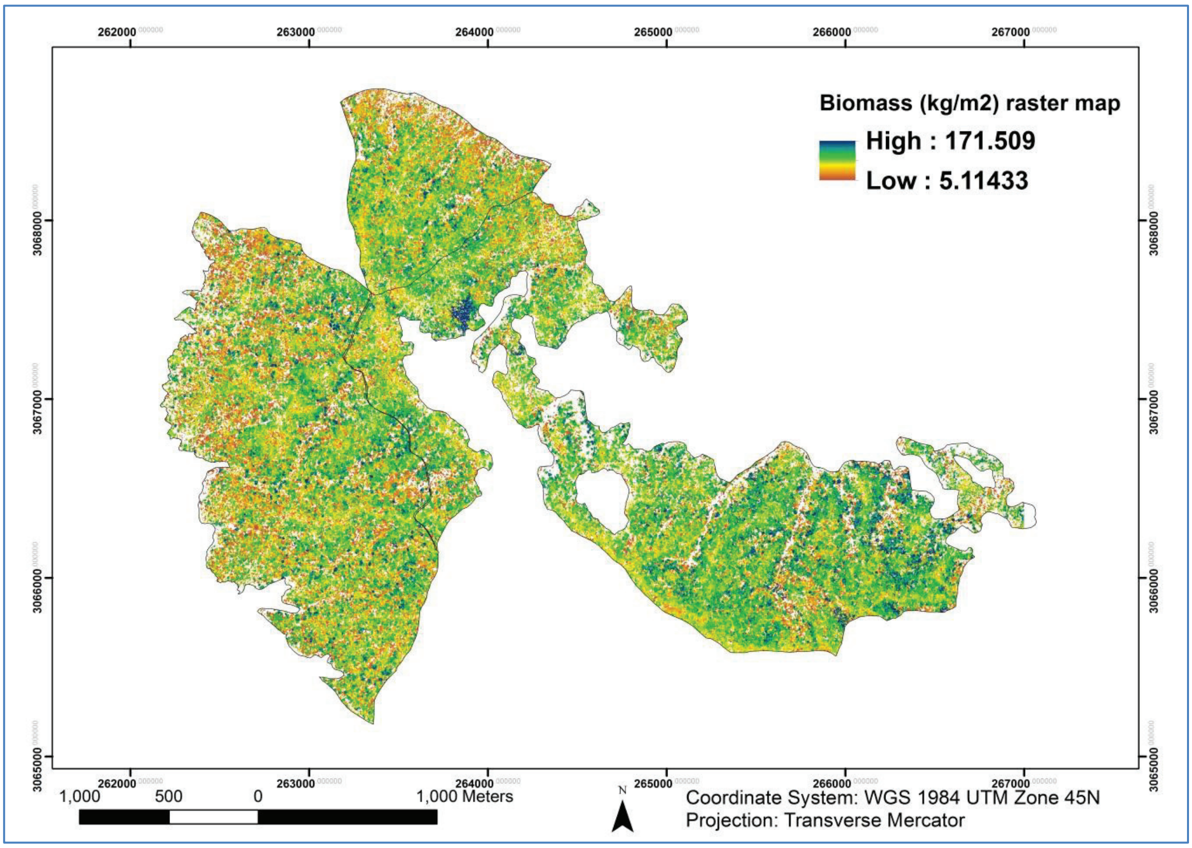


Figure20. Map of above ground biomass stock in the study area.

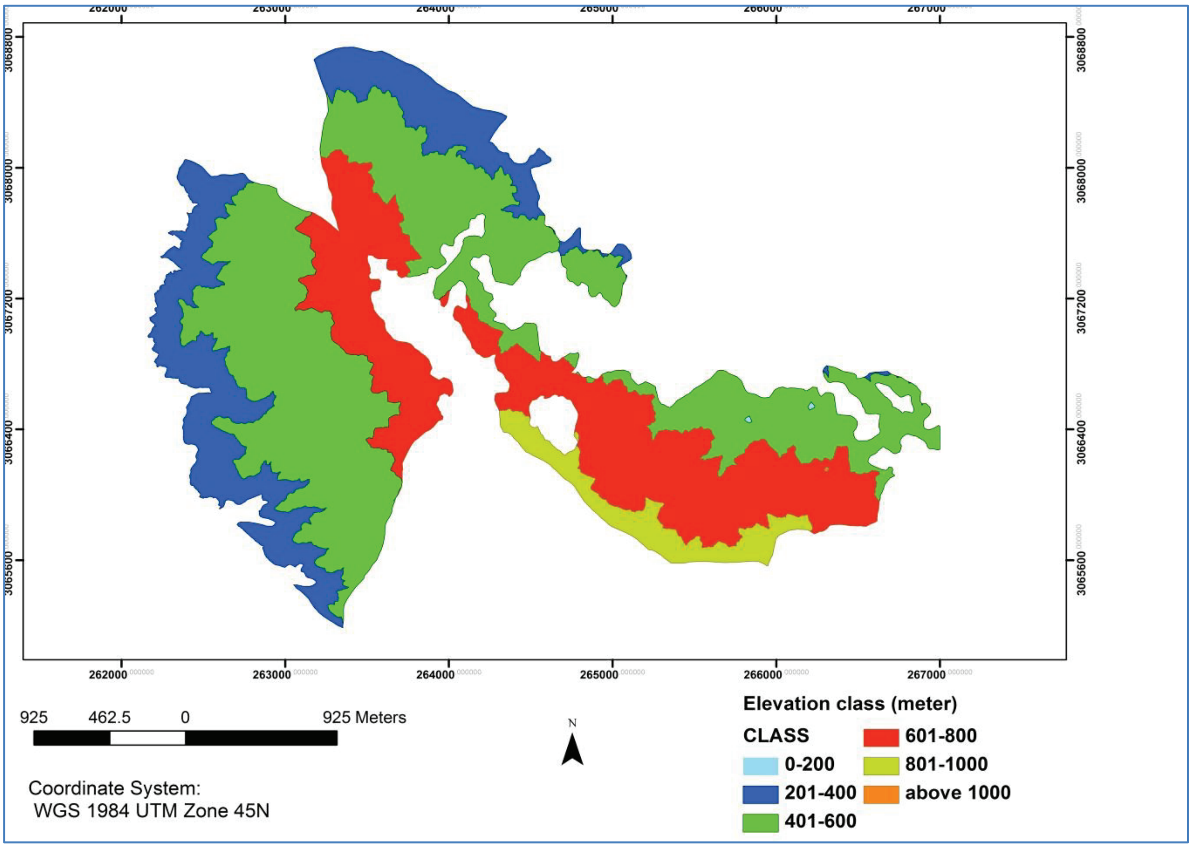


Figure21. Different elevation range within the study area.

Conversion of biomass polygon map to raster map

Per tree biomass polygon map (figure 19) was further converted to raster (figure 21) to extract the information of AGB per pixel level. It was performed to make consistency with other variables. For example litter quality data was extracted from species classification map. This classification map provides information per pixel. According to the figure 20, the lowest biomass content is 5.11 kg/m^2 . The maximum value for pixel is 171.50 kg/m^2.

3.5. Multicollinearity Issue: Relationship between above ground biomass and elevation

Before starting stepwise regression analysis, VIF value for all variables was checked one by one to check collinearity. Correlation matrix was also prepared for all variables. From the correlation matrix (Table 15), it was found that above ground biomass and elevation are highly correlated. Correlation coefficient value (r=0.84), indicates there is a positive correlation. It means that higher elevation has higher biomass. If VIF exceeds 10, it is a matter of concern to think about the variable selection. Here VIF for respective variables (table 15) were always lower than 10. So it may be summarized that although AGB and elevation are highly correlated, they will not reduce the reliability of predicted model.

When biomass raster map (figure 20) was checked with elevation range map (figure 21) it also showed that in most of the places higher elevation range showing higher biomass value compare to lower elevation range. Because in higher elevation (more than 600 m) it was found that trees are taller, with big canopy and wide DBH. But density is lower compare to low elevation. At low elevation range, tree density is higher but trees are smaller with a small DBH. But this is not true for the whole study area. According to the ground sample, this strong relationship is valid but in field there are other places where high elevation with low biomass might be possible. This would be a study of further investigation.

Table15. Summary of correlation matrix showing the correlation value and VIF among variables.

	SOC	Elevation	AGB	Variance Influence factor (VIF)		
SOC	1.00	0.74	0.79	SOC	Elevation	AGB
Elevation	0.74	1.00	0.84	3.19	4.11	5.39
AGB	0.79	0.84	1.00			

3.6. Stepwise regression to select the best model

Forward and backward stepwise regression procedure were performed to select the best predictor and found that soil organic carbon can be predicted with four variables AGB, elevation, species diversity and bulk density (Table 16) in backward regression. But bulk density is not a remotely sense able variable also. Results of forward regression selected two variables AGB and litter quality for SOC prediction. Finally based on Akaikes Information Criterion (AIC) and R^2 value, best fit model was selected (Table 16).

It is clear from the model summary table 16 that model-2 (forward regression) has the lowest AIC (-136.05) value and 66% of soil organic carbon (kg/m^2) can be explained from the above ground biomass (kg/m^2) and litter quality data by using this model. RMSE was calculated from the 61 plot data. If RMSE is considered, for model-2 error will be 0.29 kg/m². Instead of forward, if backward model is used to predict soil organic carbon (model -1, Table 16) only 68% soil organic carbon can be predicted by this model where error will be 0.27 kg/m².

In both cases, the second variables are marginally significant. For example in case of model-1, elevation and species diversity are marginally significant and bulk density cannot be measured through remote sensing data. Due to four explanatory variables, model-1 is showing the higher coefficient of determination (R^2) and lowest RMSE (Table 17). In both cases, the AIC value is quite similar. So model-2 refer to the parsimony principle, to explain as much as possible with as few explanatory variables as possible. Here, both variables can be measured through remote sensing data, which is the main stream of this research. So model-2 is selected here as the best fit model to explain SOC.

Table16. Represents the summary of forward and backward regression model with AIC and p value.

Model type	Predicted model	AIC value	R2	P value
Model-1: Forward regression	SOC ~ AGB(p<0.001) + elevation(p=0.07)+ Species diversity(p=0.11) + bulk density(p=0.15)	AIC=-136.15	0.68	P<0.001
Model-2: Backward regression	SOC ~ AGB (p<0.001) – Litter quality(p=0.07)	AIC=-136.05	0.66	P<0.001

Model -1: Summary of backward stepwise linear regression with all variables

SOC ~ AGB + Elevation + Species. Diversity + Bulk. Density

Now the model can be reformed with the value of intercept and coefficient those came from the summary of regression and it will be

SOC = 0.9242 + 0.0186*AGB + 0.000928*elevation + 0.1312 * Species diversity – 0.4043 * Bulk density

Model 2: Summary of forward stepwise linear regression with all variables

SOC ~ AGB + litter quality

Now the model can be reformed with the value of intercept and coefficient and it will be

SOC = 0.8417 + 0.0262*AGB – 0.4043 * Litter quality ………………Equation 12.

Table17. Represents the RMSE for SOC predicted model.

Model Name	RMSE	RMSE %
Model-1	0.27	17.00
Model-2	0.29	18.14

3.7. Model Validation through bootstrapping

Bootstrapping replicates the original dataset with a replication factor 1000 and provides the mean for each variable and intercept for selected model. From the bootstrapped replication result, it was found that the mean value of AGB is far from value 0 and is normally distributed. From the stepwise regression result it was also found that AGB is significant ($p<0.001$). As the coefficient of above ground biomass (Figure 23 a) is normally distributed and comes from 1000 replication dataset with the original dataset, this coefficient value is much more robust. So the coefficient value of 61 original observation data set can be replaced by the coefficient comes after bootstrapping. It will increase the strength of the model.

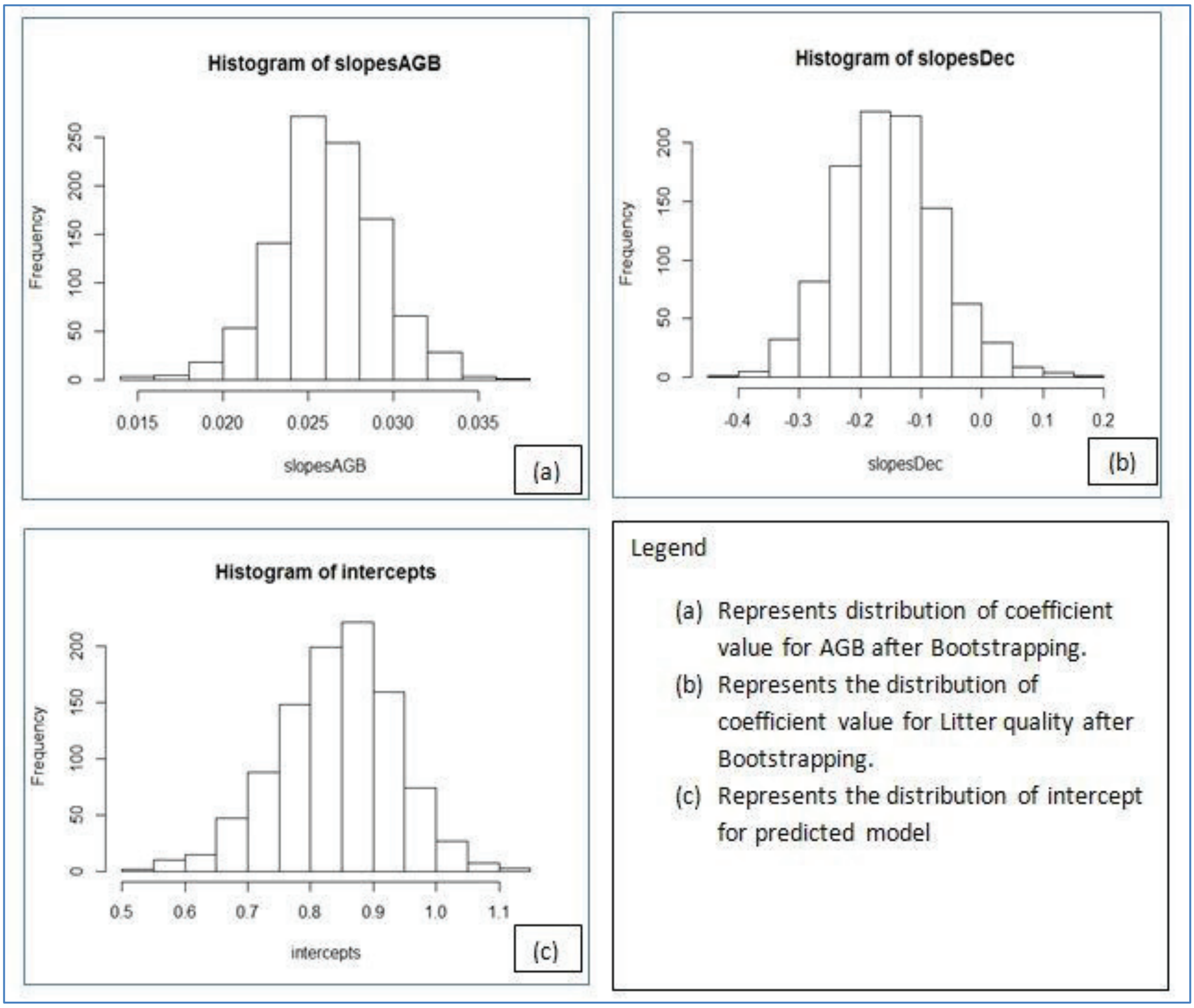


Figure22. Distribution of intercept and coefficient of the selected model after bootstrapping.

But in case of litter quality, the mean value is very close to 0 (Figure 22.b). From the result of stepwise regression, it was also found that litter quality variable was marginally significant ($p = 0.07$). Here it also makes sense to change the stepwise regression coefficient of litter quality with the bootstrapped coefficient of litter quality as it stronger due to replication issue.

In same way the intercept distribution (figure 22.c) of selected model comes from bootstrapping also more reliable. In table 18 the distribution of two variables mean and intercept are presented to support the argument of mean and intercept distribution pattern (figure 22).

The mean, intercept and confidence interval (95%) calculated from bootstrapping are presented below

Table18. Represents the bootstrapped statistics for selected model.

Bootstrapping	Variable name	Estimated mean	Confidence interval (95%)
	Intercept	0.8411201	0.02027681-0.03224018
Variable 1	AGB	0.02621259	-0.3108681 -0.04289908
Variable 2	Litter quality	-0.1529002	0.6561247-1.007624

So the selected model can be rebuilt with the bootstrapped value and it will be

SOC = 0.8411 + 0.0262* AGB – 0.1529*Litter quality ……………………….Selected model.

To validate the selected model , the soil organic carbon data collected from the field was contrasted with soil organic carbon data from the predicted model 2 and calculated a linear regression. The regression line show coefficient of determination of 0.68 and correlation values 0.82.

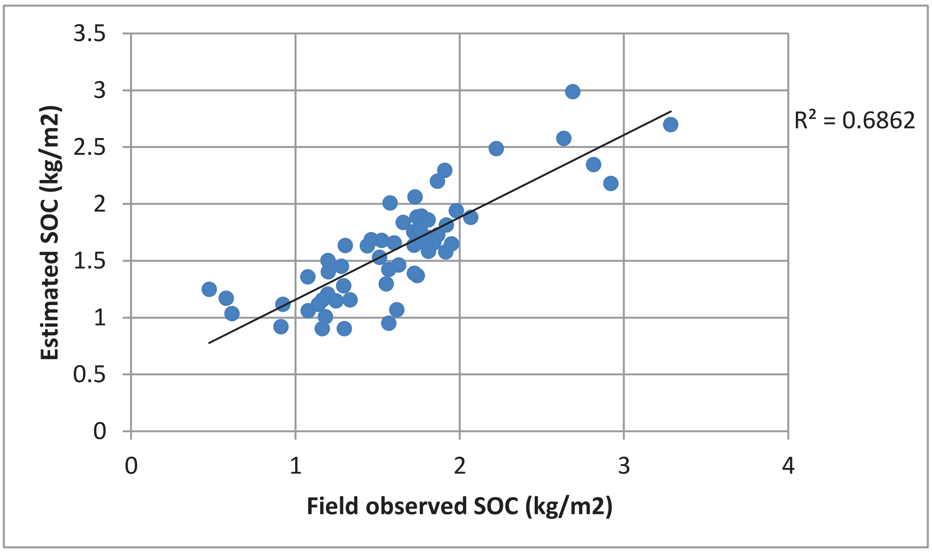


Figure23. Comparison between SOC measured in the field and SOC predicted by the model.

3.8. Soil Organic Carbon estimation and making of spatial SOC distribution map.

Above ground biomass data (kg/m²) was extracted from the AGB raster map (Figure.21). The litter quality value (1, 0) was extracted from the classified image map. When AGB polygon map converted into raster the cell size put similar with classified map image. So that each pixel of AGB raster map match with the pixels of the litter quality raster map. Here each pixel of AGB map provides the value of AGB (kg/m²). In similar way each pixel of litter quality map also provides the information of litter quality showing value either 1 or 0. In both map has some non-forest area which is showing no data when they combined. It means this area has no index (1 or 0) for litter quality and no data for above ground biomass to predict soil organic carbon. So this is a limitation of this model. Another important limitation is related to litter quality index rule that was prepared based on plot level ground data. But species classification result was showing the value per pixel. In this study plot wise pixel number according to species class was not counted to make litter quality index.

In figure 24, steps of soil organic map preparation are illustrated with related concept behind each step. How above ground raster map and litter quality map are used with their value to prepare a soil organic carbon spatial distribution map.

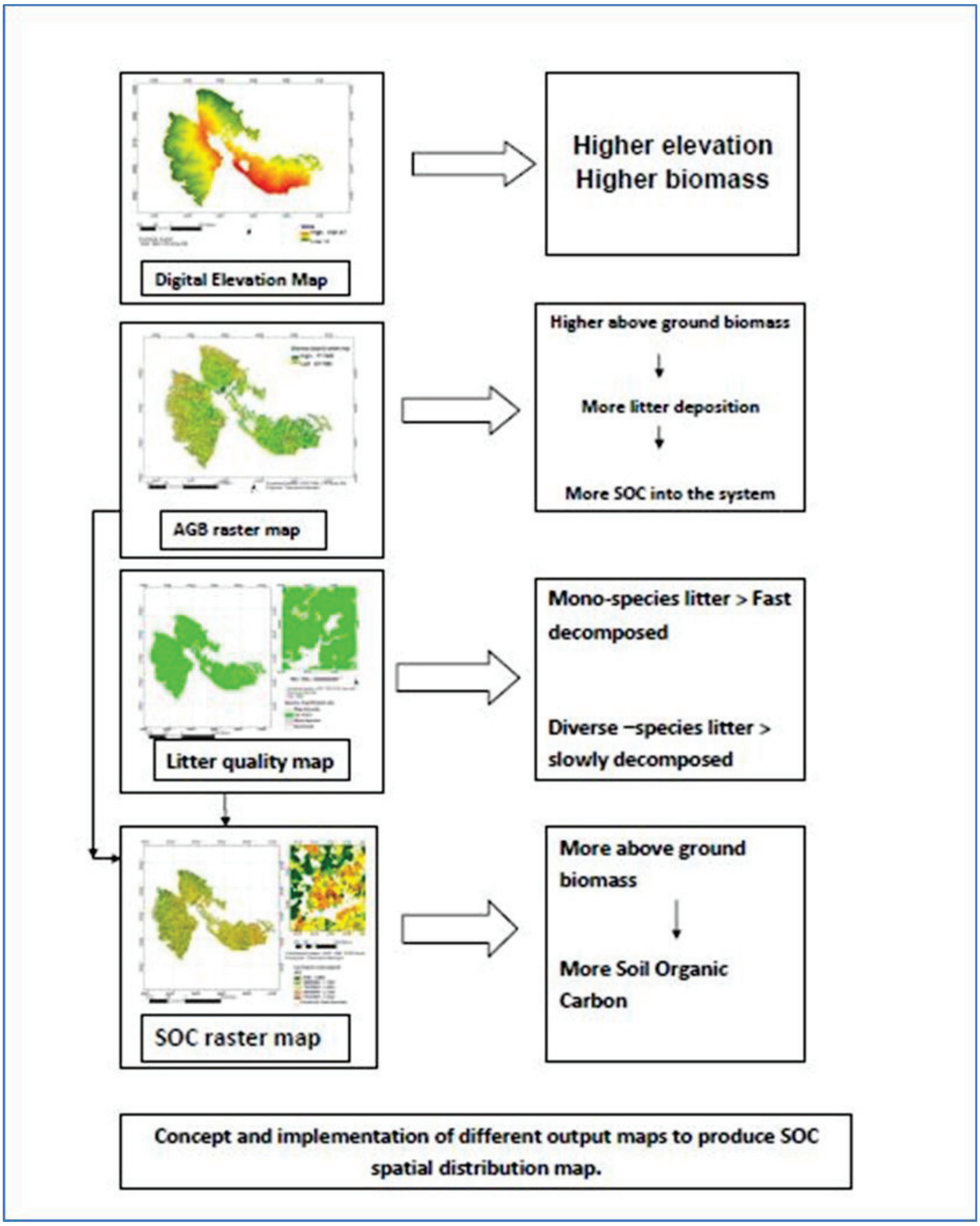


Figure24. Concepts and implementation of different output maps to produce SOC map

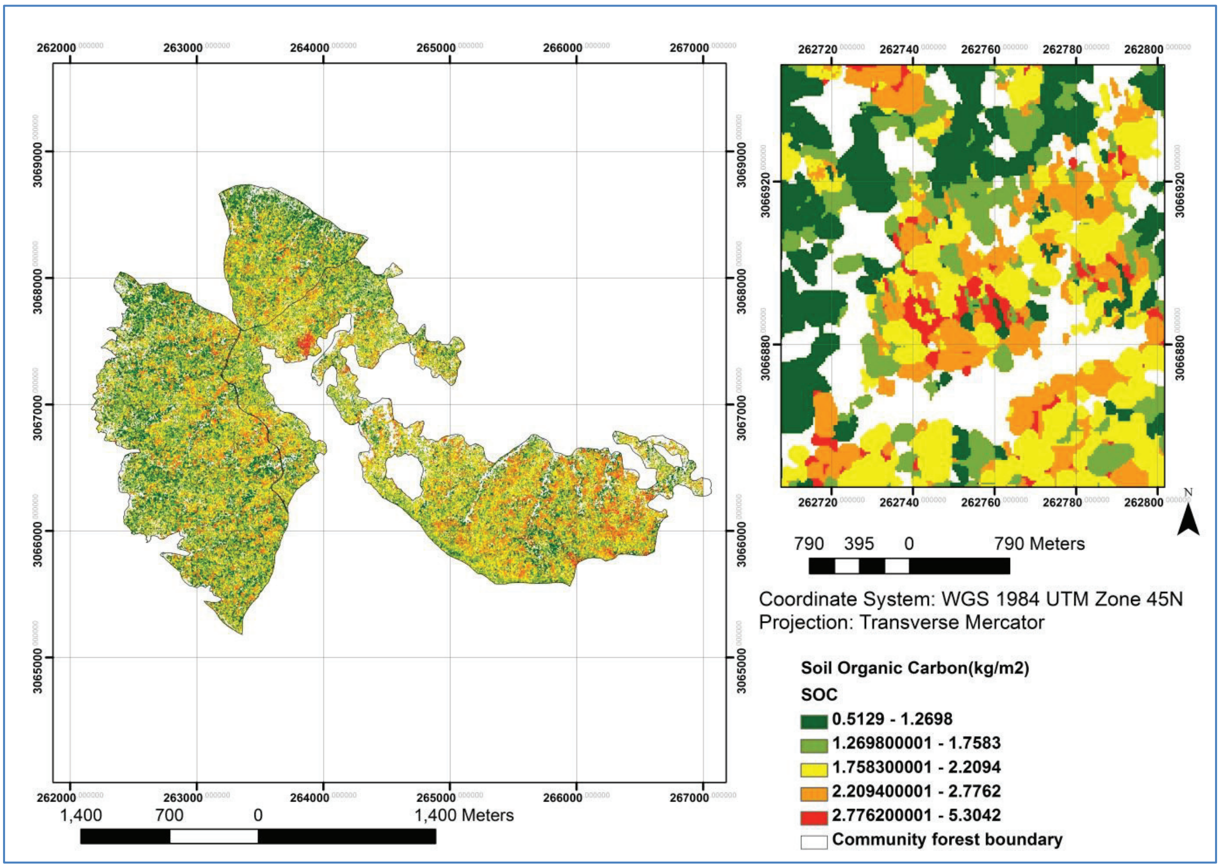


Figure25. Map of predicted soil organic carbon by using selected model.

The selected model was used in field calculator option in ArcGIS to calculate the soil organic carbon (kg/m^2). A result was showing that the minimum value of SOC is 0.51 kg/m^2 and the maximum value is 5.30 kg/m^2. The average SOC is 1.77 kg/m^2 within 0-10 cm layer.

4. DISCUSSION

4.1. Relationship between SOC analysed by LOI and WB methods

Generally Walkley-Black method is widely used and considered as more accurate compared to Loss on Ignition (LOI) method. In LOI method due to high temperature some inorganic carbon is also burnt and increases the value. So to test the accuracy of both methods, they both were contrasted in linear regression line. In this study the R^2 value was 0.71 ($p<0.001$). It means there is a strong correlation ($r = 0.84$) between this two methods.

Schulte (1995)reviewed the comparison between loss on ignition and Walkley black method and highlighted the correlation between the results of these methods. In a study, (Ball, 1964.)compared the weight loss of some organic soils of North Wales at 850 and 375°C with organic matter determined by a modification of the Walkley and Black (1934) procedure. At both temperatures the value of LOI was highly correlated with organic matter by the Walkley and Black procedure, but the lower temperature was deemed preferable. So the issue here is that the determination of temperature in LOI method affects the accuracy of SOC measurement. In another study Goldin (1987) compared the loss of weight on ignition of 60 non-calcareous soils of north western Washington and British Columbia with organic carbon determined with a Leco carbon analyser and found a strong correlation ($R^2=0.98$).

In this research, the overestimation derived from LOI method is a systematic error. The error value between LOI and WB (27.05%) is originated from different sources: (1) Determination of temperature for ashing (2) Placement and management of desiccator before weighing (3) Interfering inorganic constituents of the sample (4) Reagent composition and sample to sample variation in wet oxidation process.

The most important but unidentified factor is variability of organic matter composition itself for each and every sample.

4.2. Image classification and litter quality data

4.2.1. Image classification and accuracy assessment

The overall classification accuracy 72.85% can be discussed with the descriptive statistic of section 3.1. The achieved accuracy can be analysed in two ways: producer's accuracy and users' accuracy. In descriptive statistics, it was showed that about 68% of trees are *shorea robusta* (sal tree) in the respective area. Because almost everywhere is *Shorea robusta*, so when classification suggests 'mixed species', there is still a chance it is actually *Shorea robusta*. This is the reason for high producer accuracy in case of sal tree classification. 'Mixed species' class means a cluster of all minor species. This is very difficult to differentiate them through the spectral characteristic. Although they counted as a single class their spectral signature still different and produces error in classification process.

This result is comparable with Baral Jamarkattel (2011) who obtained 66.7% accuracy by classifying worldview-2 images into two classes. In same study area Mbaabu (2012) got the same accuracy level for community managed forest when she classified the accuracy of managed forest with government forest. She achieved 70% accuracy for community forest classification and 82% accuracy for government forest classification.

4.2.2. Transformation of species data into litter decomposability and litter quality

From the classified image it is clear that the whole study area is dominated by mono species *shorea robusta*. Litter quality index 1 is defined for *shorea robusta* species. 'Mixed species' classification class occupies quality index value 0. When litter is composed of mono species leaf (index 1), it decomposed faster compare to diversified litter (index 0). The reason why diversified litter is expected to take time to decompose is related to different chemical composition of leaf and foliage, diversified leaf shape and structure. In a study Joffre and gren (2001) showed that litter quality and quantity are determined by some above ground factors i.e. above ground plant composition, canopy structure, photosynthesis and respiration rate.

As the litter quality is marginally significant (p=0.07) it indicates that species diversity has a very marginal effect on litter quality. It makes sense, because 32 % area of interest is represented by 'mixed species' class. They are distributed within the whole study area and may be played a very minor role in decomposability. Litter decomposition is mainly dominated by sal species. This decomposition may be differed from species to species due to the composition of leaf and foliage. Several different studies, it has been revealed that litter decomposition rates and nutrient release in mixed-species forests increased, decreased, or even not differs, when compared to monocultures. Sometimes it happens due to site specificity and species selection (Mund & Schulze, 2005).

For example Kasel et al. (2011) observed the effects of site and tree species on SOC (0–10 cm depth) by using soils occurring under four species (*Acacia implexa, Acacia mearnsii, Allocasuarina verticillata* and *Eucalyptus melliodora*). They revealed the positive effect of tree growth on SOC. The relationship between SOC and species diversity was examined by Saha et al. (2009) for home gardens (HG), in Thrissur district, Kerala, India. They also measured tree density and plant-stand characteristics like species diversity (Shannon Index) of the HG. According to their result, the soil C stock was directly related to plant diversity of HG. Plant stand characteristics and species types are related with litter quality. Due to this, litter quality value was assessed from the species class to know the effect of litter quality on SOC. It was found a marginal effect of litter quality on SOC.

4.3. CHM segmentation and accuracy assessment to extract AGB data.

4.3.1. **Model validation: Relation between LiDAR height and observed tree height.**

Model validation is not the aim of this study but a complementary part. As biomass estimation depends on the sole input of trees height, it is essential to test accuracy of estimated height. The accuracy result and RMSE value indicate how much confident the data are for further analysis and discussion. The RMSE of 3.84 and $R^2 = 0.71$ shows that there is a strong positive correlation. The result of this study is comparable with the result of Kwak et al. (2007). He found 0.77, 0.88 and 0.70 coefficient of determination for two coniferous and one deciduous species respectively by using 1.8 m point density. Kraus et al. (2004) claimed the point density for DTM accuracy which influences the over and underestimation of height. In same study area, Karna (2012) dealt with same LiDAR data (0.8 m point density) and got 76% accuracy for segmentation. Takahashi. et al. (2005b) in their study found the overestimation of LiDAR tree height with an error of 0.90 meter in mountainous area of Sugi plantation in Japan when he dealt with high density LiDAR data. In another study, Kumar (2012) achieved 84% accuracy in tree canopy detection by using 164 points / m^2 density LiDAR data. So the density of LiDAR point cloud is a factor for segmentation accuracy.

In this study out of 185 trees 63% (117 trees) were overestimated and rest 37% are under estimated with an error 18.74%. The error comes from the predicted height is correlated with different sources like (i) Identifying the exact tree into the image and correlate it with the field observation (ii) Haga altimeter is not too much accurate for height measurement (iii) I-paq also produces some error to identify the exact tree (iv) Another important error is related to image noise. This noise comes from interpolation technique of LasTools software. LasTools only uses the TIN interpolation method and makes a prediction based on three close neighbour points on triangle and fit them into a model. So larger distance between the points makes larger error. So why the maximum height of observed tree (32 m) is far from the maximum height of LiDAR height (40 meter). In same study area, Karna (2012) dealt with same LiDAR data (0.8 m point density) and concluded that 0.8 m point density is good for plot level biomass estimation but not for accurate estimation of individual trees. This over and underestimation of tree height ultimately affects the estimation of biomass accurately. Finally this error also influences the spatial SOC distribution map.

4.3.2. Delineation of tree crowns from LiDAR CHM:

Accuracy test of segmentation was done by the procedure described by Zhan. et al. (2005), which applies to objects that are matching an object in the reference by at least 50%. Accuracy result (section 3.4.3) shows 70% canopy was properly segmented. At first this accuracy need to be compared with the result of Lopez Bautista (2012), as his rule set was applied in segmentation process. He got 76% accuracy in LiDAR CHM segmentation. This result can be compared with the result of Wang et al.,(2004) who used a high spatial resolution imagery to segment individual trees in a forest composed of *Picea glauca*, *Pseudotsuga menziesii* and *Abies lasiocarpa* and achieved 75.6% accuracy.

In this study accuracy reduced due to some causes. For example, (i) the study area is a natural broadleaf forest and the shape and ages of trees vary from one place to another. So why individual tree canopy delineation is a difficult task. (ii) Sometimes due to complex geometric shape of trees, species variability and height variation, surrounding trees are also delineated as a single canopy. (iii) As the study area is 38.42% matched with the working area of Lopez (2012), his rule set was applied to delineate the CPA. In some places, this rule set didn,t differentiate the intermingled tree and produces a larger CPA to create an error in biomass estimation. (iv) Another cause of moderate accuracy is that terrain shape and slope of forest. Some parts are very much steep (>75%). Due to steep slope canopy differentiation from CHM was difficult and reduces the accuracy as well.

4.4. Modeling DBH for AGB estimation

4.4.1. Model development and validation: Relation between CPA and DBH

In section 3.4.4, four different types of model have shown to predict the DBH from CPA. Based on lowest RMSE power model was selected for DBH prediction. In same study area, Lopez Bautista (2012) tested power and exponential model to calculate the AGB separately. He concluded that exponential model was unable to predict the DBH for those trees with a CPA more than 250 m^2. During field work, it was not found any tree with a CPA more than 250 m^2. Current sampling plot were completely different from (Lopez Bautista, 2012)'s plot. But it can't be denied that trees with a large CPA were there.
Results may be compared with the finding of Shimano (2000). He found that for broadleaved forest power model fits better for DBH measurement compares to exponential model. He classified six broadleaf species. For all species power model give better correlation coefficient ($r > 0.86$ for all species) compare to exponential model. The result of power model relationship between CPA and DBH ($r=0.82$) can be compared with the result of Shah (2011) who found a good correlation using linear model. He made a linear correlation between CPA and DBH for three species named *Shorea robusta, Schina walichii* and *Termanalia alata* and found a strong r value i.e. 0.83, 0.80 and 0.86 respectively which is similar with the current study's r value (0.81).
In this study, the coefficient of determination (R^2) for power model is 0.66 and $r = 0.81$; it means 66 % of DBH can be predicted from the segmented CPA. When a correlation was developed from training data, it was showing R^2 value 0.68. Power model was demonstrating a large residual error which comes from the different sources.

4.4.2. Biomass or AGB stock estimation

From the result of biomass map it was found that the mean value of biomass was 592.33 kg/tree which one is closer to mean biomass of observed value in the field. From field observed data, biomass of major species was estimated and it varied from 374.98 kg/tree (*Lagerstromia parviflora*) to 1263 kg/tree (*Semicarpus* *anacardium*). The segmented image was not classified into species level. During the implementation of allometric equation, wood specific gravity was counted as common for all species (0.88 gm/cm^3). This is one of the limitations of this estimation. The other uncertainty occurred in this map due to the segmentation accuracy.

As the rule set used to make this map was secondary data source (Lopez Bautista, 2012), the present result was compared with his findings. Lopez Bautista (2012) got 181.34 MgC /ha but according to this finding the value is 163 Mg C/ha (carbon conversion factor 0.47) which is lower than his findings. This deviation may be due to the selection plot. Because the ground sample plots were completely different from that one of Lopez Bautista (2012). As the ground plots are different, it also yields different regression model which effects on the calculation of AGB. Another important cause related to segmentation accuracy. Lopez Bautista (2012) got 6% more accurate result compare to this study.

The maximum range of biomass observed in field was 11,198 kg/tree (based on height and DBH). But in case of segmented biomass map, the highest range was 30237 kg/ tree (based on LiDAR height and predicted DBH). So this uncertainty comes from three sources one from the error of predicted DBH, another one from the LiDAR height accuracy and last one big crown or cluster of tree crowns which appeared as a single tree crown in segmentation.

Therefore this result was validated with the field observed biomass. The correlation coefficient value ($r=0.84$) indicates that there is a strong positive correlation between field observed biomass and LiDAR CHM segmented biomass. So the biomass extracted from LiDAR data is reliable to predict soil organic carbon.

4.5. Stepwise regression and selection the best fit model

To select the best model stepwise forward and backward regression automatic procedure was followed. Finally forward regression was selected based on AIC and p value of each variable. Here the variables are three types some of them are related with tree parameters, some of them are related with soil and topography. All parameters are not possible to measure through remote sensing. The main stream of this study was to predict soil organic carbon based on remote sensing variables.

A problem with stepwise regression is that the use of algorithm (forward selection, backward elimination or stepwise), the order of variable entry (or deletion), and the number of parameters, can all affect the selected model (Derksen & Keselman, 1992). This problem is more severe where the variables or predictors are correlated (Grafen. & Hails., 2002). Based on the problem of correlation between two predictors, model selection based on Akaike's Information Criterion (AIC) has increased substantially over recent years (Johnson. & Omland., 2004). In this study, above ground biomass and elevation are correlated. During the model selection process, AIC value was quite nearer from model -1 (AIC=-136.15) to model-2(AIC=136.15). But model -1 includes one marginally significant (elevation) and two non-significant variables (species diversity and bulk density). Bulk density is not possible to measure through remote sensing data. So why model -2 is preferred although it shows low R^2(0.66) compare to model-1(R2=0.68). But the variables of model -2 can be easily measured by remote sensing data.

From the model it is observed that AGB is positively correlated. It can be discussed in this way that higher above ground biomass means larger canopy and high shoot/root ratio. It causes more leaf and foliage deposition at the soil surface. Based on decomposition rate, litter turns into soil organic carbon. So AGB and SOC are positively correlated. On the other hand, litter quality is showing negative correlation with soil organic carbon. Based on litter quality index, it can be illustrated that monoculture species provide more homogenous litter which is easily decomposable. Fast and easy decomposable litter produce less SOC. In similar way, mixed species litter produce heterogeneous litter and takes time to decompose.

Ultimately more SOC deposited into the soil. According to the rule of litter quality index, it makes sense that litter quality and SOC are negatively correlated.

4.5.1. Bootstrapping for model validation

Bootstrap is a computer intensive method used frequently in applied statistics. This model was first described by Bradley Efron (1979). Now a lot of papers has been published based on bootstrapping and model slection. It normally used for robust estimation of sample standard error and confidence interval. It is believed that at least 1000 bootstrap samples are needed for replication but sometimes 10000 samples replicates also used (Kenneth & David, 2002).

In this study 1000 samples replication number was used to calculate a robust mean of all variables and confidence interval. From the result of bootstrapped validation it was found that mean of AGB and intercept of the model were normally distributed. This is the main strength of bootstrap method. Instead of making assumption about the distribution of values, bootstrap goes back to the original data and makes a replacement with original one. Bootstrap shows the distribution of mean, intercept and confidence interval for all variables. So the recalculated coefficient from additional samples is much more robust compare to original dataset mean. So the coefficient of both variables (AGB and litter quality) determined by bootstrapped are more robust and give more confidence about the strength of the model.

4.6. Soil Organic Carbon (SOC) estimation

According to the current study, the average value of soil organic carbon was 1.77 kg/m^2. Some uncertainty and errors are involved with this result. In result section 3.8, it was discussed that the developed model can predict only 66 % of SOC. This large amount of residual started from field measurement and continued up to model development. At section 4.8 all error propagation will be listed.

These results are comparable to some published research reports from Nepal. For example, B.M Shrestha. et al. (2004a) reported that the SOC pools at 0–20 cm soil depth was 2.6 kg C m^2 in forest in a mid-hill watershed in Nepal. Their study area was Pokhare khola watershed in Nepal. The elevation range of Pokhare khola watershed varies from 400 m to 1100 m, similar to the current study. Similarly Awasthi. et al. (2005.) reported SOC pool in 0–15 cm soil layer of forest area was 4.0 kg C /m^2 which is higher than this results. The difference might be due to location of the study areas, model development and accuracy.

This results can be compared with B. M. Shrestha. et al. (2009) who developed a CENTURY model to predict SOC in different land uses in Nepal watershed. According to their model the value of SOC for 2004 was 3.8 kg/m^2 for a managed forest soil. This value was for 2004 but the current study was showing a moderate value compare to that. The strength of model is not only an issue for this moderate SOC value. SOC also lose due to different causes. For a subtropical deciduous forest, SOC loss and depletion is related with two important issue, such as i) soil erosion ii) human disturbance. Due to mountainous and hilly slope (some place more than 80%), Chitwan watersheds are facing loss of soil C from the system due to erosion. Gerrard (2002), found moderate erosion of 5.6 Mg/ha/year through their model. Another SOC loss comes from human disturbance which one is related with i) Firewood and biomass removal ii) Removal of leaf litter and iii) Selective logging. Both may be happened at a same time, it may be reduced, washed out and dissolved due to different causes and at a same time SOC can be regain due to managed forest activities. Current study shows a range of SOC up to 5 kg/m^2. In different location, it gives different value. But the average value is 1.77 kg/m^2.

4.7. Error propagation and uncertainty

Sources of errors those propagated in different stages are shown in a flowchart (Figure 27). Some important issues are discussed bellow

4.7.1. **Field data collection: location of trees**

In any field survey, GPS and I-PAQ calibration and find out the exact location is a tough work. Especially in case of natural broadleaf forest where canopy are intermingled with each other and the shape and size of trees are heterogeneous. Due to this problem exact location of sample plot could be a little bit deviated. Even finding the location of reference tree is also deviated. It also effects on segmentation accuracy and model development. So why compare to coniferous forest, broadleaf forest classification show less accurate result. The impacts of field data collection errors are showed in figure 26.

4.7.2. **Laboratory method**

Laboratory work was done with full precautions. In both methods, quality control samples were used to check the analysis accuracy. 'Laboratory Duplicate sample' was used to count the methodological error. In same way the 'Field duplicate sample' were analysed separately. Field duplicate means repeated sample from same plot. Laboratory duplicate samples showed minor error. But some field duplicate sample showed different results. This error originated from sample collection, drying, processing and continued up to analysis. It also indicates that samples are not homogenous everywhere and their composition is different.

4.7.3. **Image segmentation: Noise of LiDAR data**

Noise rectification of LiDAR data is an important task to develop a CHM model. Some negative values and extra higher values were eliminated to make it noise free. Still some pixels with noise value remain in CHM image and causes error in segmentation process.

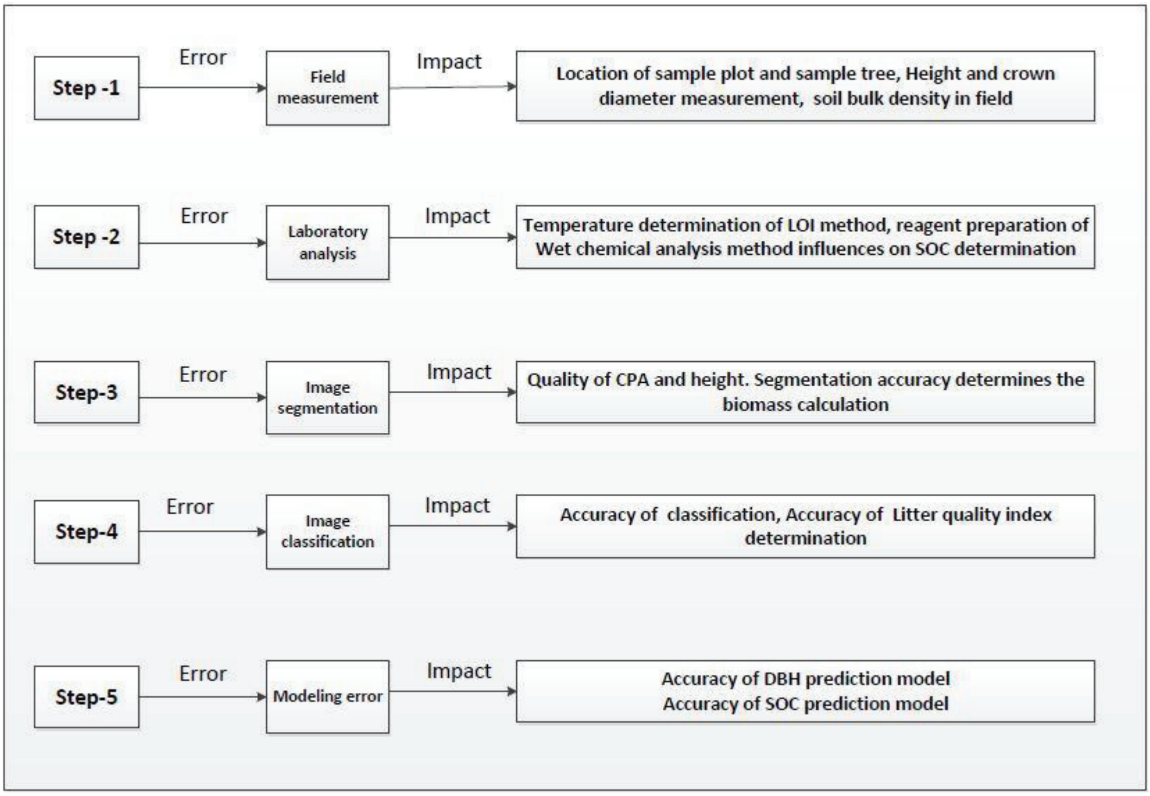


Figure26. Flow diagram showing the steps of error propagation

4.7.4. **Model development**

Two different types of models were used in this study. One is regression model to predict DBH based on CPA. Another one is stepwise regression model to predict SOC based on different variables. The results of those variables come from either laboratory data or image processing works. So they are not completely error free.

4.8. Novelty of this research work

Direct laboratory analysis of soil organic carbon is accurate but needs soil sample as well. Instead of collecting soil sample and laboratory analysis, a prediction may be done based on other variables. Direct remote sensing quantification of soil organic carbon still needs laboratory facility and not suitable for small SOC variability sample of same location due to large error of the model (Bartholomeus et al., 2008). So this method is not suitable for soil analysis of same location with small variability in SOC composition. In a recent study, Nocita et al. (2012) concluded that SOC quantification from reflectance suffered from uncontrolled soil surface, vegetation cover and moisture content. So they are also trying to find out the alternate solution. A very few works have been done in this field and few works have been published. But research interest demands more concern in this field concerning the global warming and mitigation issue. This research revealed this opportunity. Here based on remotely sensed variables a statistical model is developed to estimate soil organic carbon.

4.9. Limitations of this study

As it was a new work related to quantitative estimation of soil organic carbon through remote sensing data, some limitations need to discuss in a proper way for further improvement. One of them is related to the raster value of AGB and litter quality map. Both of the maps originated from different sources and non- forest area was not matched completely on each other. It was not possible to omit the cloud cover in Worldview image classification. So in output result, cloud cover was showing as a shadow area. It means this area has no index (1 or 0) for litter quality to predict soil organic carbon. But in CHM, cloud cover area was showing AGB raster value. In that case only AGB value was taken to prepare the SOC map. Here litter quality index was ignored. Multiresolution segmentation of worldview image integrated with LiDAR CHM can solve this limitation. In case of integrated segmentation, litter quality index classification will match completely with CHM data.

Another limitation related to the common non forest area. In both maps, common non forest area was showing no data. It means this area has no data for above ground biomass and litter quality to predict soil organic carbon. But in reality everywhere there should be some SOC either more or less compare to forest cover area. This is a limitation of this predicted model.

Other limitation need to be acknowledging that litter quality index rule was prepared based on plot level ground data. But species classification result was showing the value per pixel. Pixel number per plot according to species class was not counted to make an accurate litter quality index. For further study total number of pixels within the plot and their classification is necessary to make a more accurate index for litter quality.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusion

The main aims of this research were to assess the effect of elevation, aboveground biomass and tree species diversity on Soil Organic Carbon (SOC) and to develop a model to estimate SOC stock using airborne LiDAR and high resolution Worldview image -2 measured variables. In his regards, conclusions are based on the research questions as follows:

Is there any strong relationship between soil organic carbon and elevation?
Based on the correlation matrix and stepwise regression, it was found that elevation and SOC are both positively correlated. It was expected that this strong correlation ($r = 0.74$) was the reflection of the fact of that there was a strong correlation between above ground biomass and elevation. All variables are interconnected within the system and it is difficult to measure the individual influence on each other. So why in a backward regression model, elevation individually didn't predict soil organic carbon. When elevation is related to species diversity, AGB and soil bulk density, it can predict soil organic carbon. But in this model species diversity and bulk density are non-significant (p value is 0.11 and 0.15 respectively).

Is there any positive correlation between soil organic carbon and above ground biomass?
From this study, it was proved that there is a strong positive correlation between soil organic carbon (SOC) and above ground biomass (at 95% confidence interval, p value < 0.001). Based on a backward stepwise regression model, above ground biomass can predict SOC when it is correlated with elevation, species diversity and bulk density. In similar way based on forward stepwise model, AGB can predict SOC when AGB is correlated with litter quality ($p=0.07$).

Is there any positive correlation between soil organic carbon and species diversity?
Based on the result of the stepwise regression, it was found that litter quality is very marginally correlated ($p = 0.07$) with soil organic carbon and can predict SOC in relation with AGB. In this study, litter quality is the representative index of species types. From this, it can be concluded that there is a marginal correlation between SOC and species types. However, there was a poor correlation between species diversity and soil organic carbon. When species diversity predict soil organic carbon in relation with AGB, bulk density and elevation, the significance level of species diversity is very low ($p= 0.11$).

Which regression model best explain the relationship between SOC and all other remotely sensed variables measured by LiDAR and Worldview image?

From the findings of the study it can be summarized that the following model is the best fit model based on AIC and p value of stepwise regression procedure.

$SOC = 0.8411+0.0261*AGB – 0.1529*litter\ quality$
So soil organic carbon can be measured by using two remotely sensed variables, above ground biomass and litter quality. The coefficient of determination (R^2) indicates that 66% soil organic carbon can be measured by using this model. This model predicted the average value of 1.77 SOC (kg/m^2) within 0 to 10 cm layer in Chitwan District, Nepal.

5.2. Recommendation

Estimation of soil organic carbon (SOC) based on remote sensing variables is a new and emerging field of work. As the sufficient amount of remote sensing data was available for this study area, the work was conducted based on the available remote sensing data. But litter quality is not a direct ground sampling representative data. Plant litter sample was collected during field work to make a relationship with the SOC and litter quality. Only litter index was prepared based on species class to make dummy variables for stepwise regression. So for further improvement of SOC estimation through RS data the following works may be recommended

i) At first a relation between species types and plant litter of respective place needs to be analysed with remote sensing data and plant litter laboratory analysed (chemical composition of litter) data. A plant litter sample is need in the same location of soil sample to determine C: N ratio. Afterwards it may be analysed if there is any relation between litter quality and soil organic carbon.

ii) Sampling was designed based on elevation strata. It is recommended to collect more samples from a higher elevation. More samples from a higher elevation range will help to make a reliable decision about the elevation variable as a predictor for soil organic carbon.

iii) Soil samples were collected from the randomly selected forest area to make a correlation with above ground biomass. Sampling from non-forest area was ignored. As the biomass is a RS predictor for SOC, non-forest area showing zero (null) carbon value for soil. This is an error. So non – forest area soil sampling is recommended to develop a new algorithm for SOC estimation of non-forest area.

iv) Species diversity should be a criterion in sampling design to judge the correlation between soil organic carbon and species diversity. This current study ignored this criterion. For further investigation and study area selection, species composition and diversity should be analysed before selecting the sampling design.

v) The ground samples represent a linear positive relationship between biomass and elevation. Further sampling is recommended at a location with low biomass and a higher elevation or vice versa.

vi) Soil Organic Carbon estimation was the main stream of this work. Remote sensing related work like segmentation and image classification were a complementary part. This meant that improvements to the segmentation and image classification accuracy were ignored due to time limitations. For further work, this should be considered. Because accurate biomass estimation depends on accurate segmentation and accurate image classification and it needs time and proper ground sampling..

LIST OF REFERENCES

Annonymous. (2011). *The core of GIScience a process based approach*. University of twente, ENSCHEDE, The Netherlands: Faculty of geo information science and earth observation.

ANSAB., ICIMOD., NORAD., and FECOFUN. (2010). Guidelines for measuring carbon stocks in community-managed forests, Nepal, Asia.

ARBONAUT. (2010). Monitoring changein carbon stock and achieving REDD+ targets. . *Arbolidar,8*.

Arvidsson, J. (1998). Infuence of soil texture and organic matter content on bulk density, air content, compression index and crop yield in feld and laboratory compression experiments. Uppsala, Sweden

Received 26 September 1997; accepted 22 June 1998: Department of Soil Sciences, Swedish University of Agricultural Sciences, P.O. Box 7014, S-750 07.

Awasthi., K. D., Singh., B. R., and Sitaula., B. K. (2005.). Profile carbon and nutrient levels and management effect on soil quality indicators in the Mardi watershed of Nepal. *Acta Agric. Scand., B 55,* , 192–204.

Ball, D. F. (1964.). Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils. *J. Soil Sci., 15*(84-92.).

Baral Jamarkattel, S. (2011). *Mapping carbon stock using high resolution satellite images in sub - tropical forest of Nepal.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2011/msc/nrm/baral.pdf

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, A., Hoogmoed, W. B., and Spaargaren, O. S. P. (2008). Spectral reflectance based indices for soil organic carbon quantification. *Geoderma, 145*(1–2), 28-36. doi: http://dx.doi.org/10.1016/j.geoderma.2008.01.010

Belzar, H., Rowland, C. S., and Saich, P. (2007). Forest canopy height and carbon estimation at Monks Wood National Nature Resrve,UK, Using Dual Wavelength SAR inferometry. *Remote sensing of environment, 108(3)*, 224-239.

Ben-Dor, E., Inbar, Y., &, and Chen, Y. (1997). The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500) during a controlled decomposition process. *Remote Sensing of Environment 61*(1-15).

Beyer, L., Kahle, P., Kretschmer, H., &, and Wu, Q. (2001). Soil organic matter composition of man-impacted urban sites in North Germany. . *Journal of Plant Nutrition and Soil Science 164*, 359–364.

Brian. A. S. (2002). Methods for the determination of total organic carbon (TOC) in soils and sediments. *United States Environmental Protection Agency Environmental Sciences Division National Exposure Research Laboratory. P.O. Box 93478 Las Vegas, NV 89193-3478.*, 5.

Brown, S. (2002). Measuring Carbon in Forest: current status and future chalenges *Environmental pollution, 116(3)*, 363 - 372.

CBD. (2009). Connecting Biodiversity and Climate Change Mitigation and Adaptation: Report of the second Ad Hoc Technical Expert Group on Biodiversity and Climate change. *Technical Series No41.*, 126.

Charlotte. H, M., William., D., &, and Perreault., J. (1991). Collinearity, Power and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research., XXVIII*, 268-280.

Chave, J., Andalo, C., Brown, S., Cairns, M., Chambers, J., 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

Clinton, N., Holt, A., Scarborough, J., and Yan, L., et al (2010). Accuracy assessment measures for object based image segmentation goodness. *Photogrammetric Engineering and Remote sensing, 76(3)*(289-299).

Definiens. (2011). e Cognition Developer 8.7 Reference Book, Munchen Germany.

Deqiang, Z., Dafeng., H., Yiqi., L., * and , and Guoyi., Z. (2008). Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. *Journal of plant ecology, 1*(2), 85-93.

Derksen, S., and Keselman. (1992). Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. . *British Journal of Mathematical and Statistical Psychology. , 45*, 265–282. .

DigialGlobe. (2010). White Paper " The benefits of the spectral band of Worldview -2. ". Retrieved Retreived 28-10-2012 from http://www.digitalglobe.com/products#analysis&8-band-multispectral-imagery http://www.digitalglobe.com/products#analysis&8-band-multispectral-imagery
Dixon, R. K., Brown, S., Houghton, R. A., Soloman, A. M., Trexler, M. C., and Wisniewski, J. (1994). Carbon pools and fluxes of global forest ecosystems. Science 263., 185-190.
Drouin, A., Saint-Laurent, D., Lavoie, L., and Ouellet, C. (2011). High-Precision Elevation Model to Evaluate the Spatial Distribution of Soil Organic Carbon in Active Floodplains. *Wetlands, 31*(6), 1151-1164. doi: 10.1007/s13157-011-0226-z
Dubayah., R. O., &, and et. al. (2000). Land surface characterization using LiDAR remote sensing In: Hill MJ, Aspinal RJ (Eds). Spatial Information for landuse management. (pp. 25-38). Australia: Gordon and Breach Science Publisher.
Esteban., G. J., and Robert., B. J. (2000). The vertical distribution of soil organic carbon and it's relation to climate and vegetation. *Ecological Applications., 10*(2), 423-436.
Gautam., B. R., and Kandel. P.N. (2010). Working paper on LiDAR mapping in Nepal. Retrieved from
Gerrard, J., Gardner, R.,. (2002). Relationships between landsliding and land use in the Likhu Khola Drainage Basin, Middle Hills, Nepal. . *Mt. Res. Dev. , 22*(48–55.).
Goldin, A. (1987). Reassessing the use of loss-on-ignition for estimating organic matter content in non-calcareous soils. *Commun. Soil Sci. Plant Anal. , 18*, 1111-1116.
Gonzalez, P., Asner, G. P., Battles, J. J., Lefsky, M. A., Waring, K. M., and 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
Grafen., A., and Hails., R. (2002). Modern Statistics for the Life Sciences. *Oxford University Press, Oxford.* .
Grizelle., G. T., R S. (2001). Soil fauna and plant litter decomposition in tropical and subalpine forest. *Ecology by the ecological society of America, 82(4)*, 955-964.
Guggenberger., G., Zech, W., Haumaier, L. a., and Christensen., B. T. (1995). Land-use e ects on the composition of organic matter in particle-size separates of soils. II. CPMAS and solution 13C NMR analysis. *European Journal of Soil Science., 46,* , 147±158.
Gulledge, J., and Schimel, J. P. (2000). Controls on soil carbon dioxide and methane fluxes in a variety of taiga for stands in interior Alaska. *Ecosystems 3*, 269-282.
Hardon, H. J. (1936). Factoren, die het organische stofen het stikstof-gehalte van tropische gronden beheerschen. *Landbouw XI(12)*(517±540.).
Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R. (2000). Controls over carbon storage and turnover in high latitude soil. *Global Change Biol.Suppl. , 6(8)*, 196-210.
Hunt., C. A. G. (2009). Carbon sinks and climate change: forest in the fight against global warming. . Retrieved from http://books.google.nl/books?hl=en&lr=&id=FBV004gcEaMC&oi=fnd&pg=PR1&dq=Carbon+sinks+and+climate+change:+forest+in+the+fight+against+global+warming.+&ots=oA0vuPVo0L&sig=6EAcjOEthYHz7KSNwJAI29rOyeY
Husch, B., Beers. T.W, and Kershaw, J. A. (2003). Forest mensuration .Wiley.
IPCC. (2007). Summary for Policy Makers.In : Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel On Climate Change. Cambridge University Press, Canbridge, United Kingdom and Newyork, NY, USA.
IPCC. (2001). Intergovernmental Panel on Climate Change, 2001. *Climate Change : The Scientific Basis, Cambridge Univ. Press, Cambridge , UK.*
Joffre, R., and gren, G. I. (2001). 5 - From Plant to Soil: Litter Production and Decomposition *Terrestrial Global Productivity* (pp. 83-99). San Diego: Academic Press.
Johnson., J. B., and Omland., K. S. (2004). Model selection in ecology and evolution. . *Trends in Ecology and Evolution. , 19*, 101–108.
Karna, Y. K. (2012). *Mapping above ground carbon using worldview satellite image and LIDAR data in relationship with tree diversity of forests.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2012/msc/nrm/karna.pdf

Kasel, S., Singh, S., Sanders, G. J., and Bennett, L. T. (2011). Species-specific effects of native trees on soil organic carbon in biodiverse plantings across north-central Victoria, Australia. *Geoderma, 161*(1–2), 95-106. doi: 10.1016/j.geoderma.2010.12.014

Katherine, L. B., Mary, E. M., and John, D. A. (1996). Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectances: a comparison of statistical methods. *Canadian Journal of Forest Research 26(4)* 590-600. doi: 10.1139/x26-068

Kenneth, P. B., &, and David, R. A. (2002). *Model Selection and Multimodel Inference A Practical Information-Theoretic Approach.*

Kraus, K., Wagner, W., Ullrich, A., &, and Melzer, T. (2004). From single pulse to full waveform Airborne Laserscanner: Potential and practical challenges. *from Institute of photogrammetry and Remote sensing, Viena University of technology, Austria.*

Kumar, V. (2012). *Forest inventory parameters and carbon mapping from airborne LIDAR.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2012/msc/nrm/vinodkumar.pdf

Kwak, D. A., Lee, W. K., Lee, J. H., and Biging, G., et al. (2007). Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research, 12(6)*(2839-2849).

Lal, R. (2004). Soil carbon sequestration to mitigate climate change. *Geoderma, 123*(1–2), 1-22. doi: 10.1016/j.geoderma.2004.01.032

Lal, R. (2005). Forest soils and carbon sequestration. *Forest Ecology and Management, 220*(1–3), 242-258. doi: 10.1016/j.foreco.2005.08.015

Lal., R. (2001). *Assessment methods for soil carbon.* Boca Raton etc.: Lewis.

Lopez Bautista, A. A. (2012). *Biomass carbon estimation and mapping in the subtropical forest of Chitwan, Nepal : a comparison between VHR geo-eye satellite images and airborne LIDAR data.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2012/msc/nrm/lopezbautista.pdf

Lu, D. (2006). The potential and challenge of Remote sensing based biomass estimation [Review]. *International Journal of Remote Sensing, 27(7)*, 1297-1328.

Malhi, Y. (2010). The carbon balance of tropical forest regions, 1990–2005. [Review from 1990-2005]. *Current Opinion in Environmental Sustainability, 2*(4), 237-244. doi: 10.1016/j.cosust.2010.08.002

Martin., M. E., & , and John., D. A. (1997). HIGH SPECTRAL RESOLUTION REMOTE SENSING OF FOREST CANOPY LIGNIN, NITROGEN, AND ECOSYSTEM PROCESSES. . *Ecological Applications 7*, 431–443. . doi: http://dx.doi.org/10.1890/1051-0761(1997)007[0431:HSRRSO]2.0.CO;2

Mbaabu, P. R. (2012). *AGB carbon mapping using airborne LIDAR data and geo-eye satellite images in tropical forests of Chitwan, Nepal : a comparison of community and government managed forests.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2012/msc/nrm/mbaabu.pdf

Michael., R. C., &, and Robert., A. L. (2011). An Introduction to Bootstrap Methods with Applications to R. Retrieved from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470467045.html

Moller, M., Lymberner, L., &, and Volk., M. (2007). The comparison index: A tool for assessing the accuracyof image segmentation. *International journal of Applied Earth Observation and Geo information, 9*(3), 311-321.

Motavalli., P. P., Palm, C. A., Parton, C. A., Elliott, E., and Frey, S. D. (1995). Soil pH and organic C dynamics in tropical forest soils: evidence from laboratory and simulation studies. *Soil Biology & Biochemistry, 27*, 1589±1599.

Mund, M., and Schulze, E. D. (2005). Silviculture and Its Interactions with Biodiversity and the Carbon Balance of Forest Soils. In M. Scherer-Lorenzen, C. Körner & E.-D. Schulze (Eds.), *Forest Diversity and Function* (Vol. 176, pp. 185-208): Springer Berlin Heidelberg.

Naesset., E. (2009). *Some challenges in forest monitoring, proposing a " research platform" for development, training and validation of spaceborne technologies in Brazil.* . Paper presented at the A presentation Brazilian Norwegian on forest and monitoring., Ilhabela, SP, Brazil. 24 slide. retrieved from http://xa.yimg.com/kq/groups/6172281/911844978/name/Working+Paper+on+LiDAR+Mapping+in+Nepal.pdf.

Nelson., D. W., &, and L.E., S. (1996). Total carbon, organic carbon, and organic matter. In: Methods of Soil Analysis. *A.L. Page et al., Ed. Agronomy. 9:961-1010. Am. Soc. of Agron., Inc. Madison, WI., Part 2, 2nd ed., .*
Nguyen, T. T., Janik, L.J and Raupach, M. (1991). Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Aust. J. Soil Res, 29*(49-67).
Nocita, M., Stevens, A., Noon, C., and van Wesemael, B. (2012). Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma*(0). doi: http://dx.doi.org/10.1016/j.geoderma.2012.07.020
P Lerouge, M. C.-M., C Rayon, A.C Fischette-Lainé, V Gomord, L Faye. (1998). N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol Biol, 38*, pp. 31–48.
Patenaude., G., Milne, R., and & Dawson, T. P. (2005). Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto protocol. *Environmental Science and Policy, 8*(2), 161-178.
Popescu., S. C., Wynne, R.H., & Nelson, R.F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote sensing, 29*(5), 564-577.
Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G. (1982). Soil carbon pool ad world life zones. *Nature 298*, 156-159.
Prentice., I. C. (2001). The carbon cycle and the atmospheric carbon dioxide. Climate Change 2001. *The Scientific Basis. Intergovernmental Panel on Climate Change. Cambridge Univ. Press. UK.*, pp. 183– 237.
Prichard, S. J., Peterson, D.L, and Hammer, R., D. (2000). Carbon distribution in subalpine forests and meadows of the of the Olympic Mountain, Washington. *Soil Sc. Soc. Am. J. , 64*, 1834-1845.
Rumpel, C., Janik, L. J., Skjemstad, J. O., and Kögel-Knabner, I. (2001). Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. *Organic Geochemistry, 32*(6), 831-839. doi: http://dx.doi.org/10.1016/S0146-6380(01)00029-8
Rumpel, C., Knicker, H., Kögel-Knabner, I., Skjemstad, J. O., and Hüttl, R. F. (1998). Types and chemical composition of organic matter in reforested lignite-rich mine soils. *Geoderma, 86*(1–2), 123-142. doi: http://dx.doi.org/10.1016/S0016-7061(98)00036-6
Saha, S., Nair, P., Nair, V., and Kumar, B. (2009). Soil carbon stock in relation to plant diversity of homegardens in Kerala, India. *Agroforestry Systems, 76*(1), 53-65. doi: 10.1007/s10457-009-9228-8
Salah, Y. M. S., and Scholes, M. C. (2011). Effect of temperature and litter quality on decomposition rate of Pinus patula needle litter. *Procedia Environmental Sciences, 6*(0), 180-193. doi: http://dx.doi.org/10.1016/j.proenv.2011.05.019
Schimel., D. S., House., J. I., Hibbard., K. A., and Bousquest., P., et al.,. (2001). Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature 414.* , 169–172.
Schulte, E. E. (1995). Recommended Soil Organic Matter Tests.
Shah, S. K. (2011). *Modelling the relationship between tree canopy projection area and above ground carbon stock using high resolution geoeye satellite images.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede. Retrieved from http://www.itc.nl/library/papers_2011/msc/nrm/shah.pdf
Sharma, E. R., and Pukkala, T. (1990). Volume Equation and biomass prediction of forest trees of Nepal.
Shimano, K. (2000). A power function for forest structure and regeneration pattern of pioneer and climax species in patch mosaic forests. *Plant Ecology, 146*(2), 205-218. doi: 10.1023/a:1009867302660
Shrestha., B. M., Sitaula.B.K., Singh., B. R., &, and Bajracharya., R. M. (2004a). Fluxes of CO2 and CH4 in soil profiles of a mountainous watershed of Nepal as influenced by land use, temperature, moisture and substrate addition. . *Nutr. Cycl. Agroecosys., 68*(155–164).
Shrestha., B. M., Williams., S., Easter., M., Paustian., K., and Singh., B. R. (2009). Modeling soil organic carbon stocks and changes in a Nepalese watershed. *Agriculture, Ecosystems & Environment, 132*(1–2), 91-97. doi: http://dx.doi.org/10.1016/j.agee.2009.03.003
Song, C., Dickinson,M.B.,Su,L.,Zhang, S.,et al (2010). Estimtion average tree crown size using spatial informationfrom Ikonos and Quickbird images: Across-sensor and across-site comparisons. . *Remote Sensing of Environment, 114(5)*, 1099-1107.
Spain, A. V. (1990). Infuence of environmental conditions and some soil chemical properties on the carbon and nitrogen contents of some tropical Australian rainforest soils. *Australian Journal of Soil Research 28*, 825 -839.

Takahashi., T., Yamamoto., K., Senda., Y., &, and Tsuzuku., M. (2005b). Estimating individual tree heights of sugi(Cryptomeria japonica D.Don) Plantation in mountaneous areas using small footprint airborne LiDAR. *Journal of forest research., 10(2).*

Vaze., J. a. T., J. (2001). High Resolution LiDAR DEM – How good is it? Retrieved from http://www.mssanz.org.au/MODSIM07/papers/12_s27/HighResolution_s27_Vaze_.pdf

Walkley, A., and Black, I. A. (1934). An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Sci, , 37 (1934),* , pp. 29–37.

Woolard., J. W., and Colby., J. D. (2002). Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. *Geomorphology, 48 (1–3)*(269–287.).

Zhan., Q. M., Molennar., M., Tempfi, K., and Shi, W. Z. (2005). Quality assessment of geospatial objects derived from remotely sensed data. *International journal of remote sensing, 26*(14), 2953-2974.

GLOSAARY

Litter: Plant litter is dead or live part of plants (leaves, twigs, stem and roots) and characterized as fresh, undecomposed, and easily recognizable (by species and type) plant debris.

SOM: Soil organic matter (SOM) is the organic matter component of soil and considered an important part of soil for its high contribution to soil productivity. Generally, SOM contains two main fractions: humic substances and labile soil organic matter. But intensively SOM can be classified into three groups

i) Labile SOM – This part is very quickly reactive and provides energy and nutrients for soil micro-organisms. It releases part of the nutrients for plant usage and its half-life is between days and few years. It produces short-term organic matter turnover during the year.

ii) Stable SOM – this is the less decomposable part. Cation-exchange capacity is the main and the most important function of this pool.. Its half-life is between years and decades.

iii) Inert SOM – the non-reactive part of organic matter and physico-chemically protected against decomposition . Its half-life is between decades and centuries.

SOC: Soil Organic Carbon (SOC) is the common name for carbon held within the soil. Soil carbon is the largest terrestrial pool of carbon (2,200 Pg). Soils contain carbon (C) in both organic and inorganic forms. In most soils (with the exception of calcareous soils) the majority of C is held as soil organic carbon (SOC).

Soil bulk density (Db) is a measure of the mass of soil per unit volume (solids + pore space). This is generally done on an oven-dry basis. Soil bulk density was measured in the field. After collecting the soil sample from filed, it was keep 3-4 days open to make it air dry. Regular weighing helped to assume the moisture content of the soil. When the weight of the sample became more or less fixed, it was counted as the air dried weight of the sample. Then the bulk density was determined by using the following equation

$$Db = (\text{mass of dry soil}) / (\text{volume of solids \& pore spaces})$$

Above round biomass (AGB): The term biomass is related with the dry weight of trees. It includes above ground biomass and below ground biomass as well. Above ground biomass includes all living materials on the surface such as stem, stump, branches, bark, seeds, and leaves of vegetation from both strata of trees and vegetation strata below the forest floor.

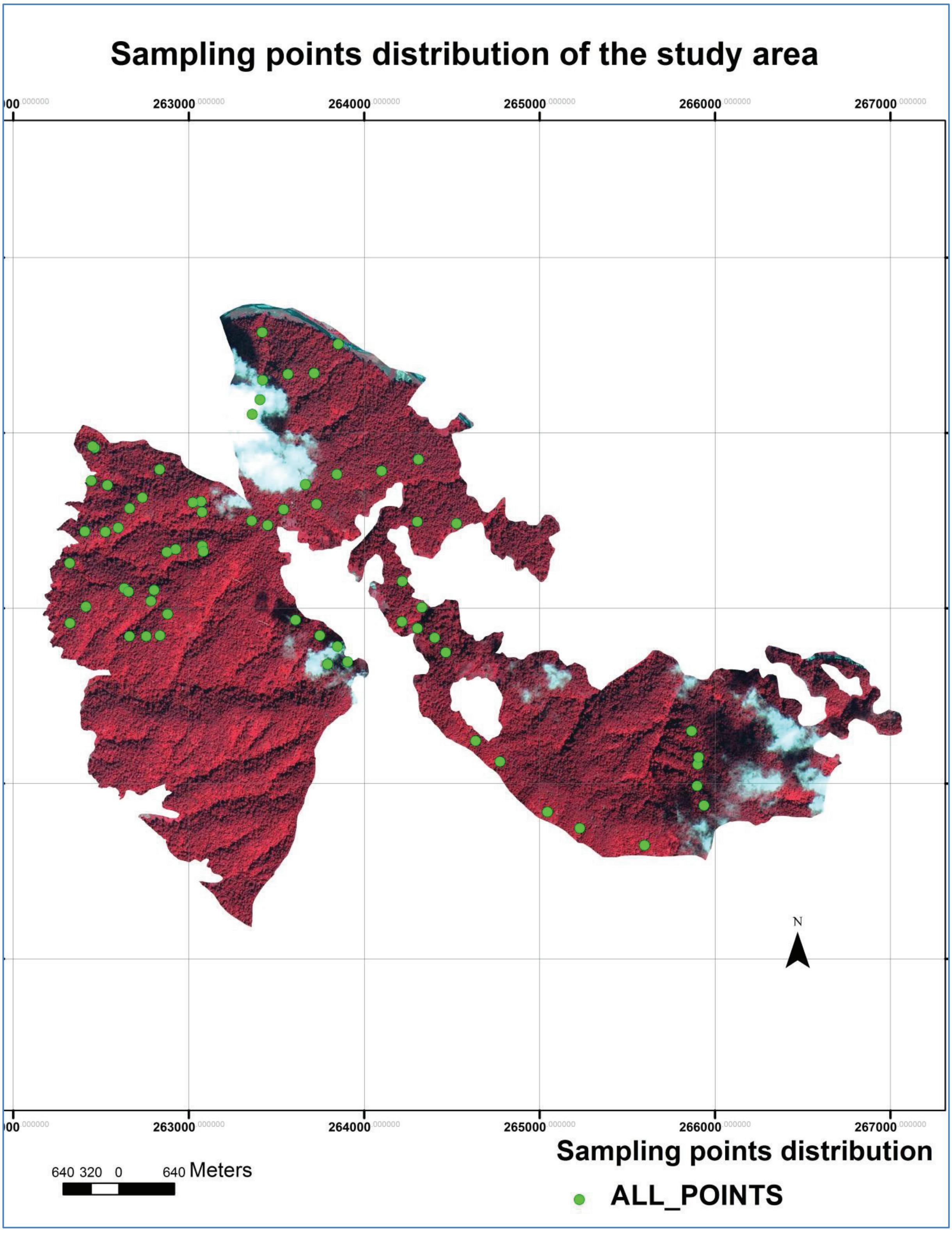
Crown Projection Area (CPA): The crown projection area of a tree is the area of vertical projection of the outermost perimeter of the crown on horizontal pane. Crown size, which is closely related to the photosynthetic capacity of tree, is an important parameter to characterize tree biomass.

Diameter at Breast Height (DBH): This is the diameter of a tree over bark measured perpendicular to the stem axis of at breast height.

Species Diversity:Species diversity is the number of different species that are represented in a collection of individuals or any population data set. Here the number of species means the number of equally-abundant species. It is needed to obtain the same mean proportional species abundance as that observed in the dataset of interest (where all species may not be equally abundant). Species diversity consists of two components, species richness and species evenness.

ANNEX

Annex 1. Distribution of sampling points within study area



Annex 2. Materials used for tree and soil parameter measurement

Worldview -2 Image and LiDAR data characteristics

Data	Spectral range of World view- 2 image	Wavelength
Worldview-2	Panchromatic	450-800 nm
	Coastal Blue	400-450 nm
	Blue	450-510 nm
	Green	510-580 nm
	Yellow	585-625 nm
	Red	630-690nm
	Red Edge	705-745 nm
	Near Infrared 1	770-895 nm
	Near Infrared 2	860-1040 nm
Data : LiDAR	Characteristics	
Date flown 20110316/20110328/20110401	Flying Speed	80 knots
	Sensor pulse rate	52.9khz
	Sensor Scan Speed	20.4 lines/second

Materials used for tree parameter measurement

List of materials used in the field work for tree parameter measurement

Name	Application
Printed Map	For plot and sampling point identification.
Measuring tape	For measuring the radius of the plot to establish a plot.
Diameter tape	For measuring the diameter of the tree at breast height.
Clinometer Haga	For measuring the height of the tree.
Clinometer Suunto	For measuring the slope of the plot.
Spherical Densitometer	For measuring the canopy density.
GPS Garmin	To identify the plot and location with accurate coordinate.
iPAQ	For navigation or positioning

Materials used for soil carbon measurement

List of items used for soil sample collection

List of items	Uses
Metal scale	for soil depth measuring
Soil sample core	to collect soil samples for bulk density
Soil sample hammer	for bearing down on the soil core while collecting sample
White cloth/ masking tape	for tightening the soil core so that no soils come out
Kuto (trowel)	for taking out soil core from the soil depth
Soil Auger	for collecting soil composite sample

Annex 3. Soil sample analysis result from Walkley-Black and Loss on Ignition

Plot	WB	LOI	plot	WB	LOI
DP_1	1.1369529	1.834037302	Nebu_11	2.0312805	2.899624754
DP_2	1.5401315	2.175139254	Nebu_12	2.3644042	2.558139535
DP_3	3.1518303	3.985846247	Nebu_13	2.6262442	3.486140024
DP_4	2.1268345	3.137067927	Nebu_14	3.9103372	4.169767442
DP_5	1.2898648	2.198108744	Nebu_15	2.6470182	3.515634116
DP_6	1.3709919	2.837209302	Nebu_16	3.1466244	3.794451661
DP_7	0.7343243	1.350052256	Nebu_17	2.585794	3.325581395
DP_8	2.3585785	3.204984031	Nebu_18	2.7104087	2.783631087
DP_9	2.2464813	3.628714017	Nebu_19	2.8060155	3.052721467
DP_10	1.4057023	2.877874641	Nebu_19-Ref	1.7474971	2.227537915
DP_11	2.3479181	3.02257706	Nebu_20	2.7710417	3.079351282
DP_12	3.0102097	4.756003597	Nebu_21	2.6727757	2.953588921
DP_13	2.4535582	3.855114432	Nebu_21_Ref	2.7538568	3.167519981
DP_14	2.2807719	3.608742452	Nebu_22	3.8591372	4.124306152
DP_15	1.5334663	2.581216558	Nebu_23	2.9981806	3.326931196
DP_16	3.2555919	4.136439948	Nebu_24	2.7102414	3.456976744
DP_17	1.557246	2.720930233	Nebu_25	3.5785395	3.808075956
DP_17_Ref	1.7256715	3.259664136	Nebu_26	2.636833	3.299060547
DP_18	2.2844943	3.356655964	Nebu_27	2.4337292	3.51366814
DP_19	2.3430413	3.627906977	Nebu_28	2.143158	2.584435654
DP_20	1.578492	2.191860465	Nebu_29	2.8372263	3.100102763
DP_21	2.0535087	3.365409771	Nebu_30	3.9813879	4.101364247
DP_22	1.6885097	2.598497542	Nebu_31	2.293094	2.533536407
DP_23	2.1064635	3.55162513	Nebu_32	2.8045165	3.515971888
DP_24	2.50536	3.112091892	Nebu_32_Ref	3.7695045	4.192257994
DP_25	3.1043213	4.092818253	Jan_1	2.257717	3.144538591
DP_26	0.6248755	1.242502136	Jan_2	1.8039279	2.393626587
DP_27	2.9479372	3.07118316	Jan_3	0.7722772	1.2
DP_28	1.9703966	2.076526901	Jan_4	1.647734	2.581354
Nebu_1	1.4735835	2.552325581	Jan_5	1.8465193	2.780239243
Nebu_2	1.7808996	2.375632521	Jan_6	2.707649	3.430535321
Nebu_3	1.7765298	2.468399816	Jan_7	1.9861512	2.694120647

Nebu_4	1.9678802	2.373627979	Nebu_9	2.1063824	2.328904163
Nebu_5	2.8549219	3.755813953	Nebu_9_Ref	0.8970444	1.552228628
Nebu_6	2.6373313	2.984802029	Nebu_10	1.7868533	2.154236835
Nebu_7	2.2753614	3.957400228			
Nebu_8	2.2443052	2.869209464			

Annex 4. Plot level database used for stepwise regression

CF	PLOT ID	X coordinate	Y coordinate	Elevation	AGB Kg/m2	SOC	Diversity	soil pH	Bulk density	Litter
Devidonga	Plot_no1	262408.5568	3067439.15	382.27	8.83	0.91	0.60	5.08	1.59	1
Devidonga	Plot_no2	262445.3984	3067727.91	347.12	8.11	1.16	0.86	4.7	1.41	0
Devidonga	Plot_no3	262663.1734	3067570.42	480.22	23.14	1.55	0.93	5.4	1.10	1
Devidonga	Plot_no4	262597.8357	3067459.88	431.88	22.56	1.29	0.84	5.3	1.22	0
Devidonga	Plot_no5	262736.922	3067632.28	476.09	16.33	0.92	0.54	5.02	1.43	1
Devidonga	Plot_no6	262524.884	3067436.31	407.28	14.22	1.08	0.76	4.72	1.57	1
Devidonga	Plot_no7	262536.5033	3067702.46	376.41	13.22	0.61	0.66	5.17	1.67	1
Devidonga	Plot_no8	262836.3457	3066845.64	519.59	20.15	1.74	1.45	4.83	1.48	0
Devidonga	Plot_no9	262662.5489	3066840.65	416.39	14.51	1.62	0.60	5	1.44	1
Devidonga	Plot_no10	263076.972	3067549.27	559.95	19.73	1.08	1.52	4.72	1.53	1
Devidonga	Plot_no11	263077.8721	3067355.89	566.91	10.01	1.57	0.20	5.06	1.34	0
Devidonga	Plot_no12	263083.712	3067325.01	594.76	39.68	2.07	1.10	5.5	1.37	0
Devidonga	Plot_no13	262926.8847	3067337.46	529.46	32.88	1.82	1.12	5.44	1.48	0
Devidonga	Plot_no14	262876.3414	3067321.75	516.95	31.3	1.78	1.16	5.32	1.56	0
Devidonga	Plot_no15	262323.7264	3066915.7	373.35	19.71	1.20	0.21	5.49	1.56	1
Devidonga	Plot_no16	262784.8837	3067041.95	514.85	29.52	1.63	0.50	5.56	1.00	1
Devidonga	Plot_no17	262414.2338	3067009.53	431.05	25.14	1.20	0.58	5.25	1.54	1
Devidonga	Plot_no18	262633.1235	3067114.18	497.8	22.13	1.57	1.49	4.96	1.37	0
Devidonga	Plot_no19	262759.07	3066841.13	478.29	20.93	1.72	1.32	5.12	1.47	0
Devidonga	Plot_no20	262880.0772	3066968.18	541.07	6.26	1.18	1.55	5.27	1.50	0
Devidonga	Plot_no21	262803.4671	3067103.58	489.92	26.23	1.51	1.17	4.86	1.47	0
Devidonga	Plot_no22	262464.5566	3067916.39	334.78	17.84	1.17	0.50	5.37	1.38	1
Devidonga	Plot_no26	262834.838	3067792.57	428.42	15.54	0.48	1.48	4.91	1.52	0
Devidonga	Plot_no27	263070.727	3067609.23	597.22	33.83	1.87	0.98	5.17	1.26	0
Devidonga	Plot_no28	263023.9645	3067604.84	576.22	30.12	1.44	1.59	5.21	1.46	0
Nebuwater	Plot_1	264527.2245	3067484.21	421.55	10.52	1.14	1.33	4.42	1.55	0
Nebuwater	Plot_2	265867.075	3066300.37	461.23	30.26	1.30	1.59	4.39	1.46	0
Nebuwater	Plot_3	265045.1369	3065838.82	540.51	31.95	1.52	1.19	4.64	1.72	1
Nebuwater	Plot_4	265230.6615	3065746.39	625.66	44.52	1.58	0.84	5.01	1.60	1
Nebuwater	Plot_5	264635.4538	3066244.67	600.79	33.95	1.92	0.30	4.45	1.35	1
Nebuwater	Plot_6	265598.7091	3065648.74	722.69	36.98	1.84	1.32	4.44	1.40	1
Nebuwater	Plot_7	264775.0235	3066125.87	708.79	38.06	1.46	0.65	5.72	1.32	0
Nebuwater	Plot_8	264309.7574	3067849.79	730.21	42.95	1.92	1.18	4.67	1.71	1
Nebuwater	Plot_9	264099.5451	3067784.05	752.96	39.79	1.74	0.00	4.82	1.63	0
Nebuwater	Plot_10	263844.8768	3067764.53	617.08	29.08	1.28	0.49	4.34	1.44	1
Nebuwater	Plot_11	263665.5795	3067706.69	752.96	34.81	1.72	0.24	4.77	1.69	0
Nebuwater	Plot_12	263729.1404	3067595.35	753.51	39.11	1.74	0.00	4.85	1.46	1
Nebuwater	Plot_13	263541.5036	3067564.48	773.86	36.61	1.95	0.00	4.94	1.50	1
Nebuwater	Plot_14	263450.6775	3067474.15	814.01	56.89	2.92	0.82	3.91	1.49	1
Nebuwater	Plot_15	263359.8973	3067499.97	861.03	66.23	2.64	1.96	5.34	1.34	0
Nebuwater	Plot_16	263848.1105	3066783.02	695.2	46.57	1.73	0.00	5.48	1.35	1
Nebuwater	Plot_17	263748.2385	3066845.18	756.36	34.16	1.81	1.56	4.66	1.40	0
Nebuwater	Plot_18	263906.583	3066695.15	798.88	41.94	1.98	1.37	4.42	1.45	1
Nebuwater	Plot_19	265902.1165	3066110.49	682.53	40.1	1.77	1.41	4.83	1.23	0
Nebuwater	Plot_20	265898.2376	3065987.94	485.55	30.31	1.72	0.74	5.54	1.25	0
Nebuwater	Plot_21	265937.5111	3065876.42	704.04	44.66	1.81	1.16	5.39	1.35	1
Nebuwater	Plot_22	265906.943	3066151.55	865.6	81.88	2.69	0.68	5.58	1.39	1
Nebuwater	Plot_23	264303.4265	3067494.15	718.76	61.36	1.91	0.54	5.17	1.28	1
Nebuwater	Plot_24	263609.5503	3066934.34	774.46	68.58	2.22	0.61	4.83	1.64	1

Annex 5. Stepwise regression table and AIC value used for model.

Model 1: Summary of backward stepwise linear regression with all variables

Last step of backward regression

```
Step:  AIC=-136.15
SOC ~ AGB + Elevation + Species. Diversity + Bulk. Density

                      Df Sum of Sq    RSS     AIC
<none>                             5.5565 -136.15
- d$Bulk.density       1   0.20894 5.7654 -135.90
- d$Species.Diversity  1   0.25642 5.8129 -135.40
- d$Elevation          1   0.32599 5.8825 -134.67
- d$AGB.               1   1.76217 7.3187 -121.35

Call:
lm(formula = SOC ~ d$AGB + Elevation + Species.Diversity + Bulk.density)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.924234   0.439156   2.105   0.0398 *
d$AGB                0.018699   0.004437   4.214 9.21e-05 ***
d$Elevation          0.000928   0.000512   1.813   0.0753 .
d$Species.Diversity  0.131221   0.081628   1.608   0.1136
d$Bulk.density      -0.404356   0.278651  -1.451   0.1523
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.315 on 56 degrees of freedom,Multiple R squared:
0.6813, Adjusted R-squared: 0.6585, F-statistic: 29.93 on 4 and 56 DF,  p-
value: 2.5e-13
```

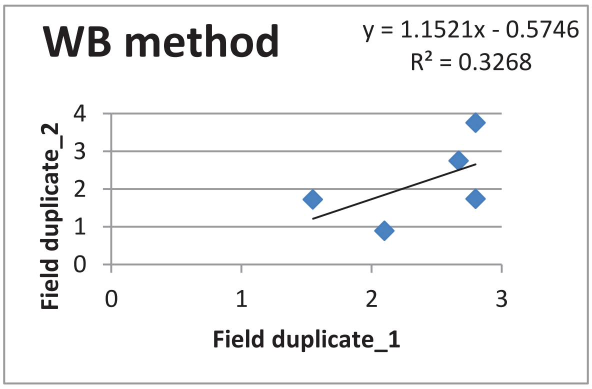
Model 2: Summary of forward stepwise linear regression with all variables

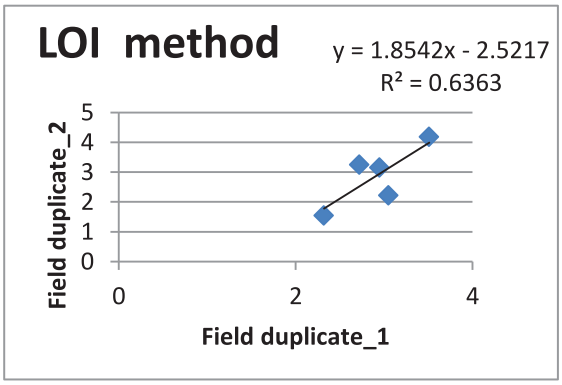
Last step of forward regression

```
Step:  AIC=-136.05
SOC ~ AGB + Litter quality

                      Df Sum of Sq    RSS     AIC
<none>                             5.9430 -136.05
+ d$Elevation          1  0.190225 5.7528 -136.03
+ d$Bulk.density       1  0.134829 5.8082 -135.45
+ d$Species.Diversity  1  0.066159 5.8769 -134.73
+ d$Species.Evenness   1  0.028851 5.9142 -134.34
+ d$soil.pH            1  0.003377 5.9396 -134.08


Call:
lm(formula = SOC ~ AGB + Decomposibility)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.841737   0.091529   9.196 6.30e-13 ***
d$AGB.             0.026233   0.002485  10.556 4.03e-15 ***
d$Litter quality  -0.156211   0.084990  -1.838   0.0712 .
---
```

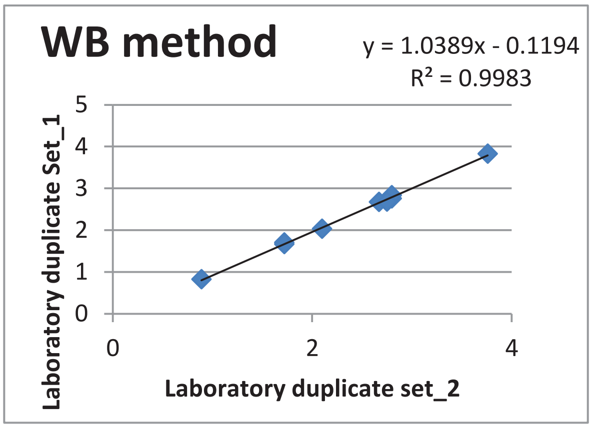
```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3201 on 58 degrees of freedom
Multiple R-squared: 0.6591,    Adjusted R-squared: 0.6474
F-statistic: 56.07 on 2 and 58 DF,  p-value: 2.79e-14
```

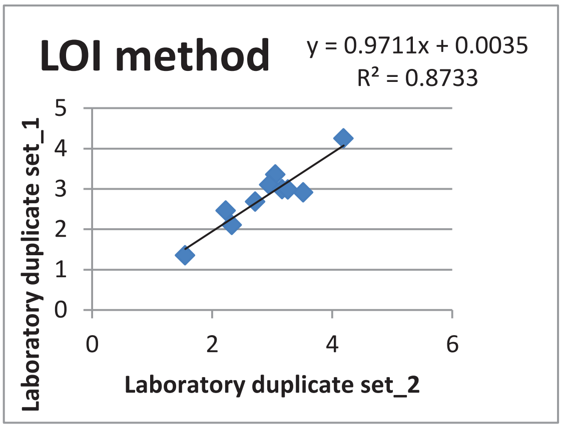
Annex 6. Regression test for field duplicate sample and laboratory duplicate sample





Above pictures represent the scenario of field duplicate samples analysed by WALKLEY BLACK and LOI method.





Above pictures represent the scenario of laboratory duplicate samples analysed by WALKLEY BLACK and LOI method.

Annex 7. Photographs from field and laboratory work