

COUPLING AIRBORNE LIDAR AND HIGH RESOLUTION OPTICAL SENSOR PARAMETERS FOR BIOMASS ESTIMATION USING MACHINE LEARNING

KASHI RAM YADAV

March, 2019

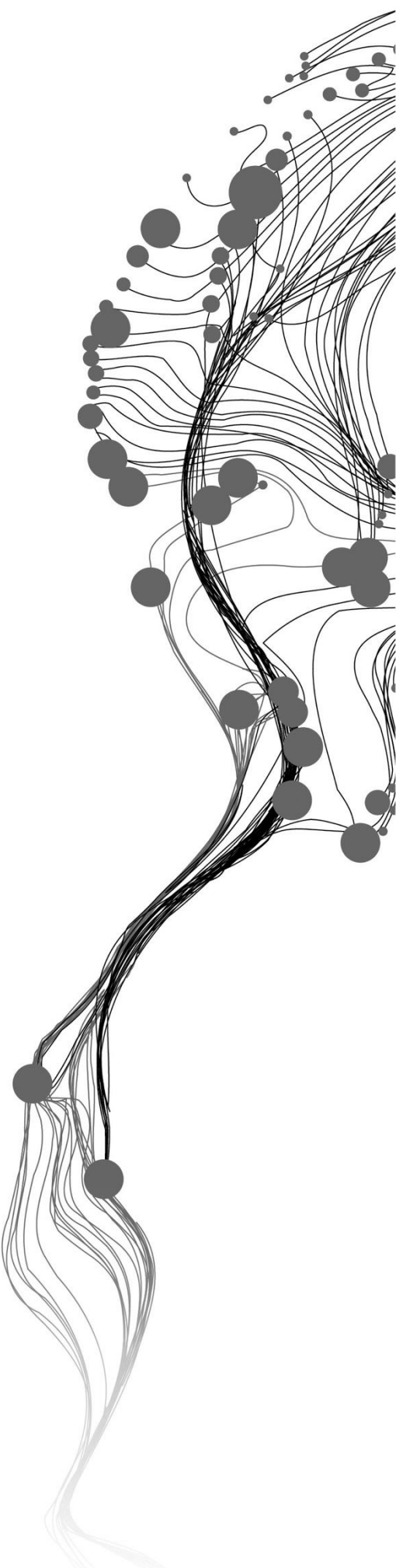
SUPERVISORS:

Dr. Subrata Nandy

Dr. Michael Ying Yang

ADVISOR:

Mr. Raja Ram Aryal



COUPLING AIRBORNE LIDAR AND HIGH RESOLUTION OPTICAL SENSOR PARAMETERS FOR BIOMASS ESTIMATION USING MACHINE LEARNING

KASHI RAM YADAV

Enschede, The Netherlands, March, 2019

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

Specialization: Geoinformatics

SUPERVISORS:

Dr. Subrata Nandy

Dr. Michael Ying Yang

THESIS ASSESSMENT BOARD:

Prof. Dr. ir. A. Stein (Chair)

Prof. Dr. ir. M.G. Vosselman (ITC Professor)

Dr. S.P.S. Kushwaha (External Examiner, Former Dean (A), IIRS)

ADVISOR:

Mr. Raja Ram Aryal

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author and do not necessarily represent those of the Faculty.

ABSTRACT

Forests play a vital role in the global carbon cycle by sequestering carbon from the atmosphere, thereby helping in the regulation of climate. The monitoring of aboveground biomass (AGB), which accounts for most of the stored carbon stock, is essential for the execution of the REDD+ (Reducing Emissions from Deforestation and forest Degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks) programme, that mandates regular, precise and reliable AGB estimation and its spatiotemporal variations. Remote sensing-based methods such as high-resolution optical sensor and Light Detection and Ranging (LiDAR) have been widely used to estimate forest AGB for resolving the limitations of traditional approaches. This study aims to integrate and optimise the parameters available from LiDAR and optical RapidEye datasets with the in-situ measurements for accurate estimation and mapping of AGB in tropical forests of Terai Arc Landscape (TAL) area in Nepal. This is performed using machine learning algorithms such as Random Forest (RF) and Support Vector Machine (SVM). 52 LiDAR metrics are extracted using height and intensity information. 27 spectral, including band reflectance and vegetation indices (VIs), variables are derived in addition to 8 texture measure of each band (i.e., 40 textural variables). Seven prediction models ((1) LiDAR metrics, (2) spectral variables, (3) textural variables, (4) spectral and textural, (5) LiDAR and spectral, (6) LiDAR and textural, and (7) LiDAR, spectral, and textural combined variables models) are formed to compare and select the best model using RF and SVM based regression algorithms. With the help of LiDAR returns, two canopy height models (CHM), normal CHM and pit-free CHM are created. It was observed that pit-free CHM gave better results, root mean squared deviation (RMSD) of 1.09 m for tree heights, than the normal CHM (RMSD = 1.46 m). We also observed that LiDAR, spectral and textural combined model with 119 variables performed best for AGB prediction using both machine learning algorithms. However, RF regression performed better with an R^2 of 0.95, RMSE of 35.15 Mg ha⁻¹ and RMSE_{rel} of 17.25 % compared to SVM regression with an R^2 of 0.40, RMSE of 48.29 Mg ha⁻¹ and RMSE_{rel} of 23.70 %. RF was also used for extracting an optimal number of predictor variables based on their importance. Next, 20 most important variables were used for generation of the forest AGB spatial distribution map. The output estimates were validated using 15 independent sample plots data, results for which were satisfactory ($R^2 = 0.72$, RMSE = 47.71 Mg ha⁻¹, RMSE_{rel} = 23.41 %). Moreover, the uncertainty of AGB estimation was found to be within the range between 0 to 34 Mg ha⁻¹ using Monte Carlo simulation. The result also shows that multi-sensor parameters such as near infra-red, red, red-edge (spectral bands), variance, contrast, dissimilarity, homogeneity, second angular momentum, mean (texture measure), bincntiles, relative height points count including other height metrics and percentile heights (LiDAR metrics) have strong relationship with the in-situ biomass. It is concluded that the combination of multi-sensor/source data using RF regression demonstrates to be a reliable algorithm for accurate estimation of tropical forest AGB. Based on the results of this study, it suggests that the estimation of biomass should be done using multi-sensor data coupled with field measurements with sufficient sample plots for improving accuracy.

Keywords: Forest biomass, Airborne LiDAR, Multi-sensor parameters, Random forests, Support vector machine, Uncertainty.

ACKNOWLEDGEMENTS

First of all, I would like to thank the almighty God/parents for their blessing in my life and throughout the study. I would express my sincere gratitude to Building Climate Resilience of Watershed in Mountain Eco-region (BCRWME) Project including Nordic Development Fund for providing me with the scholarship for pursuing MSc Degree. My sincere thanks go to my organization Ministry of Forests and Environment, the Government of Nepal for providing me with the great opportunity.

I am very appreciative to Dr. Subrata Nandy, my first supervisor, for his valuable suggestions, inspiration, useful feedback and comments from the initial phase to the completion of my research. I would like to express my sincere gratitude to my second supervisor, Dr. Michael Ying Yang, for his supervision and advice which was really thankful from the proposal writing to the final accomplishment of the research work. This study would not be so worthy without their regular support and suggestions.

I would like to acknowledge the IIRS (India)–ITC (Netherlands) JEP GFM course director, Dr. Sameer Saran, and course coordinator, Dr. Valentyn Tolpekin, for their incessant support in this period. Furthermore, I am glad to the organisers, faculties and teachers of both IIRS and ITC involved in the prestigious programme. It was a great occasion for me to study and experience in both reputed institutions of India (IIRS) and The Netherlands (ITC).

My special thanks go to the Director General, Dr. Deepak Kumar Kharal, of Forest Research and Training Centre (FRTC), Nepal and Mr. Basanta Gautam, Arbonaut Limited (Finland), for providing me airborne LiDAR and RapidEye datasets and also their valuable suggestions for this research work. I would like to show my gratitude to Mr. Raja Ram Aryal for his admirable advisory role from the FRTC for this work. Also, I would like to express my gratitude to Ms. Ritika Srinet, PhD scholar (FED, IIRS), and Mr. Surajit Ghosh for their support in the statistical analysis and valuable suggestions. Similarly, I wish to acknowledge Dr. Martin Isenburg (rapidlasso) for his help and suggestions during LiDAR metrics generations and Mr. Anish Joshi (Genesis) for his suggestions during image analysis. I am grateful to Mr. Dinesh Yadav (AFO) for his support to provide secondary data of the study area and Mr. Anurag Kulshrestha (PhD scholar, ITC) for his support, suggestions, and feedback in this study.

My sincere appreciation goes to all member of BCRWME Project including Mr. Raju Sapkota (PD), Mr. Chandra Dev Joshi (AO), Mr. Sher Bahadur Woli (AA) and Mukunda Raj Bhatta (PA) for their coordination and cooperation during my study. I am so grateful to my colleagues and classmates who support me in each and every moments of the study and the time we spent together in IIRS and ITC. In particular, I certainly obliged to Mr. Anirudha Mahagaonkar, Mr. Sayantan Majumdar, Ms. Shobitha Shetty, Ms. Anushree Badola, Mr. Utsav Soni, Mr. Charanjeet Nijjar, Ms. Shanti Kumari and Ms. Arunima Singh for the peer- discussions, and review during the research period.

Finally, I really want to acknowledge the support of my loving parents and relatives for their kind inspiration. Last but not least, my special and evergreen thanks and love goes to my beloved wife Anisha Yadav who support me in every step of my life and inspire me to study further. I want to express my love to my little princess, Nikita Yadav who is eagerly waiting for me to return with successful completion.

*.....dedicated to my Grand Father Phaudar Wadau and Late Father Khushi Lal Wadau
"The best source of inspiration"*

TABLE OF CONTENTS

1.	INTRODUCTION	1
1.1.	Background.....	1
1.2.	Problem statement and justification.....	3
1.3.	Research identification	5
1.4.	Research workflow.....	5
1.5.	Thesis outline	6
2.	LITERATURE REVIEW	7
2.1.	Overview of methods for forest biomass estimation	7
2.2.	Remote sensing approaches for biomass estimation	8
2.3.	Modelling-based approaches	11
2.4.	Review of related literature.....	14
2.5.	Knowledge Gaps	17
3.	STUDY AREA AND DATASET.....	19
3.1.	Study area.....	19
3.2.	Datasets.....	21
4.	METHODOLOGY	24
4.1.	Extraction of the variables from the RapidEye image.....	25
4.2.	Extraction of the CHM and other airborne LiDAR metrics.....	27
4.3.	Forest AGB estimation of the measured sample plots.....	31
4.4.	Machine learning algorithms and its accuracy assessment	32
4.5.	Model validation	35
4.6.	Uncertainty analysis.....	35
5.	RESULTS AND ANALYSIS	37
5.1.	Extracted variables from the RapidEye image	37
5.2.	Extracted airborne LiDAR metrics.....	37
5.3.	Forest AGB calculated from field measured sample plots	41
5.4.	Comparison of forest AGB prediction models using RF regression algorithm	41
5.5.	Comparison of forest AGB prediction models using SVM regression algorithm	42
5.6.	Selection of the best forest AGB prediction model with ML regression algorithm	45
5.7.	Variable importance using RF regression algorithm.....	45
5.8.	The spatial distribution pattern of the forest AGB.....	47
5.9.	Uncertainty mapping of the forest AGB.....	51
6.	DISCUSSION	52
6.1.	Selection of multisensor data and extraction of their parameters.....	52
6.2.	Comparison of normal CHM and pit-free CHM	52
6.3.	Performance of the RF and SVM ML regression algorithms with AGB prediction models	53
6.4.	Analysis of the best predictor variables.....	53
6.5.	Accuracy and uncertainty analysis of biomass estimation.....	54
6.6.	Overall analysis of the study.....	56
7.	CONCLUSIONS AND RECOMMENDATIONS.....	57
7.1.	Conclusions	57
7.2.	Recommendations.....	59

LIST OF FIGURES

Figure 1: Percentage of carbon, water and other elements contained in the wet and dry biomass (Walker et al., 2011).	1
Figure 2: Conceptual diagram.....	4
Figure 3: Active light transit time measurement technique (Vosselman & Maas, 2010).....	9
Figure 4: Basic principle of Airborne LiDAR (Vosselman & Maas, 2010).....	9
Figure 5: Illustration of (a) discrete return, (b) waveform, and (c) digitised waveform	10
Figure 6: Illustration of process for the training phase and classification phase of the RF algorithm, where i denotes samples, j denotes variables, p denotes probability, c is a class, s is a data, value presents various available values of variables, d denotes separate data used for classification, t denotes the number of trees (Belgiu & Drăguț, 2016).	11
Figure 7: Illustration of basic workflow of bagging.	12
Figure 8: Linear separable classification example of SVM.....	13
Figure 9: Application of the kernel trick in the SVM (a) non-linear relationship of weather class: sunny and snowy into a features space of latitude and longitude (b) linear relationship of weather class: sunny and snowy into a new dimensional features space of altitude and longitude.	13
Figure 10: The study area located in the tropical forests of TAL area in Kailali district, Nepal.	23
Figure 11: A portion of the study area showing general and profile view of the LiDAR point cloud data.	23
Figure 12: Methodological workflow diagram.....	24
Figure 13: (a) Airborne LiDAR returns from the four flight-lines on the trees, (b) All the first returns of Airborne LiDAR used for the interpolation, and (c) All the relevant returns of Airborne LiDAR used for the interpolation (Isenburg, 2016).	28
Figure 14: Pit-free CHM model workflow	28
Figure 15: A process for RF prediction uncertainty.	36
Figure 16: Selection of appropriate window size.....	37
Figure 17: The LiDAR strip overlap of the study area.....	38
Figure 18: The normal CHM extraction (a) DSM, (b) DTM, (c) CHM and (d) enlarge portion of the normal CHM (DSM-DTM=CHM).	38
Figure 19: The normal and pit-free CHM of the study area.....	38
Figure 20: A fitting line for the pit-free CHM derived tree heights.	40
Figure 21: A fitting line for the normal CHM derived tree heights.....	40
Figure 22: A distribution of field measured trees height and CHMs derived trees height	40
Figure 23: An optimal cost value for the different kernel functions that is used in the seven AGB predictions models sequentially: (a) linear kernel with textural model (b) RBF with LiDAR (c) polynomial with spectral (d) linear with spectral + textural (e) RBF with LiDAR + spectral (f) RBF with LiDAR + textural (g) RBF with LiDAR + spectral + textural.	44
Figure 24: The variables importance of LiDAR + Spectral +Textural combined model (119 variables) for forest AGB prediction model (all used abbreviations of variables are described in the section 4.1. and Table 6, 7 and 8).	46
Figure 25: The choice of an optimum subset of the predictor variables using 10-fold cross-validation	47
Figure 26: The accuracy of the selected optimum subset of the predictor variables	47
Figure 27: The top 20 selected predictor variables based on the increasing node purity	48
Figure 28: The top 20 selected predictor variables.....	49
Figure 29: The spatial distribution pattern of forest AGB over the study area.....	50
Figure 30: The regression line and accuracy for the validation.	50

Figure 31: The uncertainty mapping of the forest AGB over the study area.....51

Figure 32: (a) Normal CHM generation using all first returns only (b) Pit-free CHM generation using all relevant returns.....52

LIST OF TABLES

Table 1: Land use pattern (DDC, 2015)	20
Table 2: The forest types and their coverage in the Kailali districts (DFO, 2018).	20
Table 3: The description of rocks and soil texture found in the different region of the district.....	21
Table 4: LiDAR dataset information.....	22
Table 5: The list of the used software and their purposes	22
Table 6: The description and equations of the used vegetation indices	25
Table 7: The description and formula of the used texture variables (Haralick et al., 1973).	26
Table 8: Descriptions of the airborne LiDAR metrics	29
Table 9: The used equations for the calculation of the tree level AGB.....	32
Table 10: Comparison of the field measured forest tree heights with derived tree heights from pit-free-	39
Table 11: A summary of statistical values for forest tree heights	39
Table 12: The accuracy assessment of the CHMs derived forest trees heights.....	39
Table 13: The field measured forest AGB (in Mg ha ⁻¹) of 76 sample plots with their location.	41
Table 14: The descriptive statistics of field measured biomass.....	41
Table 15: Comparisons of the seven different forest AGB prediction models using RF algorithms.....	42
Table 16: The selected cost parameter and SVM-type under the four kernel functions.	42
Table 17: Testing performance of the models using linear and RBF kernel function of SVM.	43
Table 18: Testing performance of the models using polynomial and sigmoid kernel function of SVM	43
Table 19: Training performance of the LiDAR, spectral and textural combined model of AGB prediction	45
Table 20: The forest AGB prediction result and their validation.	50
Table 21: A list of the studies using ML algorithms for forest AGB estimation in different climatic zones.	55

LIST OF EQUATIONS

Equation 2.1: The default M_{xy} input variables	12
Equation 2.3: Linear kernel.....	14
Equation 2.4: Polynomial kernel.....	14
Equation 2.5: Sigmoid kernel	14
Equation 2.6: Gaussian RBF kernel	14
Equation 4.1: RMSE.....	34
Equation 4.2: $RMSE_{rel}$	34
Equation 4.3: $RMSE_{CV}$	34
Equation 4.4: R^2	34
Equation 4.5: RMSD	34
Equation 4.6: $RMSD_{rel}$	34
Equation 4.7: $RMSD_{CV}$	35

LIST OF APPENDICES

Appendix 1: Specification of RapidEye system.....	68
Appendix 2: Used commands script of quality checking, data preparation, and LiDAR metrics extraction using airborne LiDAR data	68
Appendix 3: The generated lasinfo report for the basic information and quality checking of airborne LiDAR point cloud data	72
Appendix 4: List of the species and their model coefficient.	73
Appendix 5: List of the species and their wood density.....	74
Appendix 6: The generated spectral variables including vegetation indices and band reflectance.....	75

LIST OF ACRONYMS

AGB	Aboveground Biomass
BGB	Belowground Biomass
Ca	Calcium
CF	Community Forests
CFM	Collaborative Forests Management
CHM	Canopy Height Model
CPA	Crown Projection Area
CO ₂	Carbon dioxide
COP	Conference of the Parties
DBH	Diameter at Breast Height
DDC	District Development Committee
DFO	District Forests Office
DFRS	Department of Forests Research and Survey
DSM	Digital Surface Model
DTM	Digital Terrain Model
FAO	Food and Agriculture Organization of the United Nation
FCPF	Forest Carbon Partnership Facility
FCD	Forest canopy Density
FRTC	Forest Research and Training Centre
GHGs	Greenhouse Gases
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GoN	Government of Nepal
GPS	Global Positioning System
IMU	Inertial Measuring Unit
INS	Inertial Navigation System
IPCC	Intergovernmental Panel on Climate Change
LAS	Log ASCII Standard
LiDAR	Light Detection And Ranging
LAMP	LiDAR-Assisted Multi-source Programme
MPFS	Master Plan for Forestry Sector
Mg	Magnesium
MoFSC	Ministry of Forest and Soil Conservation
MoFE	Ministry of Forest and Environment
MRV	Monitoring Reporting and Verification
N	Nitrogen
NASA	The National Aeronautics and Space Administration
NDVI	Normalize Difference Vegetation Index
OWL	Other wooded lands
OBIA	Object-based image analysis
P	Phosphorus
PCTMCDB	President Chure-Terai Madhesh Conservation Development Board
RADAR	Radio Detection and Ranging
RBG	Red, Green and Blue
RBF	Radial Basis Function
RED	Reducing Emissions from Deforestation
REDD	Reducing Emissions from Deforestation and Forest Degradation

REDD+	Reducing Emissions from Deforestation and Forest Degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks
RF	Random Forests
RMSE	Root Mean Square Error
R-PIN	Readiness Plan Idea Note
RPP	Readiness Preparation Proposal
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
TAL	Terai Arc Landscape
TIN	Triangular Irregular Network
UNFCCC	United Nations Framework Convention for Climate Change
UN-REDD	United Nations Collaborative Program on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries

1. INTRODUCTION

1.1. Background

Forests can store significant amounts of carbon in their biomass, contributing enormously to carbon sequestration from the atmosphere (Gibbs et al., 2007). From leaves to woody trunks and roots, every part of a tree has a certain ability to trap carbon in its tissues, which is very high when compared to other terrestrial components of an ecosystem. This phenomenon is considered to be a boon for the current climate change scenario (Gibbs et al., 2007). The forests sequester more carbon during their growth, healthy and in the sustainable-managed stage in contrast to the condition when the forest resources undergo deforestation and forest degradation (Gibbs et al., 2007; Walker et al., 2011). Food and Agriculture Organization (FAO, 2015) found that global carbon emissions from the deforestation and forest degradation are estimated to be around 2.9 billion tonnes of CO₂ per year between 2011 and 2015. Whereas, 2.1 billion tonnes of CO₂ was utilized annually by forest resources for growth during the same period. From carbon utilization and release cycle, it is evident that forest biomass plays an important role in the global carbon cycle (FAO, 2015; Walker et al., 2011).

The biomass of the forest denotes the mass of living plant tissue. Total forest biomass, consisting of aboveground biomass (AGB) and belowground biomass (BGB), together includes the stems, branches, leaves; and roots (Kindermann et al., 2008; Walker et al., 2011). FAO (2010) identified five major carbon pools of forest ecosystem namely AGB, BGB, dead wood, litter and soil organic matter. In forest carbon stock, tree biomass or AGB is recognized as the largest carbon pool and its regular monitoring is crucial for the execution of REDD+ (Reducing Emissions from Deforestation and forest Degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks), that mandates precise measurement of AGB and its spatiotemporal variations (FAO, 2010; Hajar et al., 2015). Walker et al. (2011) found that wet AGB generally contains 50% water, 25% carbon and the rest 25% are other elements like Nitrogen (N), Phosphorous (P), Potassium (K), Calcium (Ca), Magnesium (Mg) along with some additional trace elements. They also specified that dry AGB contains 50 % carbon and 50 % other additional elements. A pictorial representation of components of dry and wet AGB is presented in Figure 1 (Walker et al., 2011).

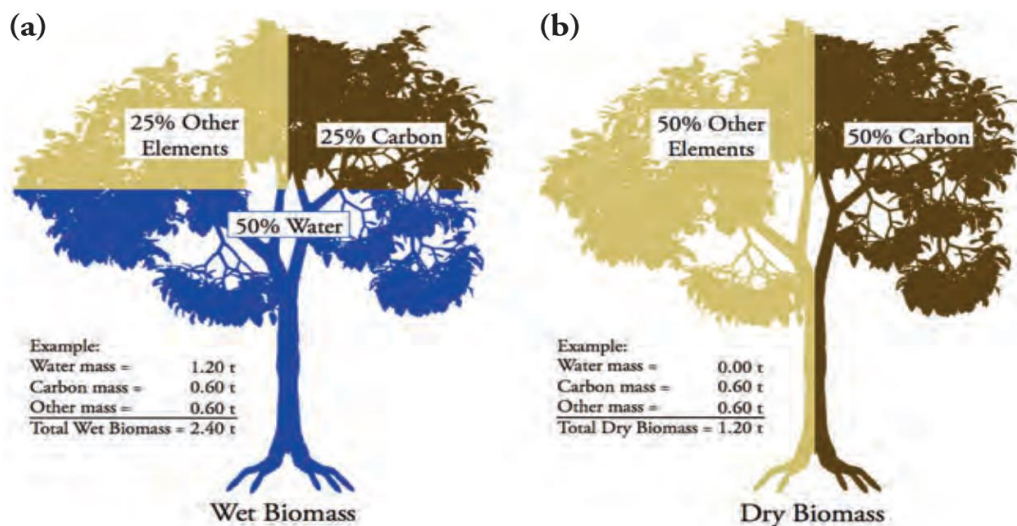


Figure 1: Percentage of carbon, water and other elements contained in the wet and dry biomass (Walker et al., 2011).

FAO (2010) reported that the world's forest area is about 4 billion hectares which accounts for almost 31% of the total land area. It is also estimated that the global rate of deforestation and loss due to natural factors is found to be around 16 million hectares annually in 1990 and about 13 million hectares annually in 2005 (FAO, 2010). Moreover, tropical forests covered about 50% of the world's total forest biomass (Kindermann et al., 2008). On the other hand, Asia and the Pacific's region hosts around 734 million hectares (i.e., 26 % of the total land area) and it is also reported that a net loss of the forests is about 0.6 million in 1990 and a net gain is over 2.2 million hectares of forests between 2000 and 2010, annually, as result of large-scale afforestation (FAO, 2009, 2010). Nepal is a small country covering 14.71 million hectares area in totality but it has 6.61 million hectares (i.e. 44.74 %) of forests and other wooded lands (OWL). In Nepal, the deforestation rate found 1.7 % and shrub or OWL coverage increasing rate was 7.4 % per year in the period between 1978/79-1994 (Department of Forest Research and Survey (DFRS), 1999). However, recent national forest inventory revealed that forest cover has increased by a rate of 2.33 % and shrub or OWL has decreased by 3.44 % during 1994-2010/2011(DFRS, 2015).

Majority of climate scientists are convinced that anthropogenically produced greenhouse gases (GHGs) are the main cause for global warming and/or climate change (National Aeronautics and Space Administration (NASA), 2018). Intergovernmental Panel on Climate Change, IPCC (2007) report indicated that the forestry sector (including deforestation) contributed 17.4 % of total anthropogenic GHGs emissions in terms of carbon dioxide equivalents in 2004. The United Nations Framework Convention on Climate Change, UNFCCC (1994) and the Kyoto Protocol (2005) are jointly initiated to mitigate the growth and stabilise atmospheric GHGs. In this protocol, for Annex I Parties (i.e., developed countries), there is defined a target to reduce the GHGs emissions at the mean rate of 5 % from the baseline year of 1990 over the commitment period between 2008-2012 (UNFCCC, 2008). The Bali Action Plan Conference of the Parties 13 (COP13) of UNFCCC in 2007 has developed a significant mechanism for reducing emission from deforestation and forest degradation (REDD) (initially started from reducing emission from deforestation (RED) in 2005 (COP11)) to ensure participation of developing countries for forest carbon financing. In the COP15 of UNFCCC in Copenhagen, the REDD mechanism was extended with the comprehensive scope as REDD+ (Cerbu, Swallow, & Thompson, 2011; UNFCCC, 2009). In the post-Kyoto climate change agreement, REDD+ is recognised as a cost-effective and efficient mechanism to limit or reduce GHGs emission. Measurement, reporting and verification (MRV) system should be reliable, credible, efficient, effective and affordable for the carbon financing or credits through the REDD+ mechanism for any country. MRV of forest carbon stock is a significant system for the proper implementation of REDD+ and it is also emphasised in the meeting on the UNFCCC after the first commitment period (2008-2012) of the Kyoto Protocol in 2012 (Hajar et al., 2015).

Nepal is a signatory of UNFCCC, the Kyoto Protocol and the Paris Agreement and has contributed to REDD (later extended as a REDD+) as one of the 14 pioneer countries for combating global climate issues since 2008. Forest Carbon Partnership Facility (FCPF) of World Bank, UN-REDD, and some additional bilateral and multilateral partners have been financially and technically supporting Nepal for its REDD readiness activities (Hussin et al., 2014; Ministry of Forests and Environments (MoFE), 2018). The government of Nepal has already submitted Readiness Plan Idea Note (R-PIN) in March 2008 and Readiness Preparation Proposal (R-PP) in April 2010 to FCPF, and FCPF Participants Committee has also endorsed the R-PP in June 2010. Additionally, a Mid-Term Report of R-PP progress is submitted to the FCPF in December 2013 which revealed progress in various regions like an assessment of land use and drivers, forest law and governance, arrangement and management in country level, and national monitoring system. Similarly, the National Forest Reference Level is also submitted to UNFCCC in 2017 for the review and Emission Reduction Program Document at sub-national level is on progress for 12 districts of the Terai Arc Landscape (TAL) (Figure 10). Nepal is actively involved in REDD+ readiness activities and its implementation process with the appropriate instructional arrangements as per its national /international level of the commitment (MoFE, 2018). Therefore, it is vital for Nepal to develop an MRV system which is

reliable, credible, efficient, effective and affordable for the carbon financing under the REDD+ mechanism. Being a member of REDD+, it is mandatory for Nepal to estimate and verify the forest biomass and also to account for its spatiotemporal change at the national level.

1.2. Problem statement and justification

MRV system is necessary to ensure transparency of the assessment and estimation process carried out under the paradigm of the REDD+ program (Hajar et al., 2015). In principle, destructive and non-destructive sampling techniques are used to estimate forest AGB. Despite high accuracy, the destructive technique needs more resources and labours. Conventionally, this technique was used for developing an allometric equation but it can be generated also from forest assessment data using a non-destructive technique. In the traditional non-destructive methods for estimation of forest biomass, field-based measurements are used in combination to species-allometric equations and extrapolated to the entire area under consideration. However, this method is limited to small forest areas (Jenkins et al., 2003; Mohd Zaki & Abd Latif, 2017). Remote sensing (RS) based methods like optical, Radio Detection and Ranging (RADAR), Light Detection and Ranging (LiDAR) have been used extensively as the best alternatives to the destructive method, to overcome the limitation of traditional approaches due to their wide-ranging spatiotemporal coverage, repetitiveness and resource efficiency features (Baccini et al., 2004; Hall et al., 2006; Jenkins et al., 2003; L. Kumar & Mutanga, 2017; Kumar et al., 2015; Mohd Zaki & Abd Latif, 2017; Nandy et al., 2017; Nandy et al., 2019).

Recommended monitoring system is a combination of RS based methods and field-based measurements for estimating greenhouse gases, forest carbon stock and forest cover changes (Hajar et al., 2015). Optical RS can deliver the different scale of imagery from low-resolution to very high-resolution for prediction of forest AGB, and accuracy estimation increases with increasing resolution and vice-versa. Combining RS techniques with field measured data provides the best results and also make it economically viable, especially in the case of large forest area (Baccini et al., 2004; Hall et al., 2006; Jenkins et al., 2003; L. Kumar & Mutanga, 2017; L. Kumar et al., 2015; Mohd Zaki & Abd Latif, 2017; Nandy et al., 2017; Nandy et al., 2019). However, optical RS is often limited by the presence of clouds that affects the penetration of radiations. Active RS based methods (like Synthetic aperture radar (SAR) and LiDAR) provide a better solution to overcome these problems due to their independence over the weather, day-night conditions, and ability to penetrate into cloud and vegetation (Dhanda et al., 2017; Sinha et al., 2015). Generally, L-band and P-band of SAR have been used to estimate forest AGB where accuracies are lower due to saturation especially in the dense forest areas and also due to sensitivity towards soil conditions where vegetation is lower (He et al., 2012; Mitchard et al., 2012). Alternatively, LiDAR is the best capable technique to measure AGB because it gives (Vosselman & Maas, 2010) more accurate (up to centimetre level) canopy height model (CHM) derived from the high-density LiDAR data. It can also provide correct and reliable estimates of biomass without saturation where high biomass is present (Dhanda et al., 2017; Mitchard et al., 2012; Næsset & Gobakken, 2008; Vosselman & Maas, 2010; Zhao et al., 2009a). However, LiDAR cannot provide sufficient spectral information as optical sensors. Therefore, utilization of multi-sensor data using modern integration algorithms can enhance the quality of estimations of the forest AGB (Dhanda et al., 2017; Lu et al., 2016; Mohd Zaki & Abd Latif, 2017).

Machine learning (ML) algorithms such as random forests (RF) and support vector machine (SVM) are non-parametric and non-linear regression algorithms that can be applied to optimise the parameters extracted from multi-sensor data (Dhanda et al., 2017). In comparison to simple regression technique, RF classification and regression algorithm are more precise and robust (Belgiu & Drăguț, 2016; Breiman, 2001; Dang et al., 2019; Dhanda et al., 2017). Also, the SVM ML algorithm has been extensively used for classification and regression purposes in forestry. It provides an accurate and reliable result of forest AGB estimation using a few samples as training data (Dhanda et al., 2017; Mountrakis, Im, & Ogole, 2011). Dhanda et al. (2017) determined that RF and SVM algorithms offer similar performance if no underlying

correlation of variables exists in the dataset. Notably, SVM offers better performance on increasing correlation with the predicted variable.

Very few studies have been carried out in tropical forests due to difficulty in accessing all parts of the forest area in addition to their complex structure, species richness, composition and diversity, making it difficult to precisely estimate and extrapolate required information (Ghosh & Behera, 2018; Goodale et al., 2002; Houghton et al., 2009; Kushwaha & Nandy, 2012; Sinha et al., 2015). Although forest AGB can be measured and extrapolated in a tropical forest by applying a multi-sensor or multi-resolution data approach, there may be additional costs involved (Sinha et al., 2015). Generally, uncertainty is involved in precisely estimating, monitoring, and reporting of forest biomass in many tropical forests. Precise and reliable information of forest biomass is mandatory for the implementation of climate change mitigation policy like REDD+ (Gibbs et al., 2007; Houghton, 2005; Kumar et al., 2014; Rodríguez-Veiga et al., 2016). Mohd Zaki & Abd Latif (2017) concluded that the precise and reliable estimation of forest AGB depends on many factors such as sources of data, used sensors, forest types and its conditions, used models, methods of processing and circumstances of climate.

This research is aimed to integrate and optimise multi-sensor data (LiDAR data and high-resolution optical image) and field-measured data using ML regression algorithms, SVM and RF, to predict the tropical forest AGB. For this purpose, multi-sensor parameters are integrated and optimized using the ML regression algorithm and compared for their quality of outputs. Moreover, the spatial distribution pattern of the AGB is generated over the study area and their uncertainty is also analysed. The conceptual diagram of the proposed research is shown in Figure-2.

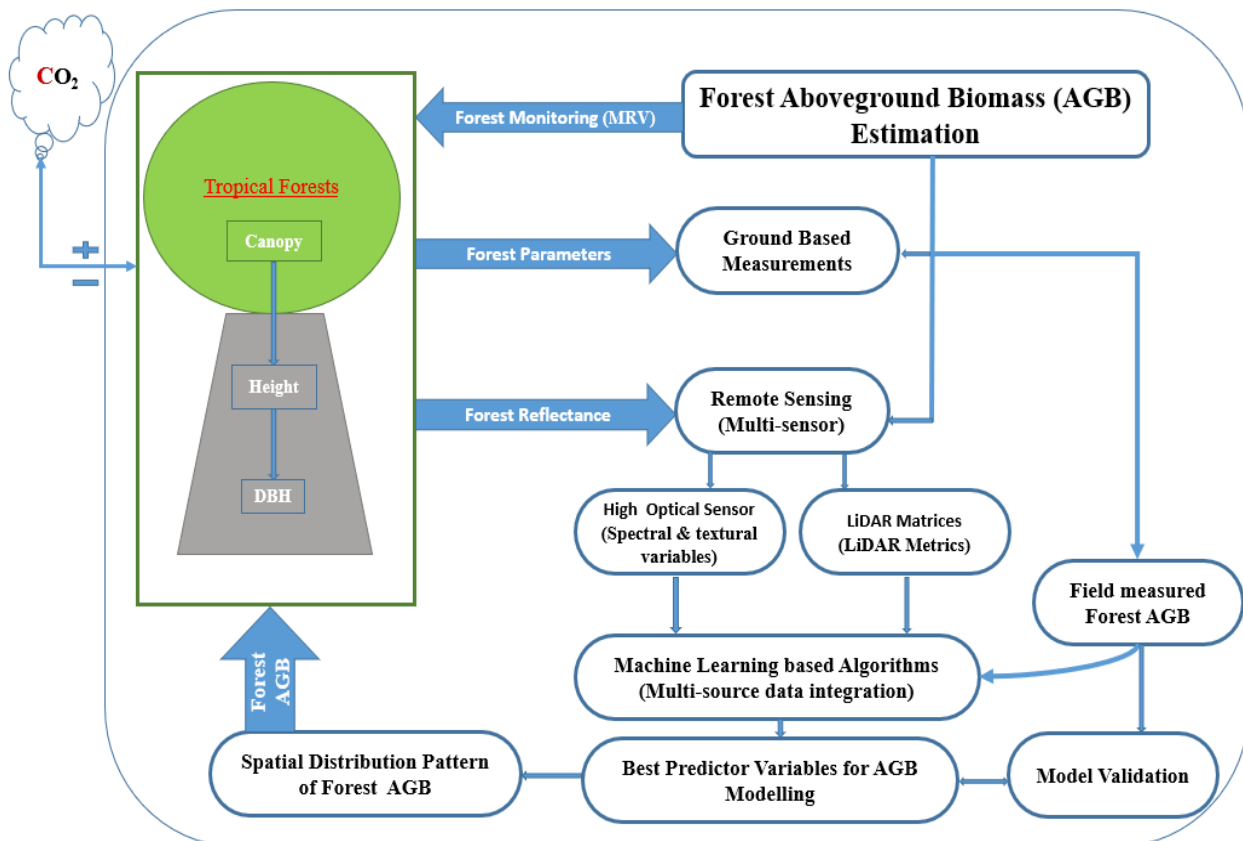


Figure 2: Conceptual diagram

1.3. Research identification

The main focus of this research is to predict the forest AGB by integrating and optimizing variables extracted from the remotely sensed LiDAR and optical RS products with reference to the field-measured biomass information using ML algorithms, and to assess the efficiency of different ML algorithms and biomass prediction models used in the study.

1.3.1. Research objectives

The overall objective of this research is to optimise the multi-sensor parameters for estimating AGB and to assess the performance of SVM and RF ML regression algorithms. The specific objectives are as follows:

1. To extract the different spectral and textural variables from the high-resolution optical satellite data.
2. To extract the LiDAR metrics from airborne LiDAR data.
3. To evaluate the efficiency of the SVM and RF for forest AGB estimation.
4. To optimise the optical and LiDAR data derived variables for forest AGB estimation using the best-performing ML regression algorithm between SVM and RF.
5. To present the spatial distribution of forest AGB and map its uncertainty over the study area.

1.3.2. Research questions

The following research questions are proposed to achieve the above objectives:

- ❖ Specific objective 1:
 - a) What are the most appropriate spectral and texture variables for forest AGB estimation?
- ❖ Specific objective 2:
 - a) What are the suitable LiDAR metrics for forest AGB estimation?
 - b) What is the difference between the normal CHM and pit-free CHM?
- ❖ Specific objective 3:
 - a) Which ML regression algorithms perform the best?
- ❖ Specific objective 4:
 - a) How can the LiDAR, spectral and texture variables be optimised for AGB estimation using the ML regression algorithm?
 - b) What is the optimal subset of predictors variables out of the optimised variables?
- ❖ Specific objective 5:
 - a) What is the spatial distribution pattern of the predicted forest AGB over the study area?
 - b) What is the accuracy of the forest AGB distribution over the study area?
 - c) What is the range of uncertainty of the estimated forest AGB over the study area?

1.3.3. Innovation

As a part of this research, it envisions to use airborne LiDAR and optical data, integrated using ML algorithms such as SVM and RF, for precise estimation of forest AGB. With a combination of newly introduced variables and those used in the past, it intends to develop effective AGB prediction models and test them for forest areas of Nepal, where a study using such a modern approach hasn't been performed so far. This addresses the novelty in our proposed research. Further, it also intends to analyse the differences between tree heights measured using in-situ methods and that obtained from normal canopy height model (normal CHM) and pit-free CHM model.

1.4. Research workflow

The research problems is defined after reviewing the relevant literature, and aimed to target the tropical forests area where limited study has been undertaken compared to other forests areas (Ghosh & Behera, 2018; Goodale et al., 2002; Houghton et al., 2009; Sinha et al., 2015) especially in the developing countries like Nepal. The formulated research objectives and research questions are based on the basis of the research

problem. Height structure information from the airborne discrete LiDAR data and spectral reflectance from the high-resolution optical sensor image are used in combination with field measured information in this study, taking into consideration the structure complexity, density variability, and heterogenous conditions of the tropical forest (Lu et al., 2016).

In this context, LAStools with different models is used to extract the optimal LiDAR metrics using both height and intensity information of the pulse returns with the help of customising command line script (Isenburg, 2016). Gray level co-occurrence matrix is used to derive textural information using all the available bands with an appropriate kernel size. The maximum reliable spectral (both the band reflectance and vegetation indices) information is also derived. Using the derived variables from both the dataset, all the possible combinations are designed to form the AGB prediction models for further processing.

The ML algorithms are executed using the different AGB prediction models with their various required parameters to get the best result. The best AGB prediction model along with the ML regression is selected based on their accuracy assessment. Additionally, selected AGB prediction model with the ML regression algorithm is further implemented to prioritise the optimal number of variables for predicting the forest AGB. By getting the optimal number of the extracted variables, they are used to map the spatial distribution pattern of the biomass. Independent field-measured plots (i.e., not used during the ML regression algorithms implementation) are used to validate the result using appropriate validation measures. Finally, uncertainty analysis is also carried out for the indication of the possible sources of the error and making the strategy in future to improve the accuracy of forest biomass estimation (Feng et al., 2017).

1.5. Thesis outline

The thesis is organised in seven chapters. Chapter 1 presents the background for the motivation of the research, statement and justification of the problem, and research identification presented as research objectives and research questions. Chapter 2 covers the literature review where an overview of the application of the remote sensing techniques with both the optical sensor and airborne LiDAR, and the modelling approach in the forest biomass estimation are included. Chapter 3 presents an explanation of the study area and dataset used to achieve the proposed objectives. Chapter 4 describes the detailed methodology for the whole research. Further, chapter 5 focuses on the outcomes of the research (result), while chapter 6 emphasises on the discussion of the achieved results. Finally, chapter 7 summarises the entire study as conclusion to the thesis . Recommendation for future studies and research in this area is also presented in this chapter.

2. LITERATURE REVIEW

This chapter summarises the science behind the application of RS techniques and modelling approaches for forest biomass estimation. A note on different applications of RS and the use of reference data is provided in the first section of this chapter. This is followed by various models that are used for obtaining the best estimates for forest biomass are presented and discussed. A review of previous studies that have worked with forest biomass modelling is presented along with a brief explanation of uncertainty analysis. Studies with a focus on Nepal have also been included at the end of this chapter.

2.1. Overview of methods for forest biomass estimation

Apart from the traditional field-based approaches for biomass estimation, conventional RS methods have taken the lead lately, being implemented extensively for region-wide estimation of forest biomass. Forest scientists are convinced in a common understating that field-based methods for estimating forest biomass are most accurate than any other. However, such an approach may be time-consuming, labour-intensive and it is limited to small geographic area (Jenkins et al., 2003; Kushwaha et al., 2014; Lu et al., 2016; Mohd Zaki & Abd Latif, 2017). Also, process-based ecosystem models and GIS-based empirical models have been used for such estimations in a limited case, but they have higher levels of uncertainty. Variable forest conditions, types of data used, quality of ancillary data, a spatial resolution of adopted products, dependency on dynamic environmental factors and inaccuracy of models are the major sources of error and uncertainties. Despite these, several studies that were based on RS have reported that they are beneficial than other approaches including the process and empirical based models (Lu et al., 2016; McRoberts et al., 2013) for their economic feasibility, repetitively and wide coverage (Hall et al., 2006; Kumar & Mutanga, 2017; Kumar et al., 2015; Mohd Zaki & Abd Latif, 2017). However, the RS-based method cannot directly measure AGB, but only provide information (like vegetation conditions, spectral characteristics and textural information) that can be used for modelling the forest biomass (Gibbs et al., 2007; Zianis et al., 2005). The modelled results need to be validated to ensure the quality of an estimate, therefore recommending the used of remotely sensed products in combination with field measure/surveyed data (Hajar et al., 2015).

Field collected information of the standing biomass is vital for accurately modelling the required information (Gibbs et al., 2007; Zianis et al., 2005). For the choice of variables, assessing the quality of model outputs, fine-tuning the models and statistical analyses field collected data is valuable and used as a reference in such cases. This data is generally collected by traditional destructive sampling, allometric equations and volume conversion methods (Lu et al., 2016; Lu & Lu, 2006). In the destructive sampling method, trees are harvested, dried and all parts of the trees are weighed to quantify the biomass. This method is very accurate and it is used for generating allometric equations. But it is physically intensive, time and resource consuming and it is not applicable for a large area. Subsequent, allometric equations are developed for each group of species with linear and non-linear regression equations based on tree height, DBH (i.e. diameter at breast height) and wood density. In this method, there is no need to destroy the trees and past field-measured data can be used for the estimation of the biomass but its major disadvantage is limited species have its own allometric equations. Similarly, conversion from volume is another method where forest biomass can be modelled using the volume of a tree-level or plot-level with the help of volume expansion factor, related wood density and AGB expansion factor, limited by species compositions and environmental circumstances (Chave et al., 2014; Henry et al., 2010; Jenkins et al., 2003; Lehtonen et al., 2004; Mohd Zaki & Abd Latif, 2017; Segura & Kanninen, 2005).

2.2. Remote sensing approaches for biomass estimation

2.2.1. Optical remote sensing

Modern remote sensing missions provide datasets of synoptic scales with high spectral, spatial, temporal and radiometric resolution. With a huge pool of datasets, this technique has emerged to be one of the most preferred for modelling and monitoring forest biomass. Apart from modelling and monitoring, these products can also be used for analysing degradation of forests and resource mapping (Kumar et al., 2015; Mohd Zaki & Abd Latif, 2017). Very high-resolution imagery like worldview-2, Quickbird, IKONOS and GeoEye-1 are suitable for identifying variable forest inputs to the development of allometric equations. Similarly, medium and/or coarse resolution imagery like LISS-III, Sentinel, Landsat, MODIS and NOAA-AVHRR imagery is more appropriate for forest monitoring and change detection on a regional and global scale (Andersson et al., 2009). Moreover, Landsat, SPOT, Sentinel, ASTER, MODIS, and NOAA-AVHRR imagery are also frequently used to estimate of forest AGB using different models and/or ancillary data at different scales (Baccini et al., 2004; Lu, 2006).

Identification, extraction, and selection of appropriate variables from RS imagery are an essential and crucial task for forest AGB estimation using these modern techniques. Although RS is a key source to provide data for AGB estimation, it does not provide a direct estimation of the forest AGB. Therefore, it is mandatory to extract different variables like spectral and textural variables using suitable techniques. Vegetation indices and texture measure, principal component analysis, minimum noise fraction transform, and spectral mixture analysis are being used as major methods for the extraction of different variables from optical remote sensing data (Lu, 2006). Lu et al. (2016) explored that near-infrared (NIR) compared to the shortwave infrared wavelength of vegetation indices do not have a stronger relationship with forest AGB due to variability and complex structures of the forest. However, NIR showed a stronger relationship with forest AGB in the forest with lower variability, complexity, and adverse soil condition.

Different statistical approaches like first order (where do not consider pixel neighbour relationship), second order (considering the relationship between two pixels) and third/higher order (considering the relationship between three or more pixels) are existing to measure texture variables. However, the second order statistical approach which is gray level co-occurrence matrix (GLCM) techniques are commonly used for texture analysis (Lu, Batistella, & Moran, 2005). The texture variables (Lu et al., 2005) including entropy, correlation, contrast, mean, dissimilarity, homogeneity, second angular moment and variance have been used with moving kernel size like 5x5, 7x7, 9x9, 11x11, 15x15, 19x19 and 25x25 for each spectral bands to find the relationship between AGB and texture in different types of forest using GLCM techniques. They concluded that texture variables have a high correlation with mature forest than spectral variables because of high variability and complexity of forest. However, it has poor correlation with comparatively simple and less complexity of forest stands. Considering the heterogeneity characteristics of the forest, it is better to use both spectral and textural variables for the estimation of the forest AGB (Lu et al., 2005).

Despite the many advantages of optical remote sensing data, it also has some limitations for quantifying forest biomass. The past study showed that the resulting forest AGB measurement is underestimated or appeared as a low performance due to the data saturation problem especially in the forests which have high-density AGB (Lu et al., 2016, 2012a). The data saturation may vary depending on the resolution of the imagery (Lu et al., 2012a). Similarly, the developed AGB model based on optical spectral variables (Lu et al., 2005) cannot be directly used in other areas because of instability in spectral signatures due to the complexity of biophysical environments and heterogeneity of forest conditions such as atmospheric condition, soil moisture, vegetation composition, growth dynamism and phenology. Moreover, optical remote sensing is more appropriate for the extracting horizontal forest features like forest coverage, types, canopy cover, and canopy density. However, it is not appropriate for the estimation of canopy height which is a significant variable for the forest AGB estimation (Lu et al., 2016).

2.2.2. Airborne LiDAR

LiDAR is the process for measuring the time delay between emission of a pulse of laser energy and reception of a returned laser pulse after being reflected by an object. It is an appropriate method for calculating the distance between the sensor and the object and works based on active light transit time measurement technique of three-dimensional surface estimation optically as shown in Figure 3 (Vosselman & Maas, 2010).

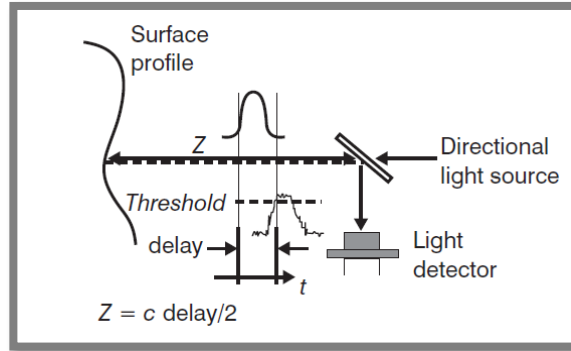


Figure 3: Active light transit time measurement technique (Vosselman & Maas, 2010)

An airborne LiDAR system operates on either an aircraft or a helicopter and comprises of three major components namely the LiDAR sensor, Inertial Navigation System (INS) or Inertial Measurement Unit (IMU) and Global Navigation Satellite System (GNSS) as illustrated in Figure 4. It measures point densities ranges from 0.2-50 points per square meter. Similarly, GNSS or Global Position System (GPS) measure the position of an aircraft with the help of both aircraft based as well as and ground-based receiver. It operates based on kinematic differential positioning at a frequency range of 1-2 Hz. Finally, INS/IMU measure the orientation of the airborne platform by using integration of acceleration with the help of gyroscopes and accelerometers at the rate of frequency ranges from 40-200 Hz. (Vosselman & Maas, 2010).

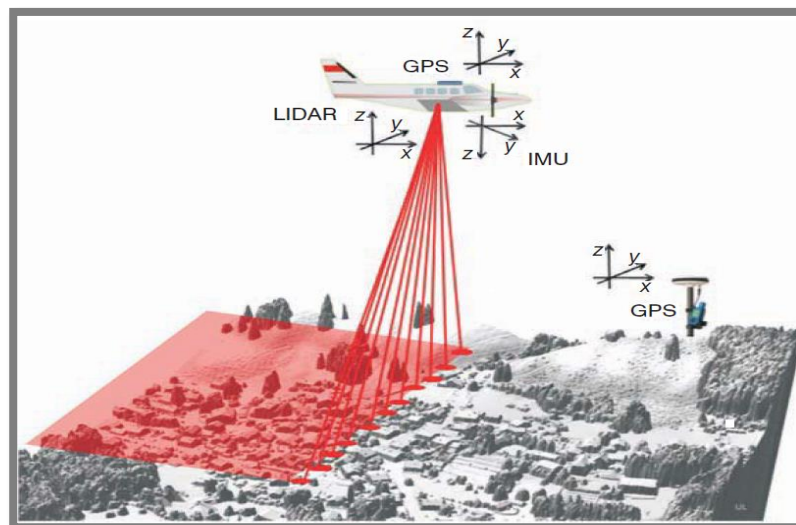


Figure 4: Basic principle of Airborne LiDAR (Vosselman & Maas, 2010).

Airborne LiDAR works at wavelength ranges from 800-1500 nanometre where forest reflectance is high for the different purpose of terrestrials applications. On the basis of recorded LiDAR returned signals, LiDAR sensor generates either discrete return or waveform. In discrete return, LiDAR

sensor delivers only peak information of the returned pulse and do not expose its shapes. On the other hand, the sensor provides the information as digitization of reflected signal of an emitted pulse where multiples echoes can be found in the full waveform, as shown in Figure 5 (Lu et al., 2012a; Vosselman & Maas, 2010).

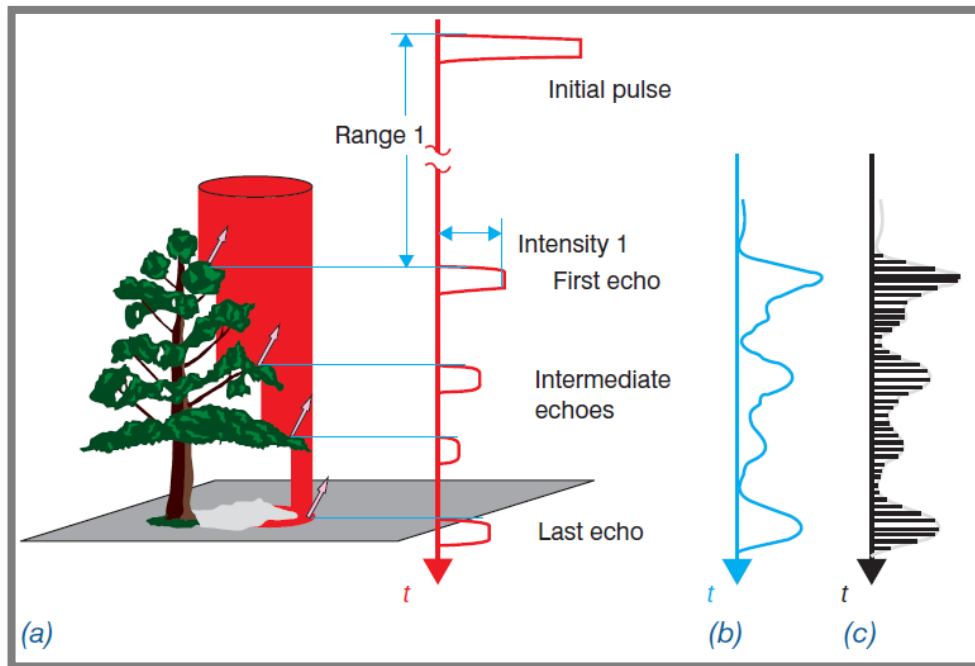


Figure 5: Illustration of (a) discrete return, (b) waveform, and (c) digitised the waveform (Vosselman & Maas, 2010).

LiDAR is a robust remote sensing technique to estimate forest AGB at the landscape level because it provides up to centimetre level accuracy of measuring vertical vegetation structures like canopy height and other relevant height metrics. It is capable to measure AGB of the high-density forest without any saturation problem where optical and RADAR RS have failed (Mitchard et al., 2012; Næsset & Gobakken, 2008; Zhao et al., 2009). Additionally, major advantages of using airborne LiDAR are high canopy penetration capacity (even a small gap is enough to detect vertical structure and obtaining ground elevation), high measurement density (about 30 measurement per square meter), higher data accuracy in both horizontal (20-100cm) and vertical (5-20 cm) measurement, fast data acquisition capacity (day and night working capability), and required least ground truth (Vosselman & Maas, 2010).

Generally, LiDAR metrics can be derived using height (i.e. height of different pulse returns or echoes) and intensity (i.e. strength of pulse backscattering) information of the point clouds on the basis of either individual tree or plot label. For the identifying individual tree from airborne LiDAR data needs point density (8-10 or more points per meter). In contrast, plot level LiDAR metrics can be extracted from sparse point density and it is widely used in the forestry applications. Normal CHM, pit-free CHM, percentile height, canopy cover, canopy density, mean, maximum, minimum, standard deviation, skewness, kurtosis are the common LiDAR metrics used in the forestry for different applications including forest AGB and carbon estimation (Casas et al., 2016; Chen Qu, 2013; Dhanda et al., 2017; Kim et al., 2009; Wing et al., 2015).

2.3. Modelling-based approaches

2.3.1. Machine learning algorithms

Parametric and non-parametric process are the major two group of algorithms which are used for forest AGB estimation. Parametric algorithms suppose fundamental statistical distribution in the data and it has a fixed number of variables with fixed meaning like simple or multiple linear equations. On the other hand, non-parametric algorithms do not depend upon any statistical distribution and it has infinite dimensional unknown variables such as the non-linear forest AGB model. Generally, parametric algorithms could not deal with the interrelation between remote sensing data variables and forest AGB due to its complexity but it can be handled by non-parametric equation (Lu et al., 2016). RF and SVM are the non-parametric algorithms generally used to integrate and optimise the extracted variables from multi-sensor data for forest AGB estimation (Dhanda et al., 2017; Lu et al., 2016; Nandy et al., 2017).

2.3.1.1. Random forests

RF is an ensemble classifier developed by Breiman (2001). An ensemble comprises a group of individually trained classifiers called decision trees which predicts based on voting and it is mostly used to improve the performance of classification and regression. The ensemble technique creates homogeneous or heterogeneous ensembles based on the use of single or multiple learning algorithm. RF ensemble classifier (Belgiu & Drăguț, 2016) creates numerous decision trees with the help of the subset of training samples and variables that are randomly chosen (Figure 6). It uses a group of classification and Regression Trees (CARTs) for making the prediction which is created by the subset of the training data using the bagging approach (Breiman, 2001). The bagging is an acronym word created from the bootstrap and aggregation which reduces the error, its basic workflow shown in Figure 7 (Breiman, 1994; Saini & Ghosh, 2017).

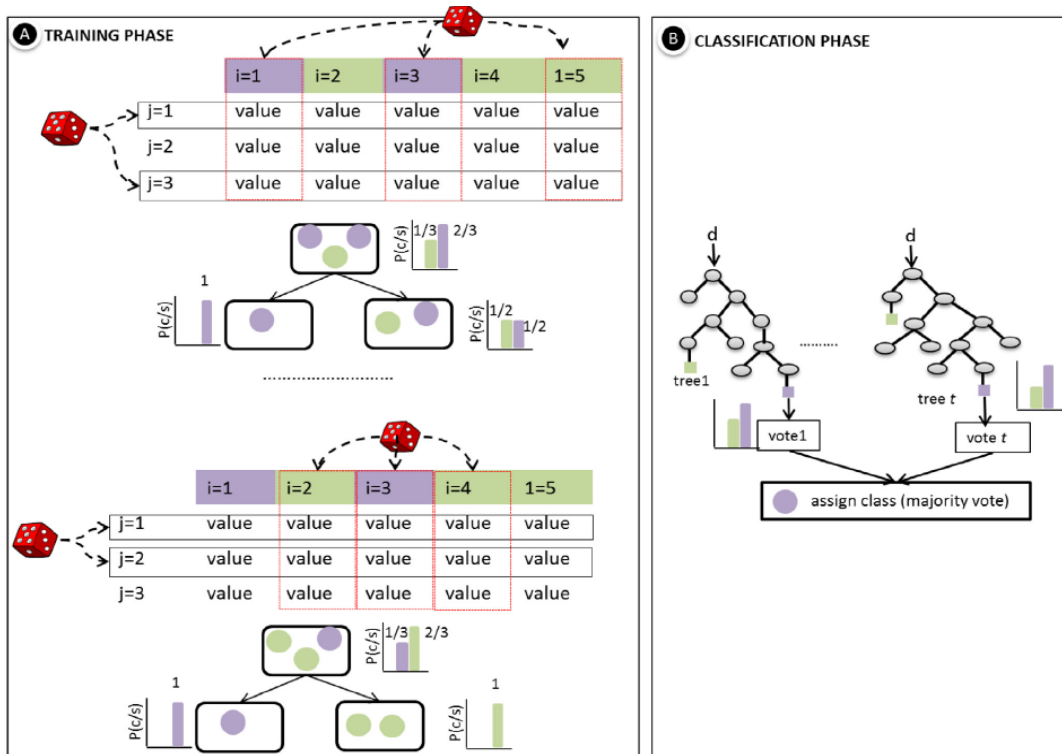


Figure 6: Illustration of process for the training phase and classification phase of the RF algorithm, where i denotes samples, j denotes variables, p denotes probability, c is a class, s is a data, value presents various available values of variables, d denotes separate data used for classification, t denotes the number of trees (Belgiu & Drăguț, 2016).

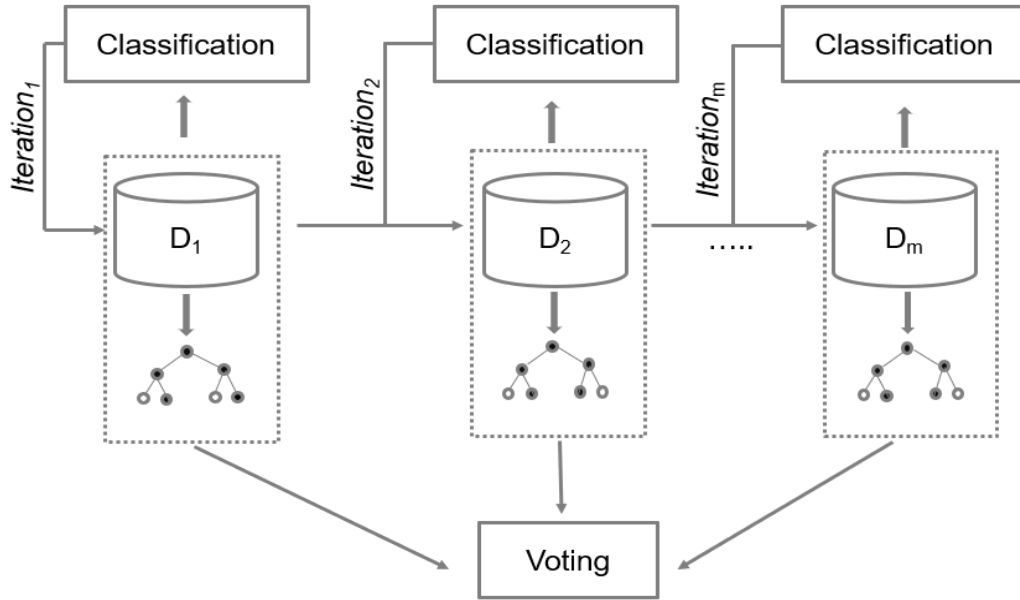


Figure 7: Illustration of the basic workflow of bagging.

Generally, $2/3^{\text{rd}}$ of the samples are called *in-bag* samples which are used for the training purpose of the trees and the remaining $1/3^{\text{rd}}$ of the samples are called *out-of-bag* (OOB) samples that are used for the internal cross-validation method to determine the model error (i.e., also called OOB error). The number of trees called N_{tree} and the number of variables called M_{try} are the two important input parameters that desire to define by the user for making the forest trees. The N_{tree} are created independently with help of subset of the training data without pruning and each node of the created decision tree is divided using available M_{try} (Belgiu & Drăguț, 2016; Breiman, 2001). The default M_{try} input parameter (Gislason et al., 2006) and the computation time needs for RF (Breiman, 2001) are given in equations 2.1 and 2.2 respectively. The RF algorithm is less sensitive to noise or overtraining, used to detect outliers, faster to train and more stable, and lighter than boosting. Moreover, it is widely used to estimate variables importance for the classification and regression (Belgiu & Drăguț, 2016; Dang et al., 2019; Genuer et al., 2010; Gislason et al., 2006; Pandit et al., 2018a).

Equation 2.1: The default M_{try} input variables

$$\sqrt{\text{Number of the input variables}} \quad (2.1)$$

The computation time needs for RF

$$T\sqrt{MN \log(N)}$$

Where the parameter T represents the number of trees, M denotes the number of variables that are used for each splitting or dividing nodes of the tree, and N represents the number of training samples.

2.3.1.2. Support vector machine

SVM is another non-parametric ML algorithm usually used for classification and regression, and it is developed by Cortes and Vapnik (1995) for binary classification problems. The binary classification using linear SVM in a two-dimensional input space is shown in Figure 8. For the linearly separable data, the goal of SVM is to find the optimal separating hyperplane for dividing the datasets into two classes with

maximizing the margin between the classes. The data points lying on the borderlines are known as support vectors (Figure 8) which define the optimal separating hyperplane and present at least one or more in each class (Mountrakis et al., 2011).

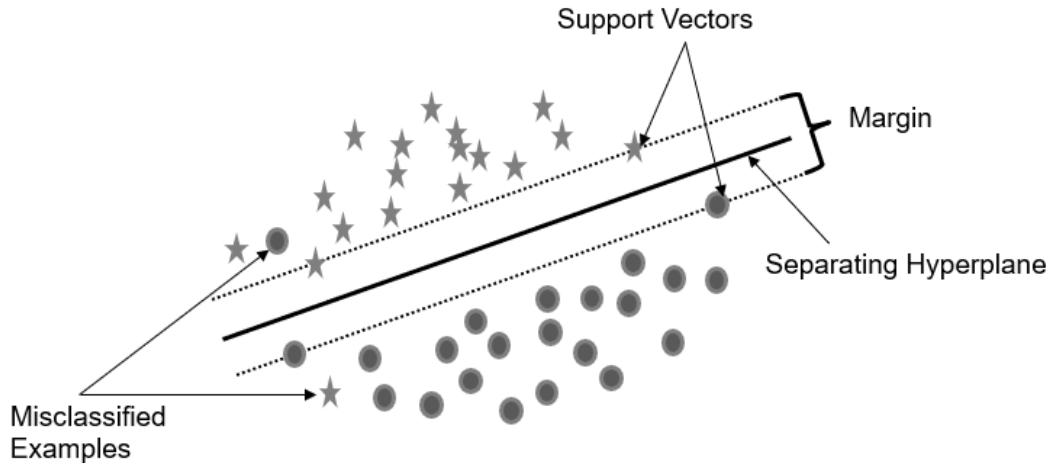


Figure 8: Linear separable classification example of SVM

In the real world, the relationship of the dataset is not only linear but also non-linear relationship exist where linear separability is problematic since the simple linear separation boundaries are often not enough for the classification with satisfactory accuracy. The techniques like soft margin method (which is created by the introduced slack variables) and kernel trick are used to solve the problem in the case of nonlinearly separable data (Bali et al., 2016; Cortes & Vapnik, 1995; Mountrakis et al., 2011). A cost parameter is denoted by C which indicates the cost for the misclassification of the dataset. By using the cost parameter, SVM tries to minimize the total cost instead of searching the maximum margin. Therefore, the C represents a trade-off between maximizing the margin and fitting the training data for minimizing the overfitting problem (Bali et al., 2016; Suresh et al., 2014). The kernel trick of the SVM is capable to map the non-linear separable data into some higher-dimensional feature space where the nonlinear relationship changes into the relatively linear relationship. For example, the scatterplot of weather class (sunny denotes with stars and snowy with circles) shows a non-linear relationship into a features space of latitude and longitude (Figure 9 (a)). In Figure 9 (b), after the application of the kernel function, the weather class between the sunny and snowy looks quite linear separable into the new dimensional features space of altitude and longitude (Bali et al., 2016).

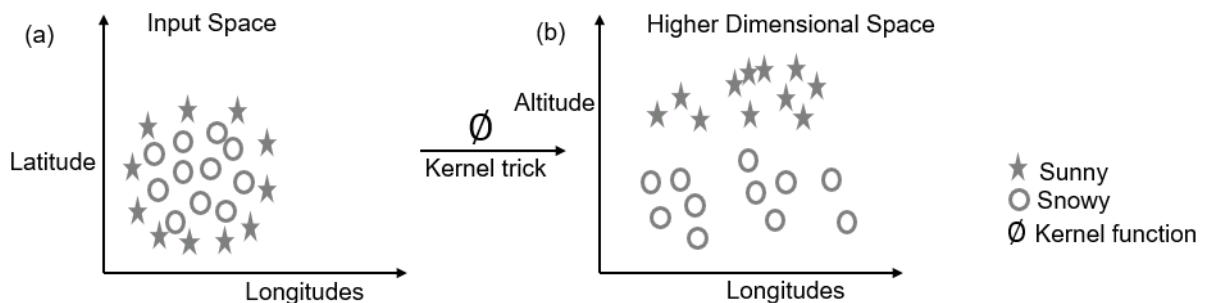


Figure 9: Application of the kernel trick in the SVM (a) non-linear relationship of weather class: sunny and snowy into a features space of latitude and longitude (b) linear relationship of weather class: sunny and snowy into a new dimensional features space of altitude and longitude.

In an SVM algorithm, four types of the kernel function namely linear, polynomial, radial basis function (RBF), and sigmoid kernel are commonly used in the practice (Equations 2.3 to 2.6) (Bali et al., 2016).

Equation 2.2: Linear kernel

In general, the linear kernel can be represented as the dot product of the features because it does not transform the data and it can be written as follow:

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j \quad (2.3)$$

Where K denotes a kernel function, and \vec{x}_i and \vec{x}_j represent the feature vectors.

Equation 2.3: Polynomial kernel

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d \quad (2.4)$$

Where K denotes a kernel function, \vec{x}_i and \vec{x}_j represent the feature vectors, d is a degree.

Equation 2.4: Sigmoid kernel

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\kappa \vec{x}_i \cdot \vec{x}_j - \delta) \quad (2.5)$$

Where K denotes a kernel function, \vec{x}_i and \vec{x}_j represent the feature vectors, and κ and δ denote the kernel parameter.

Equation 2.5: Gaussian RBF kernel

$$K(\vec{x}_i, \vec{x}_j) = e^{\frac{-\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}} \quad (2.6)$$

Where K denotes a kernel function, \vec{x}_i and \vec{x}_j represent the feature vectors, $\|\vec{x}_i - \vec{x}_j\|^2$ denotes as the square Euclidean distance between the \vec{x}_i and \vec{x}_j , and σ denotes a free parameter.

The choice of the kernel functions in the SVM algorithm relies on the learning concept, quantity of the training data, and the features space, and there is not any other specific kind of rules for selecting the appropriate kernels to any specific learning task (Bali et al., 2016). The SVM algorithm can be used in both cases either in classification or regression problem in many disciplines including forestry and agriculture using remote sensing data. It is not much affected by noisy data, less sensitive to overfitting and training data. However, the kernels and model parameters assignment is a challenging task to achieve a satisfactory outcome. Also, the algorithm can be slow to train in the case of a large number of dataset and the result interpretation task is not easy due to the use of complex black box model. The Grid-search technique is commonly used to select the appropriate model parameters (Bali et al., 2016; Feng et al., 2017; Lu et al., 2016; Mountrakis et al., 2011).

2.4. Review of related literature

2.4.1. Existing studies

Gleason & Im (2012) used LiDAR data with ML approaches to estimate forest biomass and compared the performance of four ML models including RF, Support vector regression (SVR), linear mixed-effects and

cubist regression technique at both tree and plot levels. The authors concluded that measurements of biomass at the individual tree level, all models executed worse results (RMSE: 68.1%-119.6%) in comparison to plot level (RMSE: 13.6%-34.2%). The result also showed that SVR model performed better result than other models in all the cases.

García-Gutiérrez et al. (2015) compared between the classic multiple linear regression (MLR) and ML regression approaches for LiDAR-derived measurement of forest variables. They found that SVM is statistically better than other ML techniques. Similarly, Wu et al. (2016) compared the ML algorithms (like SVM, RF, k-nearest neighbour, stepwise linear regression, and stochastic gradient boosting) using Landsat imagery combination with ground-based data for forest AGB estimation. They accepted that stepwise linear regression and RF produced a more stable performance for estimating forest AGB. Furthermore, the authors also determined that RF performed better than other ML algorithms based on RMSE (26.44 ton/ha) and R^2 (0.63).

Domingo et al. (2017) compared the regression models to measure biomass losses and carbon emission using low-density airborne LiDAR data. The study showed that low-density LiDAR data is capable to estimate pre-fire forest AGB accurately in a monospecific Aleppo pine forest. The authors summarised that MLR is a topmost model for forest AGB assessment in the pre-fire condition. They also stated that the MLR and SVM models look similar; the difference is insignificant.

Deb et al. (2017) used Resourcesat-2 data and ANN for measuring forest aboveground biomass. They also compared the study of ANN with other traditional linear and non-linear models to derive the well-suited model for estimating forest AGB in a dry deciduous forest of the tropical sub-humid or semiarid area. The authors summarised that ANN is the best model than others based on the numbers of statistical and consistency estimation measures. They also recommended that using a huge number of sample area with various sample sizes with respect to the related forest features like herbs, shrubs, and trees in addition to the LiDAR data for precise and accurately ANN modelling. Nandy et al. (2017) extracted spectral and texture variables from Resourcesat-1 Linear Imaging Self-scanning Sensor-III data and these extracted variables integrated with field-based measured data with the help of ANN to access forest biomass. The estimated forest biomass result attested admirable association between the extracted spectral and texture variables, and field measured biomass ($R^2 = 0.75$ and $RMSE = 85.32 \text{ Mg ha}^{-1}$). The authors also claimed that ANN has good ability to improve the quality of forest biomass assessment using the minimum number of suitable spectral and texture variables. They used only the ANN model of ML approaches and did not compare with the other models like RF, SVM and so on; it is one of the limitations of the study.

Space-borne LiDAR and high-resolution remote sensing data were combined in the study of Dhanda et al. (2017) for the improvement of precise measurement forest biomass and carbon stock. The authors also compared the performance of two ML algorithms i.e. SVM and RF. The result disclosed that 78.7 % adjusted variation in the estimated forest AGB ($RMSE = 13.9 \text{ Mg ha}^{-1}$) with the combination of six topmost essential parameters extracted from space-borne LiDAR and high-resolution optical sensor data. It also showed that 83 % variation in the estimated result of forest AGB with the combination of fifteen topmost essential parameters derived from the multi-sensor data. Furthermore, the authors found that RF and SVM ML models provide comparable performance if there is no underlying correlation of variables in the dataset but SVM model contributes better performance on increasing correlation. They also appealed that multi-sensor integration using ML approaches provide a better result than a single sensor approach for estimating the forest biomass.

2.4.2. Uncertainty analysis

The uncertainty analysis for forest AGB estimation specifies the major sources of the error that affect the effectiveness of the output accuracy (Lu et al., 2012b). Usually, uncertainties are concerned in a biased

decision of sampling, position uncertainties of the sample plots, uncertainties in the measurement of tree dimensions e.g., DBH and height, allometric equations, conversion factors, correction of optical image, variables selection, and used algorithms in forest AGB estimation of many tropical forests particularly in developing countries. The deforestation and forest degradation are the additional components of the uncertainty which makes a change in the forest carbon stock (Houghton, 2005; Lu et al., 2012b). Identifying the major sources of the uncertainties and their quantifications are mandatory for the reduction measure and decision making (Lu et al., 2012b).

ML algorithms (including RF and SVM) are progressively used in the field of remote sensing and geospatial data due to their ability of pattern recognition for high-dimensional features. In this context, pixel-based uncertainty analysis is required for the quantification of spatial variation. There are several methods such as U-statistics, Fourier amplitude sensitivity test, Quantile Regression, Taylor series, Jackknife-after-Bootstrap, Polynomial regression, Quantile Regression Forests and Monte Carlo simulations are used in uncertainty analysis for complex and non-linear models, and ML approaches (Dang et al., 2019; Lu et al., 2012b). Although, ML (non-parametric) approach does not provide the prediction error directly like traditional approaches such as multiple regression (Coulston et al., 2016). Monte Carlo simulation is a dominant approach to map uncertainty of forest AGB, forest carbon, and so forth, where more complex and non-linear models, and/or ML algorithms are used (Coulston et al., 2016; Dang et al., 2019; Kauranne et al., 2017). For examples, Dang (2012) used Monte Carlo simulation for error propagation in carbon estimation of Ludhikhola Watershed in Nepal with the coupling of airborne LiDAR and GeoEye data. Dang et al. (2019) used RF regression algorithm to predict the forest AGB of Yok Don National Park in Vietnam using Sentinel-2 data. Also, they found the uncertainty range between 9.87Mg ha⁻¹ to 93.27 Mg ha⁻¹ using used Monte Carlo simulation.

2.4.3. Forest biomass and carbon modelling in Nepal

The very limited forest AGB and carbon study has been conducted in Nepal with the application of remote sensing and geospatial data, particularly using ML approaches. Generally, some studies were carried out for forests tree AGB/carbon estimation using the relationship of crown projection area (CPA) and carbon with the help of the RS based data and field measured data. Most of the conducted studies are focused only in Chitwan and Gorkha districts (i.e., Central region) of Nepal. For instances, Baral (2011) combined WorldView-2 and GeoEye satellite imagery to map carbon stock with 61 % accuracy in subtropical forests of Chitwan district, Nepal. She established the relationship between the CPA and above ground carbon of the individual tree using a non-linear regression model. Also, the above ground carbon stock was measured with field data and the CPA was derived using object-based image analysis (OBIA) from the imagery. Again, Bautista (2012) compared the forest AGB/carbon estimation calculated from GeoEye and airborne LiDAR data of the same area. Further, Mbaabu et al. (2014) compared the carbon stock of two different forest management regimes using the same data and area. Likewise, Maharjan, (2012) used Digital Camera imagery and airborne LiDAR data for estimating and mapping above ground woody carbon stock in the hilly forests of Gorkha, Nepal. Similarly, Karna et al. (2015) integrated WorldView-2 and airborne LiDAR data for carbon estimation of six dominant tree species using in a subtropical forest of Chitwan district, Nepal. They used OBIA for the CPA extraction and supervised the nearest neighbour for the species level classification. Also, Shrestha, (2011) estimated carbon stock using GeoEye image and individual crown segmentation in Dolkha district of Nepal.

Kauranne et al. (2017) measured forest AGB and carbon stock using Bayesian model framework so-called LiDAR-Assisted Multi-source Programme (LAMP) to estimate the emission Reference Levels for TAL area in Nepal for UN REDD+ program. They used airborne LiDAR data of sample forest and wall to wall Landsat imagery with field sample plots. Additionally, forests were stratified using Landsat imagery and used calibrated LiDAR model with field measured plots to estimate mean forest AGB of every single stratified forest types (i.e. called LAMP2). Conversely, calibrated LiDAR model with field measured plots are initially used to estimate forest AGB on randomly generated LiDAR samples. Next, thousands of

surrogated plots on LiDAR blocks are generated and calculated forest AGB of each surrogate plots using the LiDAR-based model. Finally, the surrogate plots are used to calibrate the satellite-based AGB model and this model used to calculate forest AGB of each satellite pixels of the entire area which named as LAMP3. In a hector resolution, relative RMSE of forest AGB calculation was obtained between the range of 35 to 39 %.

Pandit et al. (2018a) measured landscape label of forest AGB in Buffer Zone Community Forests of Central Nepal using Landsat8 together with field data. They used MLR and RF for estimating AGB from the satellite imagery and field data and compared the result. They found that RF provided better result ($R^2=0.87$ and $RMSE = 20.50$ t/ha) than MLR ($R^2= 0.56$ and $RMSE = 37.01$ t/ha). Similarly, they (Pandit et al., 2018b) were also estimated AGB of the same area using Sentinel 2 with field measured data and obtained $R^2= 0.81$ and $RMSE = 250.57$ t/ha. However, they did not carry-out the uncertainty analysis in any study.

2.5. Knowledge Gaps

Precise estimation of forest biomass is a factor of datasets, sensors, complexity, prediction models, processing tools, atmospheric conditions, and algorithms. There is no single approach that can very efficiently estimate these components. Studies from the past have asserted that usage of multi-sensor products and their integration with field variables and other derived components using machine learning algorithms can certainly perform better than a single-sensor approach. However, very few literature are available which have used a multi-sensor approach for such estimations in the tropical forests of Nepal.

In this context, it is planned to integrate and optimize LiDAR information and optical information from RapidEye towards precise estimation of forest biomass around the tropical forests of Nepal. Further, accuracies of the estimates will be verified using field-collected measurements. Different AGB predictions models are formed with the help of multi-sensor parameters, therefore this is a unique integration of the multi-sensor approach using the ML algorithms. Moreover, it is also intended to perform a pixel-level uncertainty analysis for the estimated outputs, which was not common in the referred literature for biomass estimation.

The next chapter describes the selection of the study area and used datasets for this research.

3. STUDY AREA AND DATASET

This chapter deals with the selection of the study area, and the dataset with their description.

3.1. Study area

3.1.1. Selection criteria of the study area

The portion of the Terai Arc Landscape (TAL) area in Nepal is selected for this research. The research area is carefully chosen based on the following criteria:

- The area represents a major portion of the TAL area and biological corridors connectivity between Nepal and India.
- The forest area truly characterises the tropical forests of TAL area in Nepal.
- The area also represents the part of the National Forest Reference Level.
- The area covers with Terai hardwoods mixed forests with dominant Sal (*Shorea robusta*) species and it's associated with *Terminalia alata* (Asna), *Mallotus philippinensis* (Sindure), *Terminalia bellirica* (Barro) and so forth.
- Airborne LiDAR point clouds and high-resolution optical sensor image (i.e. RapidEye image) are also available for this area.

3.1.2. Description of the study area

This research is conducted in the tropical forest of TAL area in Nepal. TAL is a conservation approach of the landscape label to support and conserve the meta-population of significant mega-fauna (like a tiger, rhino, and elephant) through the biological corridors connectivity between Nepal and India. TAL area is extended from the Bagmati River of Nepal in the east to the Yamuna River of India in the west, covering the area of 51,002 km² with the connectivity of 15 protected areas. In Nepal, TAL is declared as a priority landscape program by the Government of Nepal in 2001. Currently, it is extended in 18 districts and covered the area of 24,710.13 km² including major river system (Narayani, Karnali, Mahakali and their tributaries). TAL inhabits over 7.5 million people of 45 different ethnic groups and its population growth is at the rate of 2.1 % per km² (i.e., high). Considerably, TAL-Nepal covers two World Heritage site (viz., Chitwan National Park and Lumbini) and three Ramsar wetland sites (viz., Bishazari, Ghodaghodi and Jagadishpur) which are significant for tourism purpose (MoFSC, 2015). Twelve major districts of the TAL area (which covers over 75 % forest of Terai) have allocated for the estimation of the National Forest Reference Level which is already submitted to UNFCCC in 2017 for the review. Additionally, Emission Reduction Program Document at sub-national level is on progress for the allocated area of the TAL (MoFE, 2018).

3.1.2.1. Overview of Kailali district

The study area is located in the Kailali district (Figure 10). This district is one of the major districts (out of 18) of the TAL area in Nepal. It is a portion of Far-western Province in Terai and one of the 77 districts of Nepal. The Kailali district lies between latitude 28° 22' N to 29° 05' N and longitude 80° 30' E to 81° 18' E. It covers an area of 3,235 km², and out of which lower tropical (below 300 m), upper tropical (300 to 1000 m), and subtropical (1000 to 2000 m) are covered with 59.3 %, 25.9 %, and 13.8 % respectively (District Development Committee (DDC), 2015).

The core study area (Figure 10) is a part of the lower tropical forest of the district and lies between latitude 28° 33' 44.19" N to 28° 36' 44.35" N and longitude 80° 52' 32.80" E to 80° 59' 04.29" E. It covers an area of 59.10 km² and out of which forest area is 42.16 square kilometres and non-forest area is 16.94 km².

❖ Land use/land cover

The district is rich in natural forest resources, covering with about 65 %. It has also fertile cultivated land (about 28 %). The area under different land use/land cover is shown in Table 1(DDC, 2015).

Table 1: Land use pattern (DDC, 2015)

SN	Description	Area covered	
		in hectare	in percentage
1	Cultivated land	89,935	27.80
2	Forest land	209,724	64.80
3	Pasture land	6268	1.90
4	River, uncultivated, build-up area	17,573	5.40
	Total	323,500	100.00

❖ Forest

The Kailali district has about 65 % of natural forest resources, therefore the district is rich in natural forest resources (including different wetlands and grasslands) and biodiversity (DDC, 2015). The major forests such as Sal (*Shorea robusta*) forest, Terai hardwood forest, mixed hardwood forest, and Riverain Khair-Sissoo (*Acacia catechu-Dalbergia sissoo*) forest are found in this district (Table 2). The available major forest resources managed under two main forest regimes (out of 4, Religious and Leasehold forest are less in size) namely government managed forest and community based managed forest. Further, community forest (CF) management and Collaborative forest management (CFM) systems have been practised under the community based managed forest system. Out of total natural forest resources, 69,820.20 hectares forest managed under CF (563 in number) and CFM (2 in number). There are 563 Community Forest User Groups (CFUGs) registered (in District Forest Office (DFO), Kailali) and involved in the CF management, 124,327 households and more than 0.75 million people are benefited directly or indirectly from the forest resources management system (DFO, 2018).

Table 2: The forest types and their coverage in the Kailali districts (DFO, 2018).

SN	Types of forests	Area	
		(in hectare)	(in percentage)
1	Sal (<i>Shorea robusta</i>) forests	70396.1	32.16
2	Terai hardwood (TH) forests (TH)	68708.6	31.39
3	Terai mix hardwood (TMH) forests	52588.3	24.02
4	Chirpine (<i>Pinus roxburghii</i>)	15967.2	7.29
5	Riverain Khair-Sissoo (<i>Acacia catechu-Dalbergia sissoo</i>) forests	10746.9	4.9
6	Shrubs or other wooded lands (OWL)	508.6	0.24
	Total	218915.7	100

❖ Social-economic and demographic information

The total population of the district is 775,709 (where male and female compositions are 48.78 % and 51.22 % respectively) with 142, 480 households. The population density is 240 per km² and the growth rate is 2.29 %. The people mainly depend on agriculture, livestock, and forest resources. The major rivers of the district are Karnali, Mohana, Khutiya, Patharaiya, Rora, Donda, Shiyganga, Gaurishankar, Kandra, Manahara, Godawari, Likma, and Glura, and major lakes are Ghodaghodi, Jokhar, Tilko, Behadababa, Koilahi, Godawari, Likma, Gulara. These rivers and lakes including natural forest resources support in people livelihood mainly ethnic group such as Tharu, Kami, Magar and other socially suppressed groups as well as

they contribute in the local and national economy. Tharu, Nepali and Doteli are the major spoken languages and the literacy rate is 66.32 % in this district (DDC, 2015).

❖ Topography and climate

The elevation of the district ranges from 109 m to 1950 m and the climate varies from tropical (more than 86 %) to subtropical (about 14 %). The average annual temperature is 24°C to 43°C during the autumn and 05°C to 19°C during the winter season. Also, the average rainfall is recorded 1840 mm and most rainfall occurs in the monsoon (June-September) (DDC, 2015).

❖ Geomorphology and soil texture

In this district, north region occupies with *Chure* (or *Churiya*) hills (foot of the *Chure* is called *Bhabar* region and together occupies with 44.80 %) and the remaining part of the district is totally plain land that lies in south part (DDC, 2015). The *Chure* is the comparatively youngest hill of the country that is stretched from the west region-Indus river of Pakistan to the east region-Brahmaputra river of India. It is composed of river's deposition products around 40 million years ago. The *Chure* is also called Shivalik and it lies between Terai (in the south) and Mahabharat area (in the north), occupies 12.78 % of the total land of Nepal. Different conservations measures have been applied to conserve *Chure* region by the President Chure-Terai Madesh Conservation Board (PCTMCDB) and their different stakeholders because it is the main water recharge zone for the Terai region (PCTMCDB, 2015). The description of the rocks and soil texture found in this district is given in Table 3.

Table 3: The description of rocks and soil texture found in the different region of the district (DDC, 2015).

SN	Region	Main characteristics	Rocks/soil texture
1	<i>Chure</i>	Young hill, fragile, and erosive	Stone, sandy rock, coarse sand, silt and sand-stone
2	<i>Bhabar</i>	Foot of the Chure hill, water recharge zone of the plain region	Gravel, sandy loamy, rough, skeletal, boulder, cobbles and pebbles
3	Terai	Productive natural forests, arable land	Fine sandy loamy soil

3.2. Datasets

High-resolution optical sensor image, airborne LiDAR data, field measured data are provided by FRTC, Nepal and Arbonaut Limited, Finland.

3.2.1. High-resolution optical sensor data

In this study, high-resolution optical sensor RapidEye image (Figure 10) acquired on 27th November 2010 is used which is provided by the Forest Research and Training Centre (FRTC), Nepal. The image has 5 spectral bands namely, blue (440-510 nm), green (520-590 nm), red (630-685 nm), red edge (690-730 nm) and near-infrared (760-850 nm) and it has 5-meter spatial resolution. The specification of RapidEye is shown in appendix 1.

3.2.2. Airborne LiDAR data

The airborne LiDAR data (Figure 11) is used for this study and it is acquired in 2011. This data is provided by FRTC, Nepal and Arbonaut Limited, Finland. The description of the airborne LiDAR data is shown in Table 4.

Table 4: LiDAR dataset information

Parameter	Descriptions
Platform	Helicopter (9N-AIW)
Laser scanner	Leica ALS50-II
Height of flight	2200 meter
Speed of flight	80 knots
Sensor pulse rate	52.9 kHz
Scan field of view half-angle	20°
Average point density	0.8 points/m ² (1.28 points/m ² for the study area)
Sensor scan speed	20.4 lines per second
Swath	1601.47 meter
Beam footprint	50 centimetre
Point spacing	Max 1.88 meter across, max 2.02 meter down
Datum	The World Geodetic System 1984 (WGS84)
Projection	Universal Transverse Mercator (UTM)
Lateral overlap (side overlap)	30 %
Average height and position accuracy	45 centimetre

3.2.3. Field measured data

The field data is collected during the airborne LiDAR campaign in March/April 2011 by the FRTC. Total 76 sample plots out of 79 plots (3 sample plots are not considered because they are laid on the river area) are used in this study. Out of 76, 55 sample plots are systematically designed and the rest of the sample plots are randomly placed over the study area. The shape of the plots is circular with a radius of 12.62 m (500 m²). Location, species, diameter (DBH has measured the species above 5cm), and height are the major plot-level information which is used for the plot-wise forest AGB calculation in this study.

3.2.4. The used Software

The list of software which is used to accomplish this research is shown in Table 5.

Table 5: The list of the used software and their purposes

SN	Name of the used software	Purposes
1	ArcMap 10.6.1	Projection, spatial analysis, map productions, and so forth
2	ERDAS Imagine 2018	Image processing and analysis
3	ENVI 5.5	Image processing and analysis
4	LAStools	Airborne LiDAR data processing and metrics generations
5	R – Statistical software	ML algorithms executions, AGB mapping and uncertainty analysis
6	Microsoft Excel	Statistical analysis
7	Microsoft Word	Thesis writing
8	Microsoft Power-point	Presentations
9	Mendeley Desktop	Citations and references

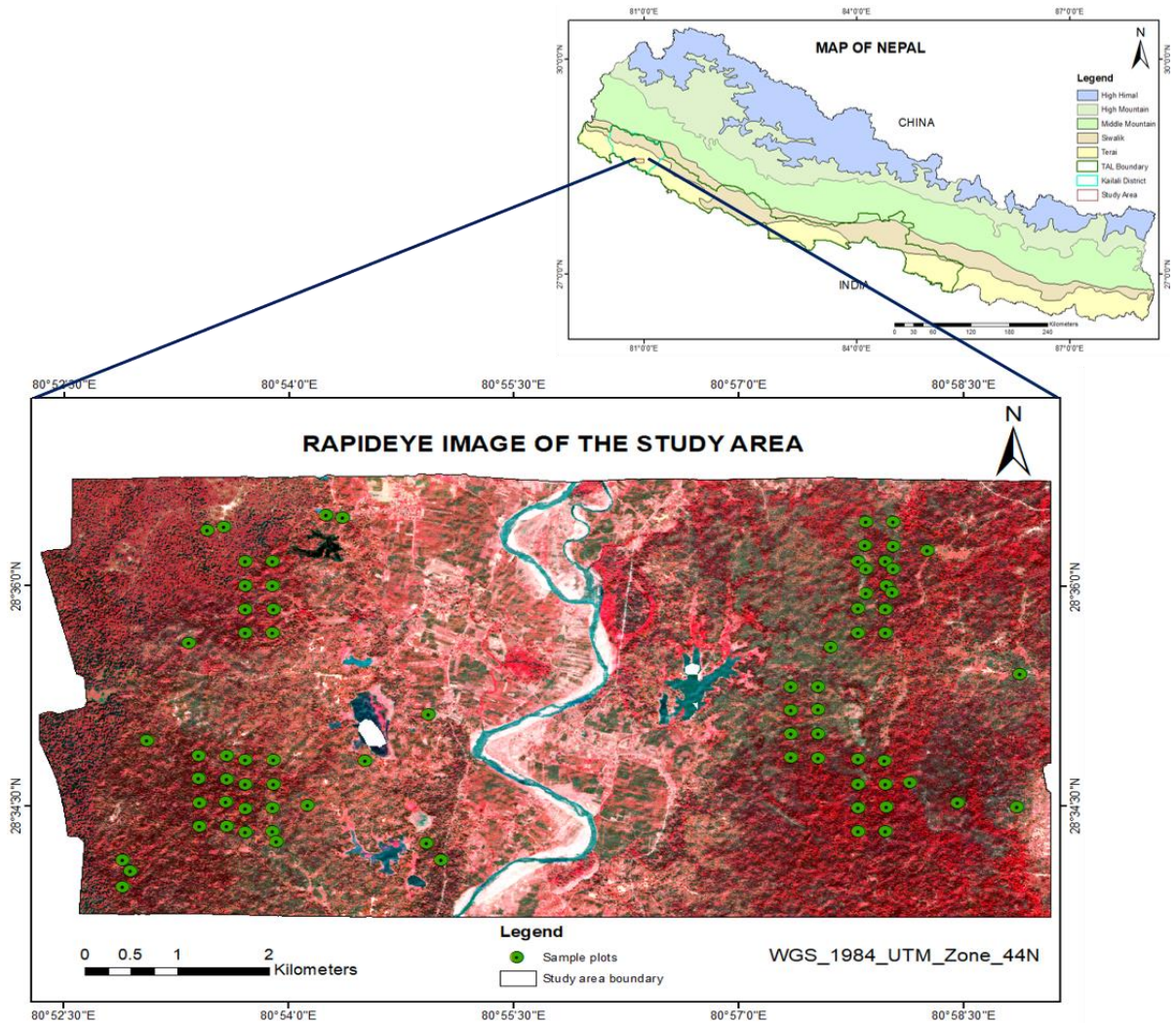


Figure 10: The study area located in the tropical forests of TAL area in Kailali district, Nepal.

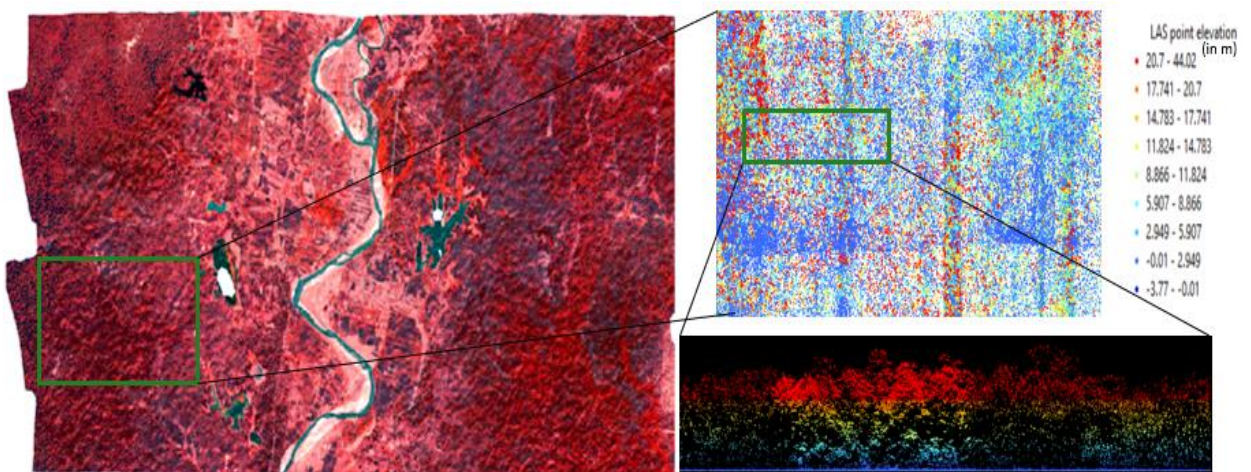


Figure 11: A portion of the study area showing general and profile view of the LiDAR point cloud data.

The next chapter explains the methodology to predict forest AGB of the study area using the above dataset.

4. METHODOLOGY

This chapter describes the method adopted to integrate and optimise the airborne LiDAR data, high-resolution data (i.e. RapidEye image), and field measured data using different ML algorithms, such as RF and SVM, to predict forest AGB. This also includes the extraction process of the spectral and textural variables and LiDAR metrics including pit-free CHM. Additionally, selection of the best AGB prediction model with ML algorithm, variables importance, the procedure of the spatial distribution of the forest AGB, its validation and uncertainty analysis also mention in this chapter. The methodological workflow diagram is given in Figure 12 and stepwise descriptions of the adopted method are mentioned in the following subsections.

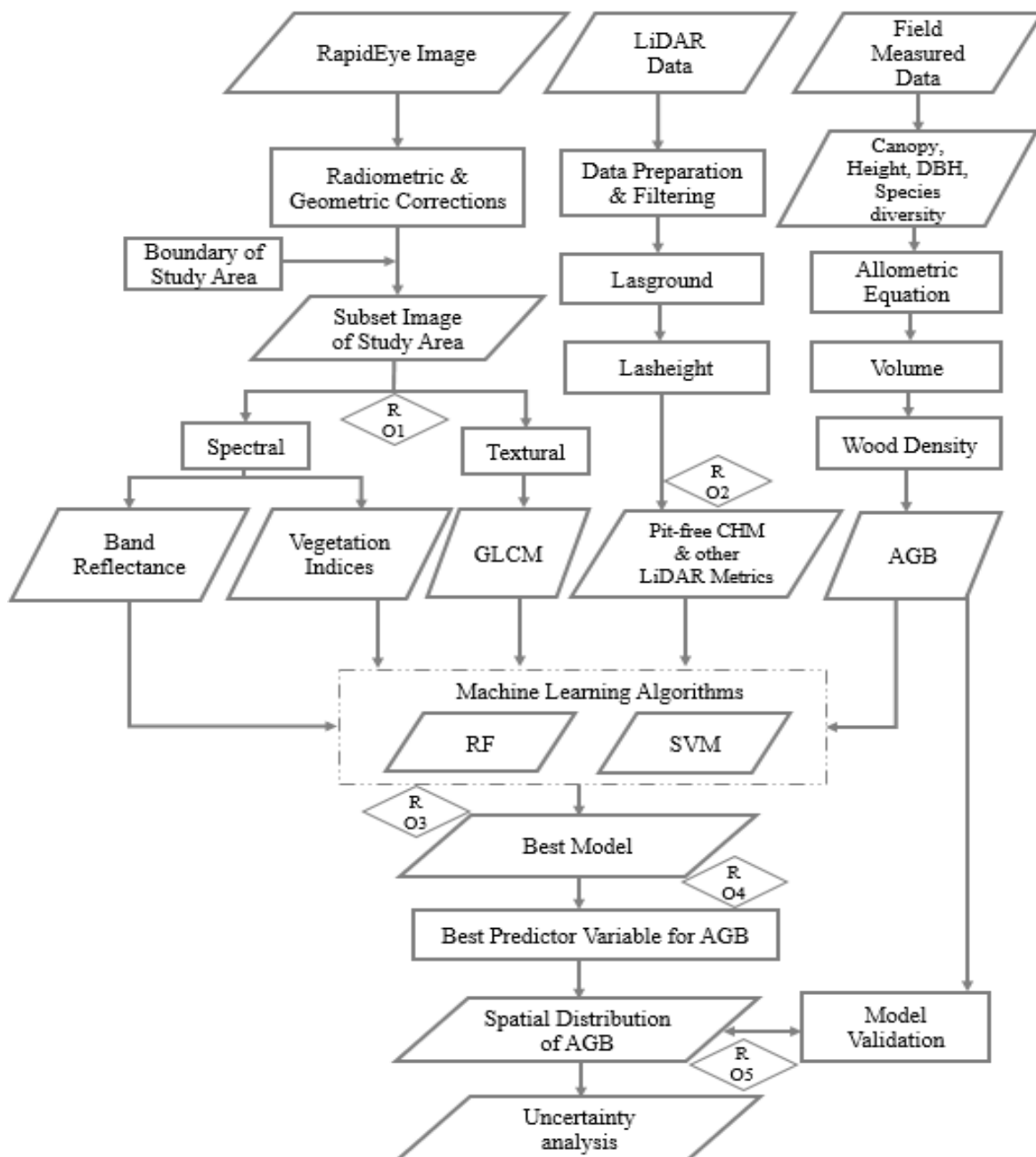


Figure 12: Methodological workflow diagram

4.1. Extraction of the variables from the RapidEye image

The acquired RapidEye image was already radiometrically as well as geometrically corrected and converted into reflectance image. Also, the image was co-registered with LiDAR data. The reflectance image was clipped to make the subset of the image of the study area for extracting spectral and textural variables (i.e., described in section 4.1.1. and 4.2.2). Further, the spectral variables were scaled down to fit with the size of the sampled plots. Also, the feature values of variable layers were extracted from the corresponding locations present in the in-situ measurements.

4.1.1. Extraction of the spectral variables

Band reflectance and vegetation indices were extracted for the spectral variables. The five bands (i.e., b1_25x25 = blue, b2_25x25 = green, b3_25x25 = red, b4_25x25 = red-edge and b5_25x25 = NIR) of the subset reflectance images were used to calculate different vegetation indices. The details of the various vegetation indices are given in Table 6.

Table 6: The description and equations of the used vegetation indices

SN	Index	Descriptions	Equation	References
1	NDVI	Normalized Difference Vegetation Index	$(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$	Rouse, Jr et al. (1974)
2	DVI	Difference Vegetation Index	$\text{NIR}-\text{Red}$	Tucker (1979)
3	GNDVI	Green Normalized Vegetation Index	$(\text{NIR}-\text{Green})/(\text{NIR}+\text{Green})$	Gitelson and Merzlyak (1998)
4	SR	Simple Ratio	NIR/Red	Jordan (1969)
5	NDVI _{Red-edge}	Red-Edge Normalized Difference Vegetation Index	$(\text{NIR}-\text{RE})/(\text{NIR}+\text{RE})$	Gitelson and Merzlyak (1994)
6	RDVI	Renormalized Difference Vegetation Index	$(\text{NIR}-\text{Red})/\text{SQRT}(\text{NIR}+\text{Red})$	Roujean and Breon (1995)
7	MSR705	Modified Red Edge Simple Ratio Index	$(\text{R750}-\text{R445})/(\text{R705}-\text{R445})$	Gitelson and Merzlyak (1994)
8	RVI	Ratio Vegetation Index	Red/NIR	Jordan (1969)
9	SARVI	Soil and Atmospherically Resistant Vegetation Index	$(\text{NIR}-\text{RB})(1+\text{L})/(\text{NIR}+\text{RB}+\text{L})$	Van Der Meer et al. (2000)
10	SAVI	Soil Adjusted Vegetation Index	$(\text{NIR}-\text{Red})(1+\text{L})/(\text{NIR}+\text{Red}+\text{L})$	Baret et al. (1989)
11	SQRT (NIR/Red)	Square Root of (Infrared/Red)	$\text{SQRT}(\text{NIR}/\text{Red})$	Tucker (1979)
12	TNDVI	Transformed Normalized Difference Vegetation Index	$(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red}+0.5)$	Baret et al. (1989)
13	MSAVI2	Modified Soil Adjusted Vegetation Index	$(2\text{NIR}+1-\text{SQRT}((2\text{NIR}+1)^2-8(\text{NIR}-\text{Red}))/2$	Baret et al. (1989)
14	MCARI2	Improved Modified Chlorophyll Absorption Ratio Index	$(1.5(2.5(\text{R800}-\text{R670})-1.3(\text{R800}-\text{R550}))/\text{SQRT}((2\text{R800}+1)^2-(6\text{R800}-5\text{SQRT}(\text{R670}))$	Haboudane (2004)

15	MTVI2	Improved Modified Triangular Vegetation Index	$(1.5(2.5(R800-R550)-2.5(R670-R550)))/\sqrt{(2R800+1)^2-(6R800-5\sqrt{R670})}$	Haboudane (2004)
16	MSR	Modified Simple Ratio-index	$((NIR/Red)-1)/\sqrt{(NIR/Red)+1}$	Chen (1996)
17	MSR _{Red-edge}	Red-edge Modified Simple Ratio-index	$((NIR/RE)-1)/\sqrt{(NIR/RE)+1}$	Wu et al. (2008)
18	CI _{green}	Green Chlorophyll Index	$((NIR/green)-1)$	Merzlyak et al. (2003)
19	CI _{RE}	Red-edge Chlorophyll Index	$((NIR/RE)-1)$	Gitelson (2005)
20	NDVI _{Red&RE}	Red and Red-edge NDVI	$(NIR-(Red+RE))/(NIR+(Red+RE))$	Xie et al. (2018)
21	MSR _{Red&RE}	Red and Red-edge MSR	$((NIR/(Red+RE))-1)/\sqrt{(NIR/(Red+RE))+1}$	Xie et al. (2018)
22	CI _{Red&RE}	Red and Red-edge Chlorophyll Index	$((NIR/(Red+RE))-1)$	Xie et al. (2018)

4.1.2. Extraction of the textural variables

GLCM of the second order statistical approach (i.e., presented in the third paragraph of Section 2.1.2.1.) was used to obtain texture variables. The different moving window sizes like 3x3, 5x5, 7x7, 9x9, 11x11 and 13x13 were tested to select the optimal window size for computing the texture measures. Next, the eight textural variables for each spectral band (i.e., B1 = blue, B2 = green, B3 = red, B4 = red-edge and B5 = NIR) were extracted using the optimal window (Table 7).

Table 7: The description and formula of the used texture variables (Haralick et al., 1973).

SN	Texture variables	Description	Formula
1	Mean	It is used to measure the gray level intensity of P_x in the image.	$\sum_{i=0}^L i P_x(i)$
2	Variance(Varrian)	It is used to quantify the image heterogeneity.	$\sum_{i,j=0}^{L-1} (i - \mu)^2 P(i, j)$
3	Homogeneity (Homogen)	It is used to quantify the image homogeneity (Inverse difference moment).	$\sum_{i,j=0}^{L-1} \frac{1}{1 + (i - j)^2} P(i, j)$
4	Contrast (Contras)	It is used to quantify the image of local variations.	$\sum_{n=0}^{L-1} n^2 \left\{ \sum_{i,j=1}^L P(i, j) \right\}, i - j = n$
5	Dissimilarity (Dissimi)	It is used to quantify the mean of the image's gray level difference.	$\sum_{i,j=0}^{L-1} i - j P(i, j)$
6	Entropy	It is used to quantify the image's disorder or complexity.	$-\sum_{i,j=0}^{L-1} P(i, j) * \log\{P(i, j)\}$
7	Second angular moment (SecondM)	It is used to quantify the image's textural uniformity.	$\sum_{i,j=0}^{G-1} \{P(i, j)\}^2$

8	Correlation (Correla)	It is used to quantify the gray levels linear dependency of the neighbouring pixels of the image.	$\sum_{i,j=0}^{L-1} \frac{\{i * j\} * P(i,j) - \{\mu_x * \mu_y\}}{\sigma_x * \sigma_y}$
Key: L = Number of gray levels, P(i,j) = Normalized co-occurrence matrix, P _x (i) and P _y (j) = Marginal row and column probabilities, μ = the mean value of P, μ_x, μ_y, σ_x and σ_y = means and standard deviations of P _x and P _y .			

4.2. Extraction of the CHM and other airborne LiDAR metrics

4.2.1. Airborne LiDAR data preparation

The airborne LiDAR data (the provided data was already pre-processed) of the study area was checked for the quality using lasinfo, lasvalidate, lasview, and lasoverlap modules of LAStools. The generated lasinfo report for the basic information and quality checking of airborne LiDAR point cloud data is shown in Appendix 3. After the confirmation of the quality of the LiDAR data, all point clouds data were tiled at the size of 1000x1000 tiles with 25 m buffer and it was also projected (WGS 84 UTM 44 N). The tiles and buffers are used to maintain the number of points per tile, use the memory resourcefully, and compute parallel tile processing (Isenburg, 2016). The produced tiles were denoised to remove outliers using the lasnoise tool and the data were categorised into the non-ground and ground returns with help of the lasground tool. The buffer was removed for the required derivatives production. The command line scripts used for the airborne LiDAR data quality checking and data preparation for the derivatives production are shown in Appendix 2.

4.2.2. CHM and other airborne LiDAR metrics extraction

Generally, the airborne LiDAR data was filtered into the non-ground points and ground points for further analysis to derive different height models and metrics. First returns (also called “non-ground points”) and last returns (also called “ground points”) were interpolated to generate a digital surface model (DSM) and digital terrain model (DTM) respectively using a specified size of the regular grid based on a Delaunay triangulation. The constructed Triangular Irregular Networks (TINs) using the LiDAR returns were rasterized for the DTM/DSM generation. DTM denotes only bare earth surface representing the height information natural topography features where the height information of other existing objects like trees and buildings are eliminated. In contrast to the DTM, DSM represents the height information of both earth surface and comprising existing things (such as trees and buildings) on it. Finally, the normal CHM was extracted after deducting DTM from DSM using normalized height points (Omar et al., 2015). Though, the only first returns interpolations are used to extract the normal CHM extraction, which has two major disadvantages. Firstly, some height information is lost especially in the case of off-nadir scan angles and the remaining LiDAR returns are not renumbered as the first returns, after the removing higher outliers (like, noise created due to cloud, dust, flying birds etc.) so they are not used in the normal CHM (Figure 13a). The second disadvantage is with regards to spike-TIN due to the construction of nettle-like triangles while interpolating all the first returns (i.e., both higher and lower first returns generated from the on-nadir and off-nadir) (Figure 13a). Irregular height variations (called ‘pits’) are seeded in the normal CHM because of the spike-TIN formation (Figure 13b) (Isenburg, 2016; Khosravipour et al., 2013). Khosravipour et al. (2013) and Isenburg (2016) created pit-free CHM by generating spike-free TIN using all relevant LiDAR returns with the help of the Delaunay algorithms (Figure 13c).

After the filtering, for sorting into the ground and non-ground points, the non-ground height points were normalized by replacing the Z-elevation and dropped the points above 45 m (since the maximum height of the measured tree was found below than 45 m) in the field measured data using the lasheight command of LAStools. Normal CHM was generated using all height points, partial CHMs were also created by calculating TIN with help of Delaunay triangulation using the LiDAR returns at different height interval from 2 m and above to 40m and above (i.e., $\geq 2m, \geq 5m, \geq 10m, \geq 15m, \geq 20m, \geq 25m, \geq 30m, \geq 35m$, and

$\geq 40\text{m}$) height points. Finally, all partial CHMs were merged together into a single CHM, by taking highest height point of the all involved CHM raster for all the location to derive pit-free CHM (Fig. 14) (Khosravipour et al., 2013). For extraction of all CHM, the blast2dem command of LAStool was used and grid size (or step) kept at 0.5 m (resolution) to reduce the uncertainties of tree height.

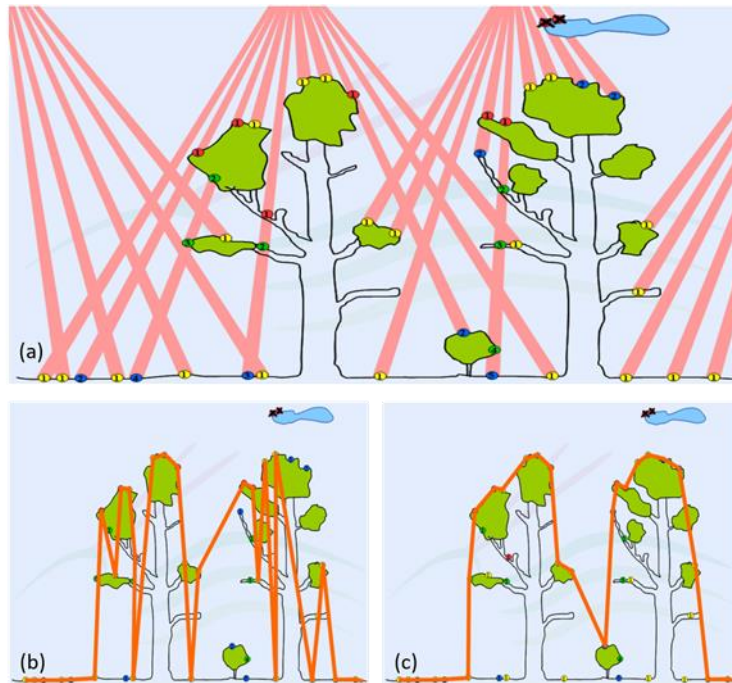


Figure 13: (a) Airborne LiDAR returns from the four flight-lines on the trees, (b) All the first returns of Airborne LiDAR used for the interpolation, and (c) All the relevant returns of Airborne LiDAR used for the interpolation (Isenburg, 2016).

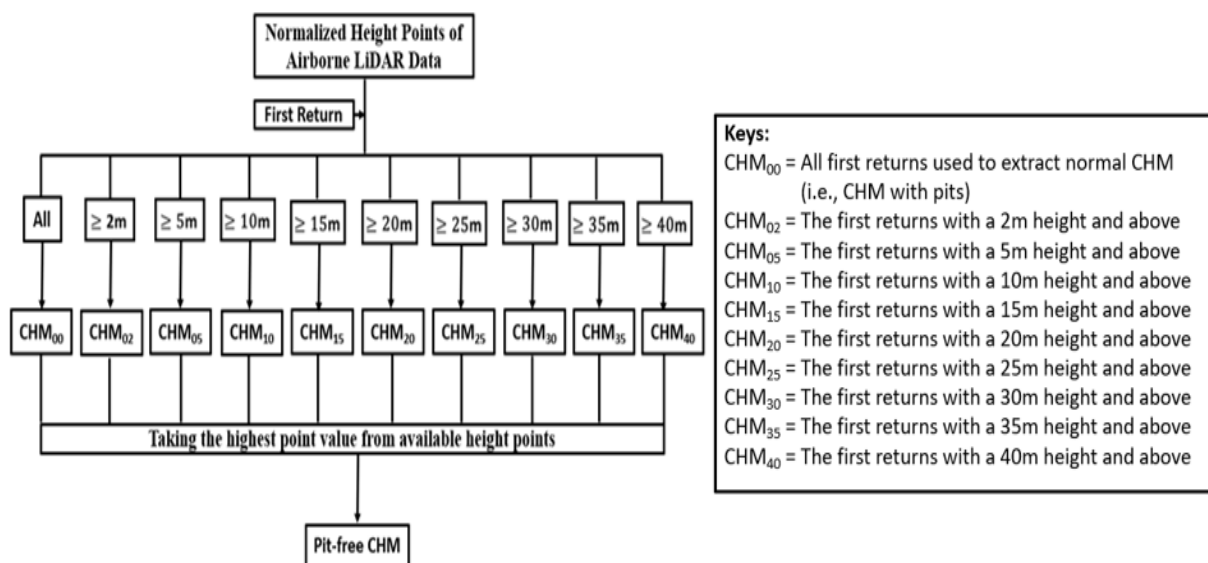


Figure 14: Pit-free CHM model workflow

The airborne LiDAR pulse returns comprise mainly the height of the object and strength of the return (or intensity) after being reflected by an object (Omar et al., 2015). Based on these two major information of the LiDAR data, 52 LiDAR metrics such as different percentiles heights, canopy cover, canopy density, bincentiles, height metrics (minimum, maximum, average, standard deviation, skewness, kurtosis, quadratic average height), different point height, and density strata were generated. The descriptions of all extracted LiDAR metrics are depicted in Table 8. The command line scripts used for generating all LiDAR metrics using different modules of the LAStools are depicted in Appendix 2. For further processing, the metrics features were extracted from the corresponding locations obtained from the field measured data.

Table 8: Descriptions of the airborne LiDAR metrics

SN	Name of the metrics	Descriptions of the metrics	References
1	BF_CHM_Pit	The pit-free CHM (i.e., generated using all relevant returns).	Khosravipour et al. (2013)
2	Cannopy_Co	The plot-level canopy cover (i.e., the percentage of the 1 st returns over a specified height (8 m)).	Barnes et al. (2017), Dhanda et al.
3	Cannopy_De	The plot-level canopy density (i.e., the percentage of the total returns over a specified height (8 m)).	(2017), and Li et al. (2017)
4	Ht_Average	The normalized average height of the pulse returns within the plot above cut-off height.	Li et al. (2017) and Nevalainen et al. (2017)
5	Ht_B10	The plot-level bincentile located below 10 % of the maximum height (i.e., the percentage of the pulse returns between the DBH or cut-off height and the specified percentage of the maximum height).	Li et al. (2017) and Nevalainen et al. (2017)
6	Ht_B20	The plot-level bincentile located below 20 % of the maximum height.	
7	Ht_B30	The plot-level bincentile located below 30 % of the maximum height.	
8	Ht_B40	The plot-level bincentile located below 40 % of the maximum height.	
9	Ht_B50	The plot-level bincentile located below 50 % of the maximum height.	
10	Ht_B60	The plot-level bincentile located below 60 % of the maximum height.	
11	Ht_B70	The plot-level bincentile located below 70 % of the maximum height.	
12	Ht_B80	The plot-level bincentile located below 80 % of the maximum height.	
13	Ht_B90	The plot-level bincentile located below 90 % of the maximum height.	
14	Ht_Kurtosi	The kurtosis of the pulse returns distribution within the plot-level above the cut-off height. It measures the tailedness of all point-height distribution within the plot.	

15	Ht_Maximum	The normalized maximum height of the pulse returns within the plot above cut-off height.	Li et al. (2017) and Nevalainen et al. (2017)
16	Ht_Minimum	The normalized minimum height of the airborne LiDAR returns within the plot above cut-off height.	
17	Ht_P01	The normalized height of the 1 st percentile of the pulse returns above cut-off height within the plot. It measures the height from which a specified % (here, 1 %) of all points located below.	Li et al. (2017) and Nevalainen et al. (2017)
18	Ht_P05	The normalized height of the 5 th percentile of the pulse returns above cut-off height within the plot.	
19	Ht_P10	The normalized height of the 10 th percentile of the pulse returns above cut-off height within the plot.	
20	Ht_P25	The normalized height of the 25 th percentile of the pulse returns above cut-off height within the plot.	
21	Ht_P50	The normalized height of the 50 th percentile of the pulse returns above cut-off height within the plot.	
22	Ht_P75	The normalized height of the 75 th percentile of the pulse returns above cut-off height within the plot.	
23	Ht_P90	The normalized height of the 90 th percentile of the pulse returns above cut-off height within the plot.	
24	Ht_P95	The normalized height of the 95 th percentile of the pulse returns above cut-off height within the plot.	
25	Ht_P99	The normalized height of the 99 th percentile of the pulse returns above cut-off height within the plot.	
26	Ht_Quadrat	The normalized quadratic average height of all pulse returns above cut-off height within the plot.	
27	Ht_Skwenes	The skewness of the pulse returns distribution within the plot-level above the cut-off height. It measures the asymmetry of all point-height distribution within the plot.	
28	Ht_Standar	The normalized standard deviation height of all pulse returns above cut-off height within the plot.	
29	Int_Averag	The intensity-based average height of the pulse returns within the plot above cut-off height.	
30	Int_Kurtos	The intensity-based kurtosis of the pulse returns distribution within the plot-level above the cut-off height.	
31	Int_Maximu	The intensity-based maximum height of the pulse returns within the plot above cut-off height.	
32	Int_P05	The intensity-based height of the 5 th percentile of the pulse returns above cut-off height within the plot.	
33	Int_P10	The intensity-based height of the 10 th percentile of the pulse returns above cut-off height within the plot.	
34	Int_P25	The intensity-based height of the 25 th percentile of the pulse returns above cut-off height within the plot.	
35	Int_P50	The intensity-based height of the 50 th percentile of the pulse returns above cut-off height within the plot.	

36	Int_P75	The intensity-based height of the 75 th percentile of the pulse returns above cut-off height within the plot.	
37	Int_P90	The intensity-based height of the 90 th percentile of the pulse returns above cut-off height within the plot.	
38	Int_P95	The intensity-based height of the 95 th percentile of the pulse returns above cut-off height within the plot.	
39	Int_P99	The intensity-based height of the 99 th percentile of the pulse returns above cut-off height within the plot.	
40	Int_Quadra	The intensity-based quadratic average height of all pulse returns above cut-off height within the plot.	
41	Int_Skwene	The intensity-based skewness of the pulse returns distribution within the plot-level above the cut-off height.	
42	Int_Standa	The intensity-based standard deviation height of all pulse returns above cut-off height within the plot.	
43	IntC_15	The intensity-based plot-level point counts between 0 to 15 m.	
44	IntC_15_30	The intensity-based plot-level point counts between 15 to 30 m.	
45	IntC_30_45	The intensity-based plot-level point counts between 30 to 45 m.	
46	IntD_15	The intensity-based plot-level percentage of points between 0 to 15 m.	
47	IntD_15_30	The intensity-based plot-level percentage of points between 15 to 30 m.	
48	IntD_30_45	The intensity-based plot-level percentage of points between 30 to 45 m.	
49	RHtC_15	The plot-level relative height point counts between 0 to 15 m.	
50	RHtC_15_30	The plot-level relative height point counts between 15 to 30 m.	
51	RHtD_15	The plot-level relative height point density between 0 to 15 m.	
52	RHtD_15_30	The plot-level relative height point density between 15 to 30 m.	
Keys: The cut-off height is breast height to 1.3 m which is used during the extraction of the LiDAR metrics.			

4.3. Forest AGB estimation of the measured sample plots

The description of the sample plots (total of 76 sample plots) is given in section 3.2.3. Initially, the 76 sample plots were divided into 53 and 23 for the modelling of the ML and validation purpose respectively. Out of 23 plots, the 8 edge influenced (due to the presence of river, lakes and arable land in the study area) randomly located sample plots were excluded from the validation measured plots to reduce the biases. DBH (measured for each tree which is ≥ 5 cm DBH), height of the tree, associated volume equation developed by Sharma and Pukkala (1990) are the main components that used to estimate the plot-wise forest volume.

Subsequently, the calculated species volume was multiplied by the related wood density to compute the tree-level AGB (Table 9) and all the tree level volume was added together to estimate plot-level forest AGB (in Mg ha⁻¹). The used wood density of the species is developed by the ministry of forest and soil conservation (MoFSC) and mentioned in the master plan for forestry sector (MPFS) in Nepal (MoFSC, 1989). The list of the species, their model coefficient, and wood density are depicted in the Appendices 4 and 5.

Table 9: The used equations for the calculation of the tree level AGB.

Equations	Sources
Stem volume (over bark) in m ³ = $\exp(a+b*\log(\text{DBH})+c*\log(h))$ Where DBH is a diameter measured at breast height (in m) h is a tree height (in m) a, b and c are model coefficients (i.e., depending on the species)	Sharma and Pukkala (1990)
Tree stem biomass = SV*WD Where SV is a stem volume (in m ³), and WD is a wood density (in kg/m ³)	Sharma and Pukkala (1990) and MoFSC (1989)
Total AGB of the individual tree = SB + BB + FB Where SB is stem biomass, BB is branch biomass, and FB is foliage biomass. The BB to SB and FB to SB ratios given by the MPFS which is used to compute the BB and FB (excluding dead trees).	MoFSC (1989)
BB-to-SB and FB-to-SB ratio Species BB-to-SB FB-to-SB Small Medium Large Small Medium Large Sal (<i>Shorea robusta</i>) 0.055 0.341 0.357 0.062 0.067 0.067 Other Terai mixed 0.4 0.4 0.4 0.07 0.05 0.04 hardwood species Where small, medium and large represent DBH<28 cm, DBH (28-53 cm), and DBH>53 cm	MoFSC (1989)

4.4. Machine learning algorithms and its accuracy assessment

RF and SVM ML algorithms were used to make a link between the dependent variable (i.e., field measured forest AGB) and independent variable (i.e., spectral, textural and LiDAR parameters) for the integration and optimisation of the independent variables with respect to the dependent variables. The spectral, textural and LiDAR variables were permuted into the seven different forest AGB prediction models: (i) LiDAR metrics/variables, (ii) spectral variables, (iii) textural variables, (iv) spectral and textural combined variables, (v) LiDAR and spectral combined variables, (vi) LiDAR and textural combined variables, and (vii) LiDAR, spectral, and textural combined variables in order to test and select the best using both ML algorithms. Finally, the best model (out of seven) and the ML algorithm (between RF and SVM) were selected based on the accuracy. The selected ML with the model was further executed to optimise the independent variables, and select the optimal predictor variables for the prediction of forest AGB and its spatial distribution pattern in the study area (Dhanda et al., 2017; Nandy et al., 2017; Pandit et al., 2018a).

4.4.1. RF regression for the forest AGB estimation

RF (the details is given in section 2.2.1.1.) is an ensemble ML technique that uses a bagging or bootstrap aggregation approach (Breiman, 2001), it is used for the classification, regression, and variable selection by making a number of decision trees (N_{tree}) using a set of samples from the training data (Belgiu & Drăguț, 2016). This ML algorithm was used to optimise the independent variables by executing the seven AGB predictions models (i.e., described in section 4.4) in the study. The RF regression was accomplished in the R statistical software (which is freely available) with the *randomForest* package. The modelling data (here, 53

sample plots) was divided into the in-bag sample data keeping the two-thirds with randomly selected, which was used as a training for making maximum N_{tree} without pruning using bagging approach, and the remaining one-third sample data was used for the out-of-bag (OOB). The OOB was executed for the internal testing (or cross-validation) of the models to determine the prediction error (i.e., OOB error) (Belgiu & Drăguț, 2016). Also, the available number of variables (called M_{try}) were randomly selected for each splitting nodes of the N_{tree} . The square root of the number of input variables (Gislason et al., 2006) is a default value for M_{try} , input. The N_{tree} and M_{try} are the major components for implementing the RF regression in R (Dang et al., 2019). It is mandatory to select the optimal value of N_{tree} and M_{try} to reduce OOB error (or RMSE). To get best M_{try} and N_{tree} , the *tuneRF* function was repeated until convergence for the M_{try} and the range from 100 to 2000 with 100 intervals was tested for the N_{tree} (Dang et al., 2019; Pham & Brabyn, 2017).

4.4.2. SVM regression for the forest AGB estimation

SVM (the details is given in section 2.2.1.2.) is another ML algorithm, applied in this research, that widely used for classification and regression purpose in forestry application. This algorithm provides comparatively higher accuracy using few training samples (Mountrakis et al., 2011). It generates hyperplane that splits the feature space to make fairly homogeneous categorization on either side. The constructed hyperplane is separated the feature space with maximum margin using the support vectors (i.e., adjacent vectors with respect to the hyperplane) in the case of the linearly separable data. The main features of the SVM is a kernel function, which is used to project the non-linear data into the higher dimensional space. This process assists to change the non-linear relationship of the data into linear and create optimal separating hyperplane (Bali et al., 2016; Suresh et al., 2014). There are four types of the kernel function namely linear, polynomial, RBF, and sigmoid kernel used in this algorithm, depending on the learning concept, quantity of the training data, and the relationship between the features.

All four kernel functions are used to optimise the independent variables using seven different AGB prediction models of LiDAR, spectral, and textural variables because there is no any consistent regulation for matching a kernel for any specific learning task (Bali et al., 2016). Therefore, training and testing of all kernels are performed to optimise the independent variables using different options of parameters (such as cost, SVM-type: eps-regression, and nu-regression) for customizations. The cost parameter denotes trade-off between the training error and testing error. Generally, the high value of the cost parameter shows lower misclassification while forming the decisions boundaries and vice-versa (Suresh et al., 2014).

This ML algorithm was also executed in the R statistical software using the *e1071* package with the SVM function. The sample data (i.e., 53 sample plots) was randomly divided into a training set (i.e., 37 sample plots) and testing set (i.e., 16 sample plots). To achieve an optimal result, the model was tuned using a different set of parameters (mainly, cost and SVM-type) until convergence.

4.4.3. Accuracy assessment

All the seven AGB prediction models of the used airborne LiDAR, spectral, and textural variables were assessed based on the root mean square error (RMSE, it provides the measure of accuracy), coefficient of determination (R^2), relative RMSE ($RMSE_{rel}$), coefficient of variance of the RMSE ($RMSE_{CV}$) to choose the best forest AGB prediction model along with the used ML regression algorithms (Equations 4.1-4.4) (Dang et al., 2019; Feng et al., 2017; Pandit et al., 2018b; Pham & Brabyn, 2017). Next, the final model and regression ML algorithm (best kernel was also decided for the SVM) were selected for further processing such as the determination of the variables importance and an optimal number of variables. Additionally, the root mean squared deviation (RMSD), that indicates the mean deviation of predicted values in relation to field measured values (Piñeiro et al., 2008), relative RMSD ($RMSD_{rel}$), coefficient of variance of the RMSD ($RMSD_{CV}$) were used to evaluate the tree heights derived from Normal CHM and pit-free CHM as compared to the field measured tree height (Equations 4.5-4.7).

Equation 4.1: RMSE

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (4.1)$$

Where RMSE denotes the root mean squared error (in Mg ha⁻¹), N denotes the number of the plots, \hat{y}_i denotes the predicted forest AGB (in Mg ha⁻¹), y_i denotes the corresponding field-measured forest AGB (in Mg ha⁻¹).

Equation 4.2: RMSE_{rel}

$$\text{RMSE}_{\text{rel}} = \left(\frac{\text{RMSE}}{\bar{Y}} \right) * 100 \quad (4.2)$$

Where RMSE_{rel} denotes the relative root mean squared error, RMSE is the root mean square error (in Mg ha⁻¹), \bar{Y} denotes the mean forest AGB calculated from the field measured forest AGB (in Mg ha⁻¹).

Equation 4.3: RMSE_{CV}

$$\text{RMSE}_{\text{CV}} = \left(\frac{\text{RMSE}}{\bar{Y}} \right) \quad (4.3)$$

Where RMSE_{CV} represents the coefficient of variance of the relative root mean squared error, RMSE denotes the root mean square error (in Mg ha⁻¹), and \bar{Y} denotes the mean forest AGB calculated from the field measured forest AGB (in Mg ha⁻¹).

Equation 4.4: R²

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{Y})^2} \quad (4.4)$$

Where N is the number of the plots, \hat{y}_i denotes the predicted forest AGB (in Mg ha⁻¹), y_i denotes the corresponding field-measured forest AGB (in Mg ha⁻¹), and \bar{Y} represents the mean forest AGB calculated from the field measured forest AGB (in Mg ha⁻¹).

Equation 4.5: RMSD

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N - 1}} \quad (4.5)$$

Where RMSD denotes the root mean squared deviation (in m), N denotes the number of the trees, \hat{y}_i denotes the derived tree heights (in m), y_i denotes the corresponding field-measured forest tree heights (in m).

Equation 4.6: RMSD_{rel}

$$\text{RMSD}_{\text{rel}} = \left(\frac{\text{RMSD}}{\bar{Y}} \right) * 100 \quad (4.6)$$

Where RMSD_{rel} denotes the relative RMSD, RMSD is the root mean squared deviation (in m), \bar{Y} denotes the mean forest tree height calculated from the field measured forest tree heights (in m).

Equation 4.7: RMSD_{CV}

$$\text{RMSD}_{\text{CV}} = \left(\frac{\text{RMSD}}{\bar{Y}} \right) \quad (4.7)$$

Where RMSD_{CV} represents the coefficient of variance of the RMSD, RMSD denotes the root mean squared deviation (in m), and \bar{Y} represents the mean forest tree height calculated from the field measured forest tree heights (in m).

4.4.4. Optimal predictors variables of the best model using ML algorithm

The selected ML algorithm was used to optimise the variables (i.e., variable importance) of best AGB prediction model and get the ideal number that was used to predict the forest AGB over the study area. The best performer ML algorithm was executed repetitively until convergence to get the optimal accuracy (Dang et al., 2019).

4.4.5. Prediction of the forest AGB and its spatial distribution

The optimal number of the selected variables were used to predict the forest AGB over the forest area using the effective ML algorithm. Also, the predictor variables containing model was executed to generate a spatial distribution of the forest AGB over the forest area (Dang et al., 2019; Deb et al., 2017; Dhanda et al., 2017; Nandy et al., 2017).

4.5. Model validation

In addition to the cross-validation during the implementation of the model, the validation measures- R^2 , RMSE, RMSE_{rel} , and RMSE_{CV} (Equations 4.1-4.4) were calculated using field-measured forest AGB of the validation plots (15 sample plots which were not used as a training and testing) to validate the model (Dang et al., 2019; Pandit et al., 2018a).

4.6. Uncertainty analysis

The uncertainty analysis of the estimated spatial distribution of the forest AGB suggests the various sources of the error that affect the accuracy and precision of forest AGB measurement (Lu et al., 2012b). There are different methods (with their merits and demerits) for uncertainty analysis, Monte Carlo simulation can provide pixel-level uncertainty for the prediction model which is very important in modelling applications such as biomass modelling (Coulston et al., 2016; Dang et al., 2019; Lu et al., 2012b).

In this study, the pixel-level uncertainty of RF-based predictions model was calculated using Monte Carlo simulation (Coulston et al., 2016). Figure 15 shows the basic process for RF prediction uncertainty, the \mathbf{p} and \mathbf{Q} represent the response and predictor variables where bootstrap (Figure 7) sample (\mathbf{p}, \mathbf{Q}) is $(\mathbf{p}^b, \mathbf{Q}^b)$, that denotes b^{th} bootstrap sample. The approximate prediction uncertainty for RF regression model was further described in the following steps (Coulston et al., 2016):

- Create 5000 N_{tree} bootstrap dataset and build RF¹, RF², RF³, ..., RF⁵⁰⁰⁰ models.
- Compute the prediction error to each replicate on the basis of related RF^b replicate of each observation in \mathbf{p}^b (i.e., the part of original data which does not participate in b^{th} bootstrap sample).
- Generate error assessment for each dataset or observation (\mathbf{p}^b) on the basis of predicted values distribution.

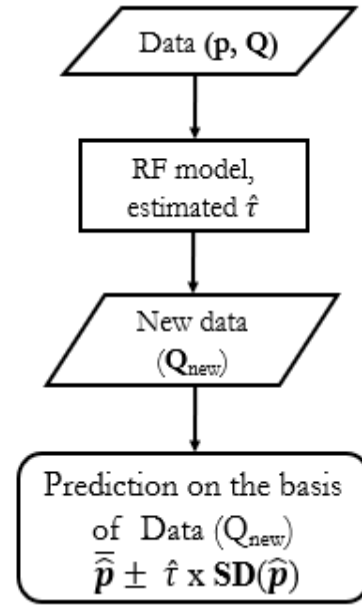


Figure 15: A process for RF prediction uncertainty.

For example, n^b records on the basis of test set samples $(\mathbf{p}^b, \mathbf{Q}^b)$ for each model,

RF¹: $\mathbf{p}^{-1}, (\widehat{\mathbf{p}})^{-1}, \mathbf{Var}(\widehat{\mathbf{p}})^{-1}$

RF²: $\mathbf{p}^{-2}, (\widehat{\mathbf{p}})^{-2}, \mathbf{Var}(\widehat{\mathbf{p}})^{-2}$

RF³: $\mathbf{p}^{-3}, (\widehat{\mathbf{p}})^{-3}, \mathbf{Var}(\widehat{\mathbf{p}})^{-3}$

.

.

RF^B: $\mathbf{p}^{-B}, (\widehat{\mathbf{p}})^{-B}, \mathbf{Var}(\widehat{\mathbf{p}})^{-B}$

where for each observation, $\widehat{\mathbf{p}}$ is a mean of prediction overall tree-level, $\mathbf{Var}(\widehat{\mathbf{p}})$ is a variance of predictions and \mathbf{p}^b is an observed value.

- The prediction of uncertainty was computed using the square root of the mean square error for each part of the test dataset. However, RF model information needed to calculate prediction error for new observation since RF creates an ensemble based on bootstrap sampling and it is also an ensemble of CART models.
- To calculate prediction error for new observation, it requires scale between $\mathbf{Var}(\widehat{\mathbf{p}})$ and $(\mathbf{p} - \widehat{\mathbf{p}})^2$ due to $\mathbf{Var}(\widehat{\mathbf{p}})$ is only available. The parameter τ measure delivers the scaling and it is equal to square root of $\{(\mathbf{p} - \widehat{\mathbf{p}})^2 / \mathbf{Var}(\widehat{\mathbf{p}})\}$
- In this simulation approach, $\widehat{\mathbf{t}}$ can be computed at 95 % confidence level and it maintains such that 95 % of the predictions occupy in $\widehat{\mathbf{t}} \times \mathbf{SD}(\widehat{\mathbf{p}})$ of observing (or true) value that is calculated using 95th percentile of all member within T (where $T = \tau^{-1}, \tau^{-2}, \tau^{-3}, \dots, \tau^{-B}$).

The above process is used to map the uncertainty of the forest AGB predictions over the study area using the variables sample with respect to their probabilistic characteristics. This approach is executed in R.

After the implantation of the above methods, the required results can be achieved that presents in the next chapter.

5. RESULTS AND ANALYSIS

This chapter presents all results derived after performing the different methods (i.e., described in chapter 4) and its analysis. The first section shows the result of the extracting spectral and textural variables from the RapidEye image. The second section presents the creation of airborne LiDAR metrics including normal and pit-free CHM. The third section shows the calculation of forest AGB using field measured data. Similarly, further sections present comparison of forest AGB estimation using different AGB prediction models and ML algorithms, selection of the best AGB prediction model along with ML algorithm, variable importance, the spatial distribution pattern of the forest AGB, and their validation and uncertainty mapping sequentially.

5.1. Extracted variables from the RapidEye image

5.1.1. Spectral variables

In addition to the 5 band reflectance, the 22 vegetation indices (section 4.1. and Table 6) were generated which is shown in Appendix 6. Out of 22 vegetation indices (VIs), red and red-edge bands were used in the equation of NDVI, MSR and CI_{green} in the place of red/green bands along with some modified and existing VIs to find the correlation with field measured biomass (Table 6).

5.1.2. Textural variable

The optimal window size of 5x5 was selected from the various moving window sizes such as 3x3, 5x5, 7x7, 9x9, 11x11 and 13x13 (Figure 16) for calculating texture variables using GLCM method. The eight texture variables (Table 7) were computed from each available band of the image using window size 5x5.

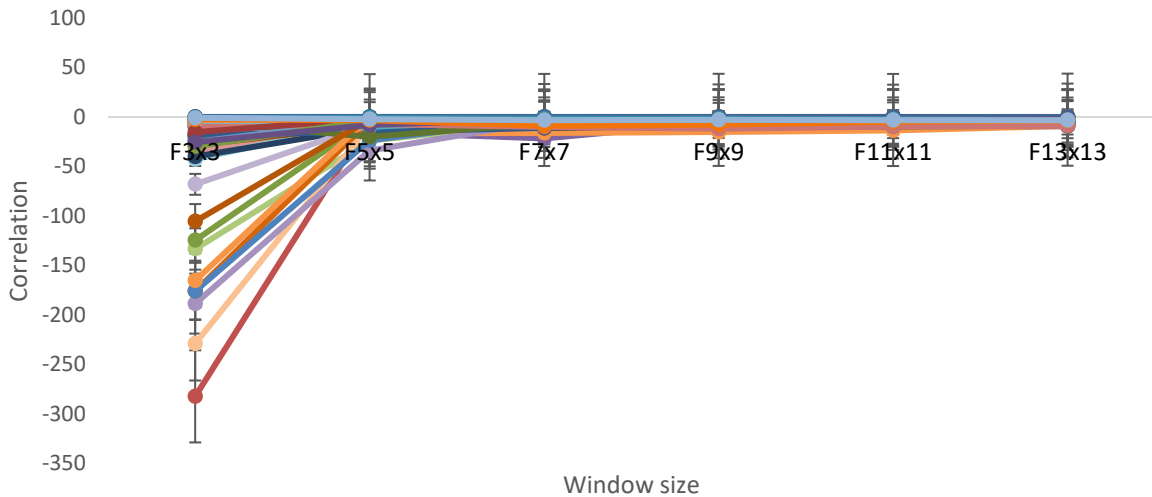


Figure 16: Selection of appropriate window size.

5.2. Extracted airborne LiDAR metrics

The quality of the airborne LiDAR data was checked (its procedure is given in section 4.2.1.) before calculating the LiDAR metrics and generated the lasinfo report (i.e., depicted in Appendix 3). The LiDAR strip overlap of the study area was showing to end lap and side overlap was 60 % and 30 % respectively (Table 4 and Figure 17).

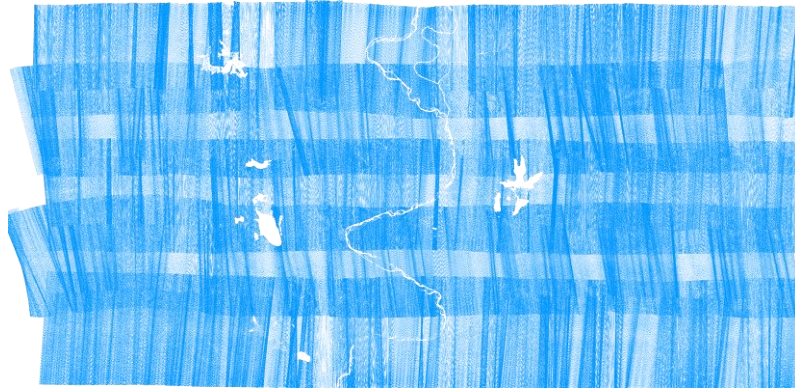


Figure 17: The LiDAR strip overlap of the study area.

5.2.1. Comparative analysis of normal CHM and pit-free CHM

The normal CHM was generated using all first returns after subtracting DTM from DSM that is shown in Figure 18. Next, pit-free CHM was created using all relevant returns of airborne LiDAR data (Figure 14). The detail procedures of normal CHM (Figure 18) and pit-free CHM are given in section 4.2.2. In normal CHM, more irregularities were appeared in canopy height compared to pit-free CHM (Figure 19 and 32).

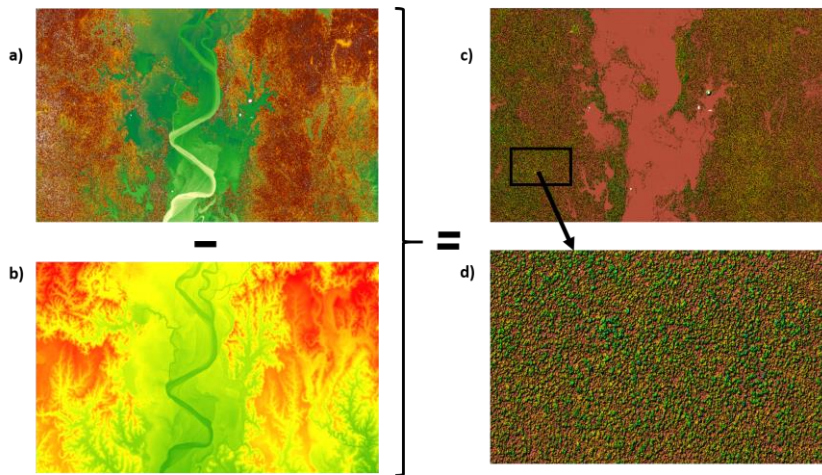


Figure 18: The normal CHM extraction (a) DSM, (b) DTM, (c) CHM and (d) enlarge portion of the normal CHM (DSM-DTM=CHM).

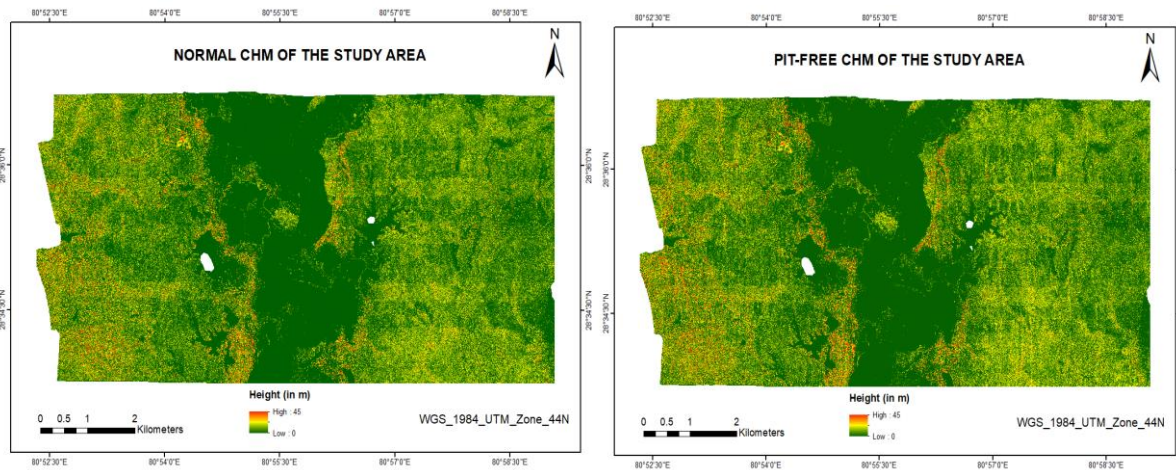


Figure 19: The normal and pit-free CHM of the study area.

Table 10: Comparison of the field measured forest tree heights with derived tree heights from pit-free-CHM and normal CHM.

SN	Longitude	Latitude	Species	Forest tree heights (m)		
				Field measured	Derived from	
					Pit-free CHM	Normal CHM
1	80.895294	28.572004	<i>Terminalia alata</i> (Asna)	5.70	4.81	4.08
2	80.895294	28.574723	<i>Shorea robusta</i> (Sal)	20.30	19.25	18.41
3	80.895299	28.577493	<i>Shorea robusta</i> (Sal)	25.10	25.83	25.81
4	80.898462	28.577426	<i>Buchanania latifolia</i> (Gayo)	9.80	11.01	11.07
5	80.898362	28.580215	<i>Mallotus philippinensis</i> (Sindure)	6.70	5.99	6.89
6	80.895303	28.597352	<i>Terminalia bellirica</i> (Barro)	7.60	8.90	6.52
7	80.895222	28.600036	<i>Terminalia alata</i> (Asna)	14.80	12.51	11.76
8	80.895225	28.602807	<i>Terminalia alata</i> (Asna)	8.80	8.87	8.23
9	80.898334	28.594638	<i>Mallotus philippinensis</i> (Sindure)	8.20	7.45	5.82
10	80.963398	28.597438	<i>Terminalia alata</i> (Asna)	16.30	14.20	15.14
11	80.966474	28.594731	<i>Terminalia alata</i> (Asna)	4.30	3.60	4.45
12	80.966524	28.600081	<i>Mallotus philippinensis</i> (Sindure)	6.00	5.62	5.50
13	80.966503	28.602801	<i>Mallotus philippinensis</i> (Sindure)	13.20	11.43	11.30
14	80.890123	28.580696	<i>Sloanea tomentosa</i> (Gobre)	6.50	5.88	5.50
15	80.893114	28.575419	<i>Careya arborea</i> (Kumbhi)	10.50	9.54	12.10
16	80.893212	28.578034	<i>Terminalia bellirica</i> (Barro)	7.30	6.69	6.14
17	80.955955	28.588591	<i>Semecarpus anacardium</i> (Bhalayo)	10.20	10.14	11.60
18	80.959022	28.580481	<i>Terminalia alata</i> (Asna)	16.10	16.78	17.08
19	80.959024	28.585969	<i>Terminalia alata</i> (Asna)	12.90	13.29	14.50
20	80.881651	28.565816	<i>Nyctantbes arbor</i> (Parijat)	3.80	4.44	3.86

Table 11: A summary of statistical values for forest tree heights

Forest tree heights from	Statistical values for the forest tree heights (in m)						
	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum	Standard deviation
Field measured	3.80	6.65	9.30	10.71	13.60	25.10	5.43
Pit-free CHM derived	3.60	5.96	9.20	10.31	12.70	25.83	5.43
Normal CHM derived	3.86	5.74	9.65	10.29	12.70	25.81	5.61

Table 12: The accuracy assessment of the CHMs derived forest trees heights

Forest tree derived heights from different CHMs	Accuracy assessment				
	Correlation coefficient (r)	R ²	RMSD (in m)	RMSD _{rel}	RMSD _{CV}
Pit-free CHM	0.98	0.97	1.09	10.23	0.10
Normal CHM	0.97	0.94	1.46	13.62	0.14

The 20 fields measured tree heights of different species were compared to the derived tree heights from normal and pit-free CHM (Table 10). The maximum and minimum tree heights were found 25.10 m (*Shorea robusta*) and 3.80 m (*Nyctanthes arbor*), mean tree height was estimated 10.71 m from the field measured data. They were also compared with normal and pit-free CHM's derived heights (Table 11) and their distribution are shown in Figure 22.

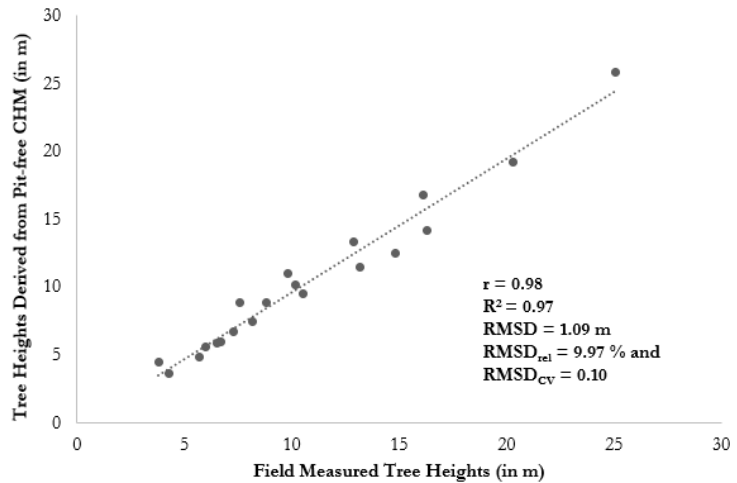


Figure 20: A fitting line for the pit-free CHM derived tree heights.

The accuracy of the derived tree heights from both CHMs was measured in terms of R^2 , RMSD, $RMSD_{rel}$ and $RMSD_{CV}$ (Equations 4.5-4.7) with respect to in-situ tree heights. Generally, the extracted heights from CHMs are not much differentiated from true (field measured) heights (Table 10). However, the statistical result showed that tree heights derived from pit-free CHM were more accurate with an R^2 of 0.97, RMSD of 1.09 m, $RMSD_{rel}$ of 10.23 % and $RMSD_{CV}$ of 0.10 than the normal CHM with an R^2 of 0.94, RMSD of 1.46 m, $RMSD_{rel}$ of 13.27 % and $RMSD_{CV}$ of 0.13 (Table 12, Figure 20 and 21).

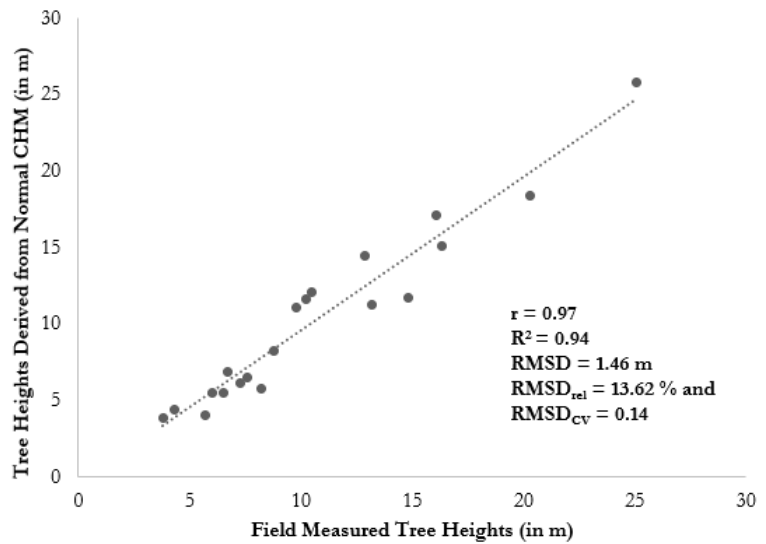


Figure 21: A fitting line for the normal CHM derived tree heights.

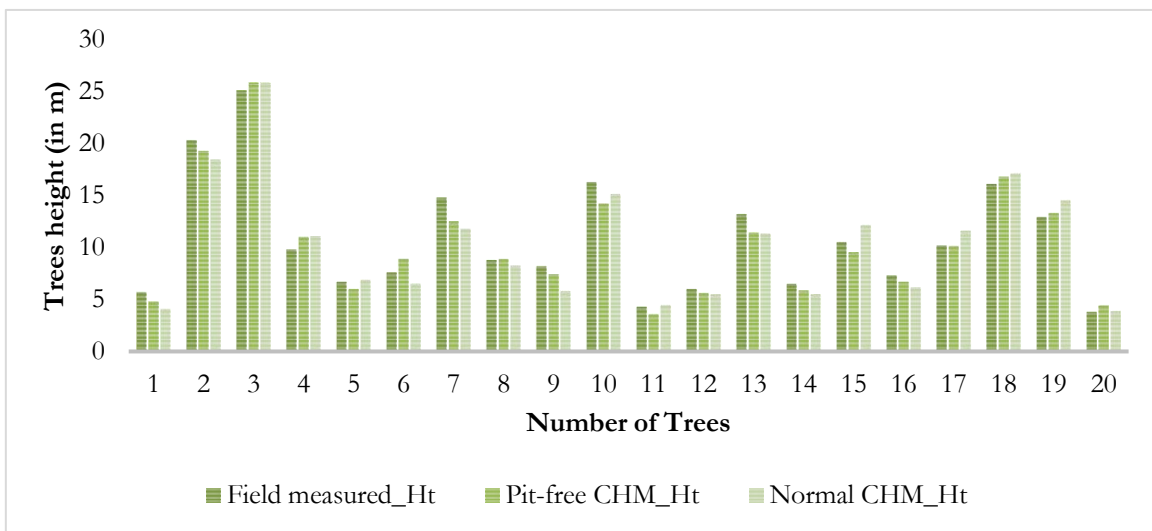


Figure 22: A distribution of field measured trees height and CHMs derived trees height

5.2.2. Extracted LiDAR metrics

Initially, 54 LiDAR metrics were extracted based on the height and intensity information. However, 52 LiDAR metrics (out of 52, 32 metrics were height based and 20 metrics were intensity based) were considered for further processing. The description of all 52 LiDAR metrics are given in Table 8 (Section 4.2.2.). Out of 52 LiDAR metrics, some new height and intensity based metrics like bincentiles, relative height points count and its density (i.e., based on both height and intensity of returns), intensity-based height and percentiles metrics were also created with other existing metrics to test the correlation with field measured biomass (Table 8).

5.3. Forest AGB calculated from field measured sample plots

The forest AGB was computed from the field measured data (the detail calculation techniques are given in section 4.3) using DBH, height, related volume equation and wood density. The calculated field measured forest AGB of 76 sample plots are shown in Table 13.

Table 13: The field measured forest AGB (in Mg ha⁻¹) of 76 sample plots with their location.

SN	Forest AGB	SN	Forest AGB	SN	Forest AGB	SN	Forest AGB	SN	Forest AGB
1	131.23	17	239.96	33	254.66	49	237.2	65	237.63
2	276.26	18	176.27	34	157.81	50	147.39	66	140.35
3	217.33	19	144.01	35	522.6	51	130.45	67	75.87
4	162.73	20	111.15	36	115.4	52	167.89	68	150.36
5	144.26	21	99.8	37	189.56	53	124.68	69	237.49
6	133.44	22	156.05	38	342.6	54	155.14	70	51.76
7	280.28	23	305.49	39	191.88	55	76.01	71	135.49
8	347.48	24	216.02	40	205.83	56	341.47	72	289.27
9	168.38	25	204.17	41	184.8	57	329.14	73	254.78
10	175.88	26	226.18	42	135.7	58	131.89	74	247.78
11	429.11	27	111.67	43	349.88	59	155.55	75	315.05
12	331.95	28	95.34	44	208.74	60	160.93	76	245.47
13	71.87	29	212.38	45	163.44	61	207.62		
14	215.52	30	236.7	46	152.23	62	161.74		
15	200.5	31	187.75	47	126.46	63	58.44		
16	283.61	32	174.18	48	222.47	64	77.51		

The value forest AGB is ranged from 51.76 to 522.60 Mg ha⁻¹ (Table 14). Also, the mean AGB is found as 197.83 Mg ha⁻¹ and it is slightly lower than the average value (202.64 Mg ha⁻¹) determined by forest resource assessment of the Terai forests of Nepal (DFRS, 2014).

Table 14: The descriptive statistics of field measured biomass

Total sample plots	Minimum (Mg ha ⁻¹)	Maximum (Mg ha ⁻¹)	Mean (Mg ha ⁻¹)	Standard deviation
76	51.76	522.60	197.83	86.41

5.4. Comparison of forest AGB prediction models using RF regression algorithm

The seven forest AGB prediction models (i.e., described in section 4.5) were run with RF regression algorithm using best M_{try} with help of executing the *tuneRF* function. Also, the values of N_{tree} were tuned from 100 to 2000 with intervals of 100 to choose optimal N_{tree} based on their accuracy. The values selected

for N_{tree} , M_{try} , R^2 , RMSE, $RMSE_{rel}$ and $RMSE_{cv}$ of the seven AGB prediction models are shown in Table 15.

Table 15: Comparisons of the seven different forest AGB prediction models using RF algorithms.

SN	Forest AGB prediction models	No. of variables	M_{try}	N_{tree}	R^2	RMSE (Mg/ha)	$RMSE_{rel}$ (in %)	$RMSE_{cv}$
1	LiDAR metrics	52	8	1300	0.93	37.59	18.45	0.18
2	Spectral variables	27	6	600	0.95	39.60	19.44	0.19
3	Textural variables	40	6	300	0.94	37.38	18.35	0.18
4	Spectral + Textural	67	10	300	0.95	37.49	18.40	0.18
5	LiDAR + Spectral	79	18	100	0.94	35.37	17.36	0.17
6	LiDAR + Textural	92	20	500	0.95	35.78	17.56	0.18
7	LiDAR + Spectral + Textural	119	18	100	0.95	35.15	17.25	0.17

The result showed that texture ($RMSE = 37.38 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 18.35 \%$) and LiDAR ($RMSE = 37.59 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 18.45 \%$) AGB prediction models accomplished better than spectral model ($RMSE = 39.60 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 19.44 \%$). It was also noticed that the LiDAR combined models were more accurate than a single model. Comparatively, LiDAR + spectral ($RMSE = 35.37 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 17.36 \%$) and LiDAR + textural ($RMSE = 35.78 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 17.56 \%$) models executed well than spectral + textural ($RMSE = 37.49 \text{ Mg ha}^{-1}$ and $RMSE_{rel} = 18.40 \%$) model. However, LiDAR + spectral + textural combined model performed best with an R^2 of 0.95, RMSE of 35.15 Mg ha^{-1} and $RMSE_{rel} = 17.25 \%$ than other model using M_{try} of 18 and N_{tree} of 100.

5.5. Comparison of forest AGB prediction models using SVM regression algorithm

In the SVM regression algorithm, the four kernels were tested on the seven forest AGB prediction models using the different parameters like cost value and SVM-type to get the optimal result. The remaining parameters had the default value as given R (i.e., $\gamma = 0.0084-0.192$, $\nu = 0.5$ and $\epsilon = 0.1$). The optimal cost value (C) was chosen after several iteration with different combination C, kernel functions and SVM-type (i.e., Eps-regression and Nu-regression) against accuracy (RMSE) of the AGB prediction models (Table 16).

Table 16: The selected cost parameter and SVM-type under the four kernel functions.

SN	Model	SVM Kernel functions and their parameters							
		Linear		RBF		Polynomial		Sigmoid	
		Cost	Type	Cost	Type	Cost	Type	Cost	Type
1	LiDAR metrics	1	Nu	4	Eps	1	Eps	2	Nu
2	Spectral variables	20	Nu	15	Nu	1500	Nu	1	Nu
3	Textural variables	2	Eps	5	Eps	20	Eps	1	Eps
4	Spectral + Textural	2	Eps	2	Eps	10	EPs	1	Eps
5	LiDAR + Spectral	1	Eps	5	Eps	1	Eps	1	Eps
6	LiDAR + Textural	1	Nu	5	Eps	1	Eps	1	Eps
7	LiDAR+Spectral+Textural	1	Eps	1	Nu	2	Eps	5	Nu

The C values selected for each model with the best kernel function are shown in Figure 23. The detail descriptions of this technique are described in section 4.4.2. The testing performance of the seven AGB prediction models and the four kernels are depicted in Table 17 and 18.

Table 17: Testing performance of the models using linear and RBF kernel function of SVM.

SN	Forest AGB prediction models	Kernel					
		Linear			RBF		
		RMSE (Mg/ha)	RMSE _{rel} (in %)	RMSE _{CV}	RMSE (Mg/ha)	RMSE _{rel} (in %)	RMSE _{CV}
1	LiDAR metrics	76.43	37.51	0.38	52.23	25.63	0.26
2	Spectral variables	68.44	33.59	0.34	69.13	33.93	0.34
3	Textural variables	58.41	28.67	0.29	69.30	34.01	0.34
4	Spectral + Textural	58.06	28.50	0.28	62.17	30.51	0.31
5	LiDAR + Spectral	88.68	43.52	0.44	52.20	25.62	0.26
6	LiDAR + Textural	153.47	75.32	0.75	52.18	25.61	0.26
7	LiDAR + Spectral + Textural	81.66	40.08	0.40	48.29	23.70	0.24

Table 18: Testing performance of the models using polynomial and sigmoid kernel function of SVM

SN	Forest AGB prediction models	Kernel					
		Polynomial			Sigmoid		
		RMSE (Mg/ha)	RMSE _{rel} (in %)	RMSE _{CV}	RMSE (Mg/ha)	RMSE _{rel} (in %)	RMSE _{CV}
1	LiDAR metrics	405.00	198.77	1.99	52.42	25.73	0.26
2	Spectral variables	67.68	33.22	0.33	69.20	33.96	0.34
3	Textural variables	72.82	35.74	0.36	69.41	34.07	0.34
4	Spectral + Textural	71.48	35.08	0.35	69.30	34.01	0.34
5	LiDAR + Spectral	404.00	198.28	1.98	52.42	25.73	0.26
6	LiDAR + Textural	310.80	152.54	1.53	52.40	25.72	0.26
7	LiDAR + Spectral + Textural	171.00	83.93	0.84	59.08	29.00	0.29

The result showed that linear kernel provided better result for texture (RMSE = 58.41 Mg ha⁻¹ and RMSE_{rel} = 28.67 %) and spectral + texture (RMSE = 58.06 Mg ha⁻¹ and RMSE_{rel} = 28.50 %) model, and polynomial kernel found good for spectral (RMSE = 67.68 Mg ha⁻¹ and RMSE_{rel} = 33.22 %) model. It was also found that LiDAR metrics model showed high accuracy than the spectral model and textural model but LiDAR combination models were carried out with more accurate than any individual models and spectral and textural combined model (Table 17 and 18). The RBF kernel function performed well for all LiDAR metrics involved models such as LiDAR metrics, LiDAR + spectral, LiDAR +textural, and LiDAR + Spectral + Textural. Remarkably, the LiDAR + spectral + textural combined model was presented better output with an RMSE of 48.29 Mg ha⁻¹ and RMSE_{rel} of 23.70 % compared to other models.

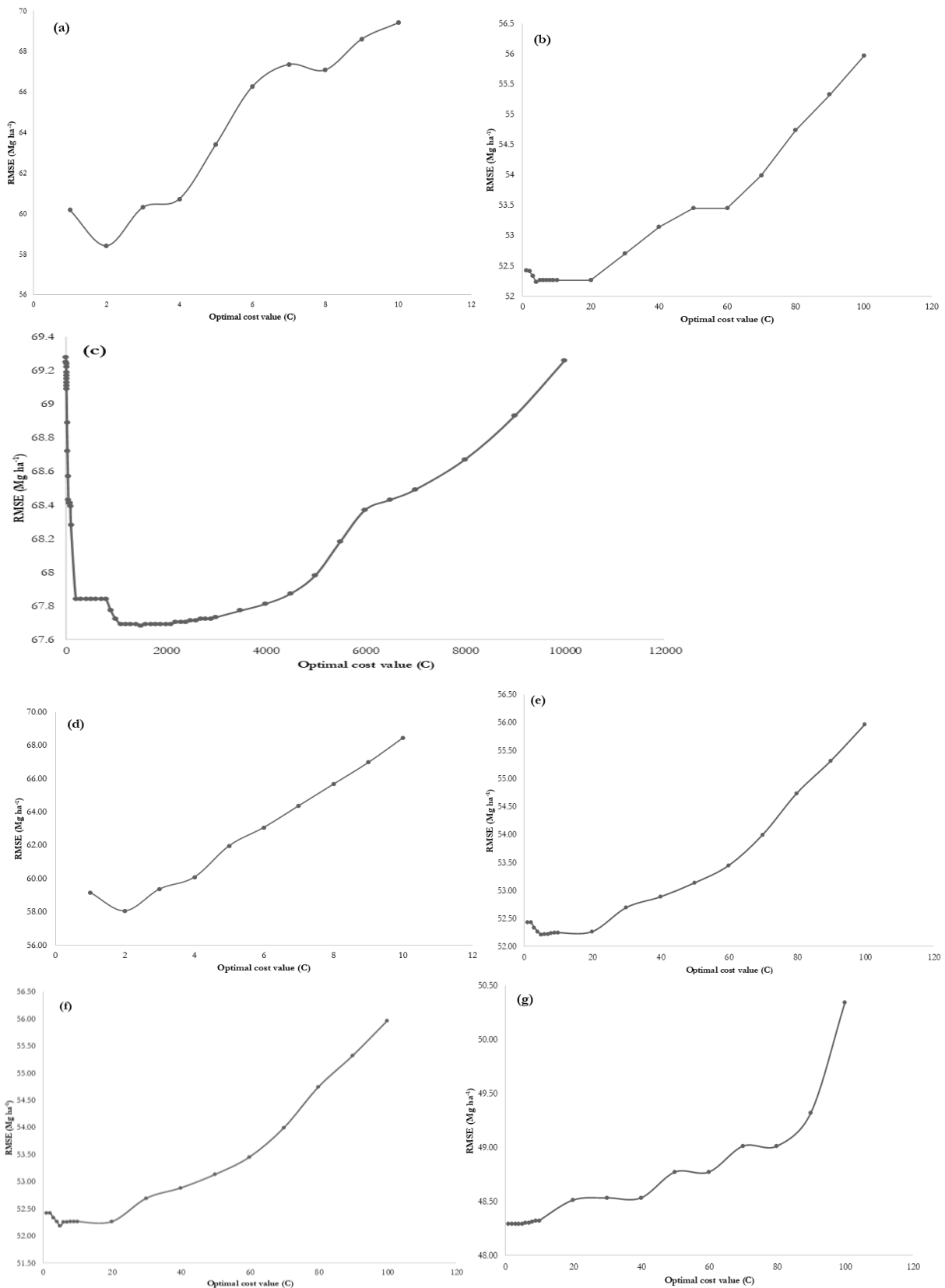


Figure 23: An optimal cost value for the different kernel functions that is used in the seven AGB predictions models sequentially: (a) linear kernel with textural model (b) RBF with LiDAR (c) polynomial with spectral (d) linear with spectral + textural (e) RBF with LiDAR + spectral (f) RBF with LiDAR + textual (g) RBF with LiDAR + spectral + textual.

5.6. Selection of the best forest AGB prediction model with ML regression algorithm

The LiDAR, spectral and textural combined model (119 variables) was selected as the best model out of seven different models using both RF and SVM regression ML algorithms. However, it was found that RF demonstrated higher accuracy with an R^2 of 0.95, RMSE of 35.15 Mg ha⁻¹, RMSE_{rel} of 17.25 % and RMSE_{CV} of 0.17 than SVM which provided accuracy with an R^2 of 0.40, RMSE of 48.29 Mg ha⁻¹, RMSE_{rel} of 23.70 % and RMSE_{CV} of 0.24 (Table 19). RF regression was used for further processing of the research such as variable importance (i.e., described in the next section) since it performed better than SVM.

Table 19: Training performance of the LiDAR, spectral and textural combined model of AGB prediction

Best forest AGB prediction model	Effectiveness of ML algorithms							
	RF regression				SVM regression with RBF kernel function			
	R^2	RMSE (Mg ha ⁻¹)	RMSE _{rel}	RMSE _{CV}	R^2	RMSE (Mg ha ⁻¹)	RMSE _{rel}	RMSE _{CV}
LiDAR + Spectral + Textural	0.95	35.15	17.25	0.17	0.40	48.29	23.7	0.24

5.7. Variable importance using RF regression algorithm

The selection of variable importance is a key feature of RF ML algorithm which is an important consideration in forest AGB estimation process (Feng et al., 2017; Pandit et al., 2018a). It assigned the score of importance for all used variables (i.e., 119 variables) using OOB sample data based on the dependent variable, i.e., field measured forest AGB, for ranking in descending order in terms of %IncMSE and IncNodePurity (Figure 24). The term %IncMSE means percentage of increasing mean square error that indicates the error quantity trend in the AGB prediction model due to the absence of the specific variable. Also, IncNodePurity represents increasing node purity that denotes the scale of purity in the node of the N_{tree} in the presence of the particular variables in the AGB prediction model (Genuer et al., 2010; Pandit et al., 2018a). Further, RF also provides the optimal subset of the predictor variables from the used forest AGB prediction model for a better estimate of the dependent variables. It determined based on increasing R^2 and decreasing RMSE using the 10-fold cross-validation (Dang et al., 2019; Genuer et al., 2010; Pandit et al., 2018a).

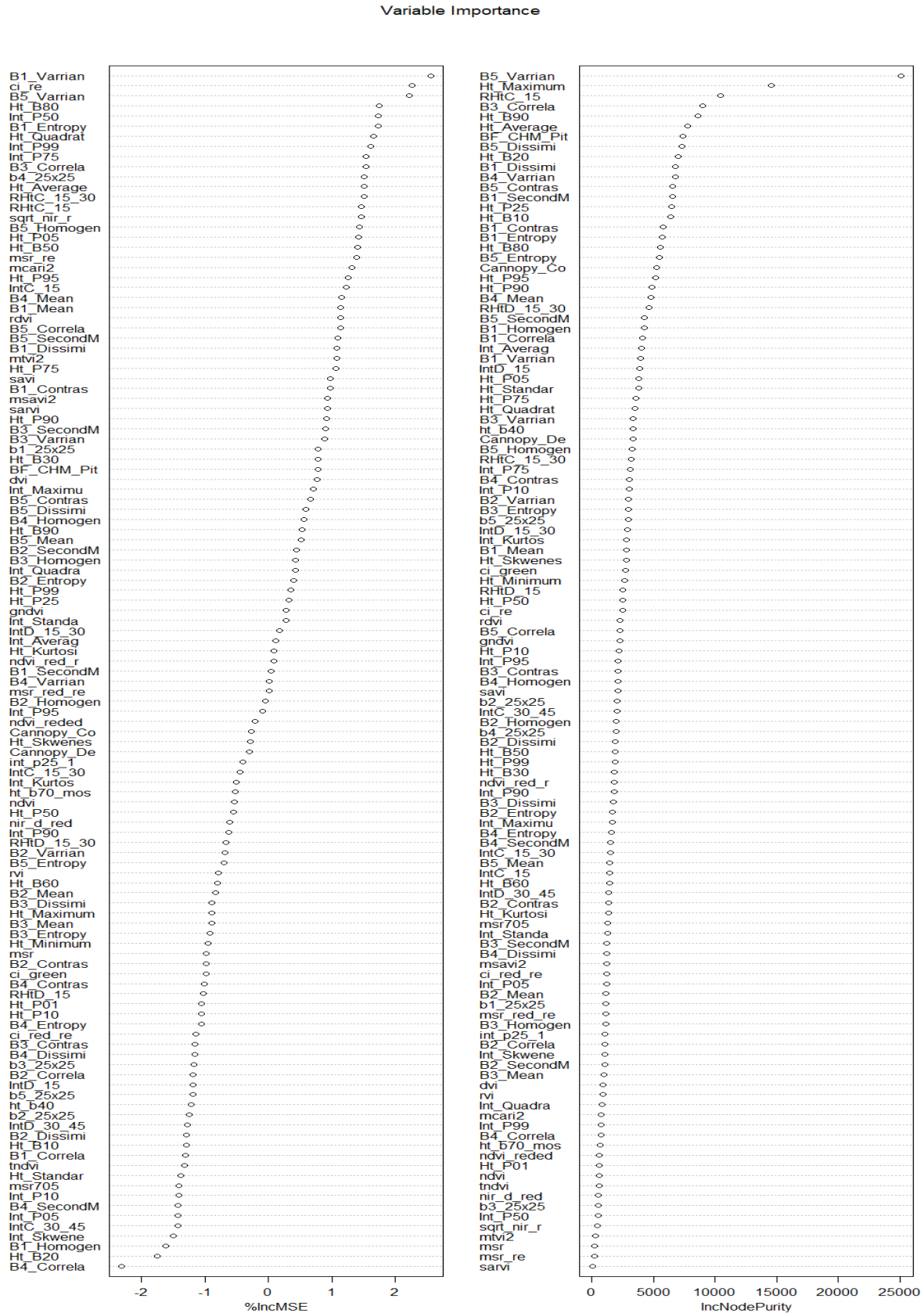


Figure 24: The variables importance of Lidar + Spectral +Textural combined model (119 variables) for forest AGB prediction model (all used abbreviations of variables are described in section 4.1. and Table 6, 7 and 8).

In this research, more than 70 iterations were performed to find the optimal subset of the predictor variables for the forest AGB prediction in R. Finally, the top 20 number of predictor variables from the 119 variables of the forest AGB prediction model (i.e., the model's number 7) were chosen with an R^2 of 0.93 and RMSE of 35.46 Mg ha^{-1} (Figure 25 and 26).

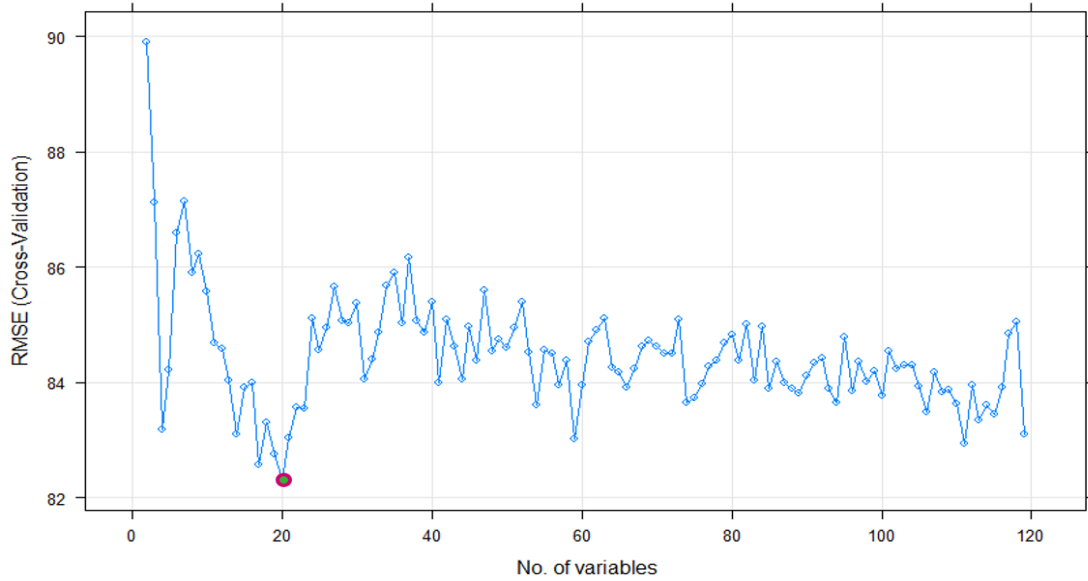


Figure 25: The choice of an optimum subset of the predictor variables using 10-fold cross-validation

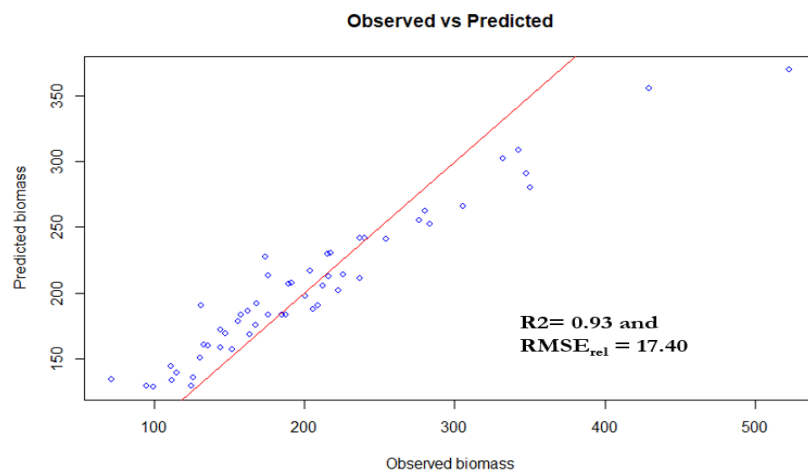


Figure 26: The accuracy of the selected optimum subset of the predictor variables

5.8. The spatial distribution pattern of the forest AGB

The selected top 20 predictor variables (Figure 27 and 28) were used to predict the forest AGB and map their spatial distribution pattern in the tropical forest of TAL area in Nepal (Figure 28) using RF regression. It was also noticed that the 10, 9 and 1 number of predictor variables were chosen from the LiDAR, textural and spectral variables for the subset of predictor variables. Moreover, LiDAR metrics and textural variables found the highly correlated with field measured biomass in comparison to spectral variables. The forest AGB was predicted over the study area at the ranged between $123.99 \text{ Mg ha}^{-1}$ to $350.64 \text{ Mg ha}^{-1}$.

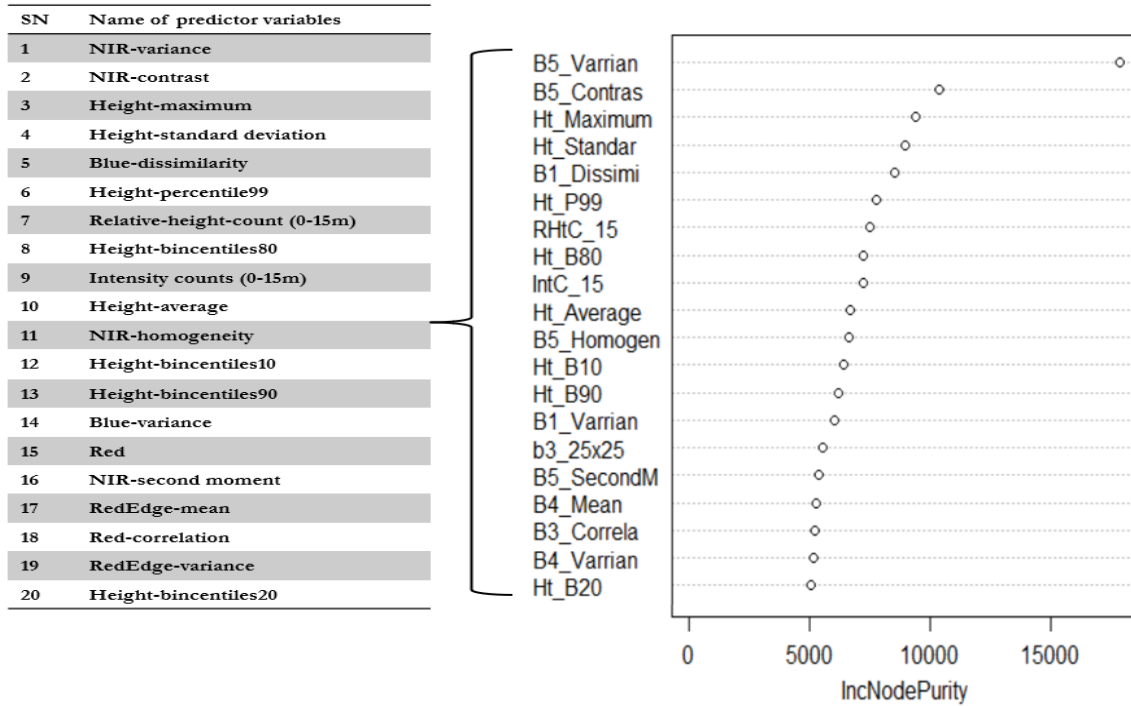
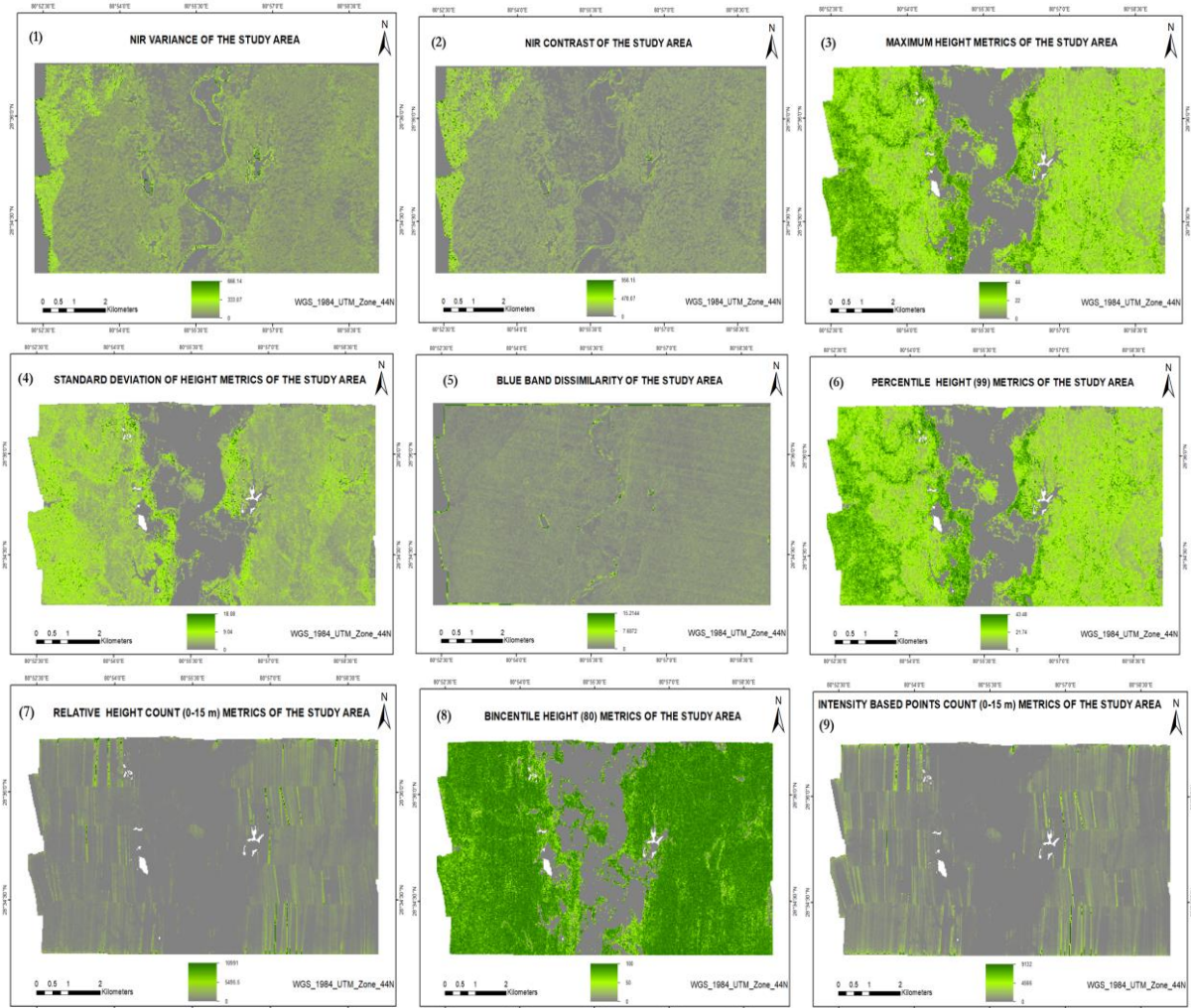


Figure 27: The top 20 selected predictor variables based on the increasing node purity



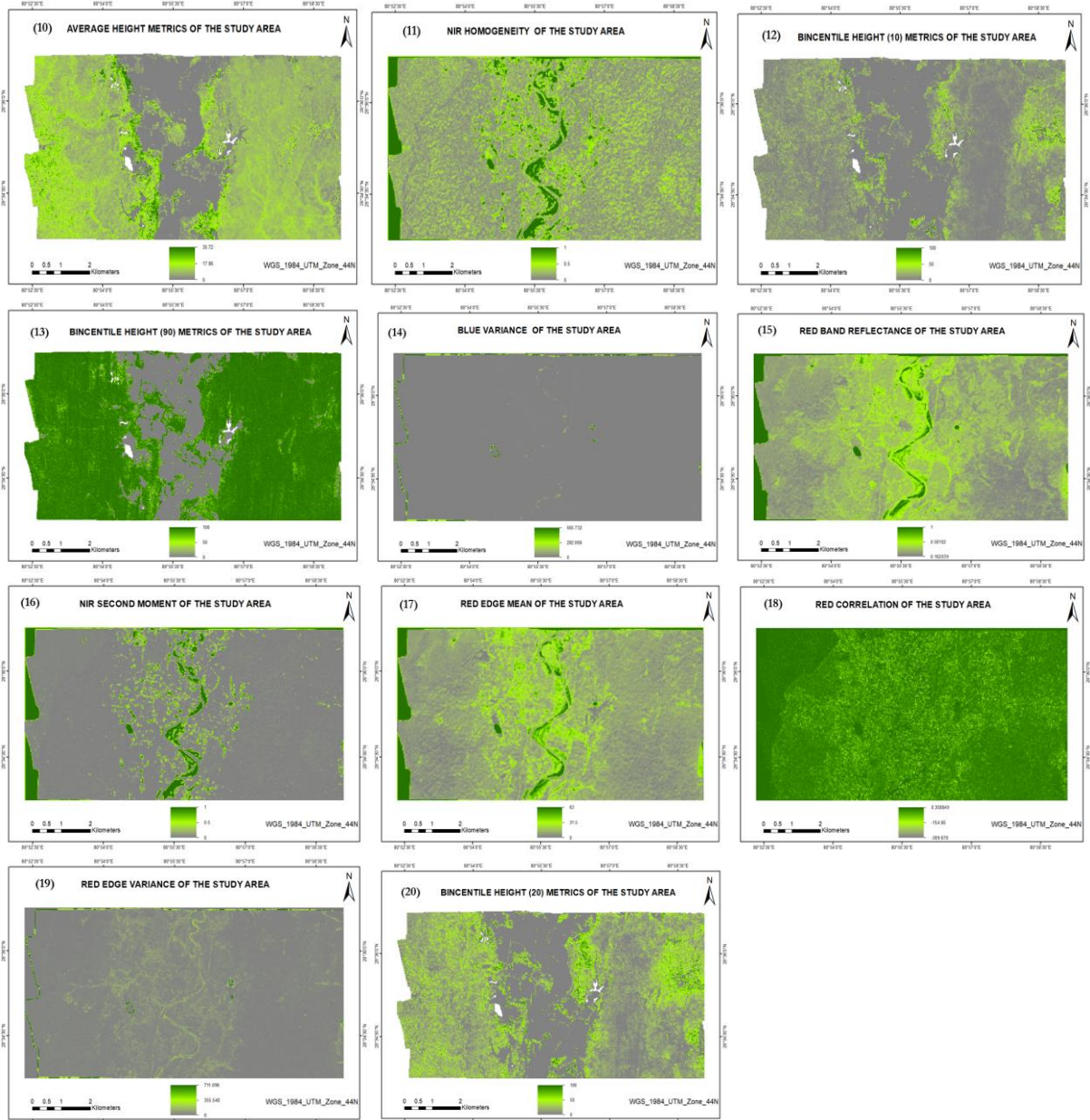


Figure 28: The top 20 selected predictor variables.

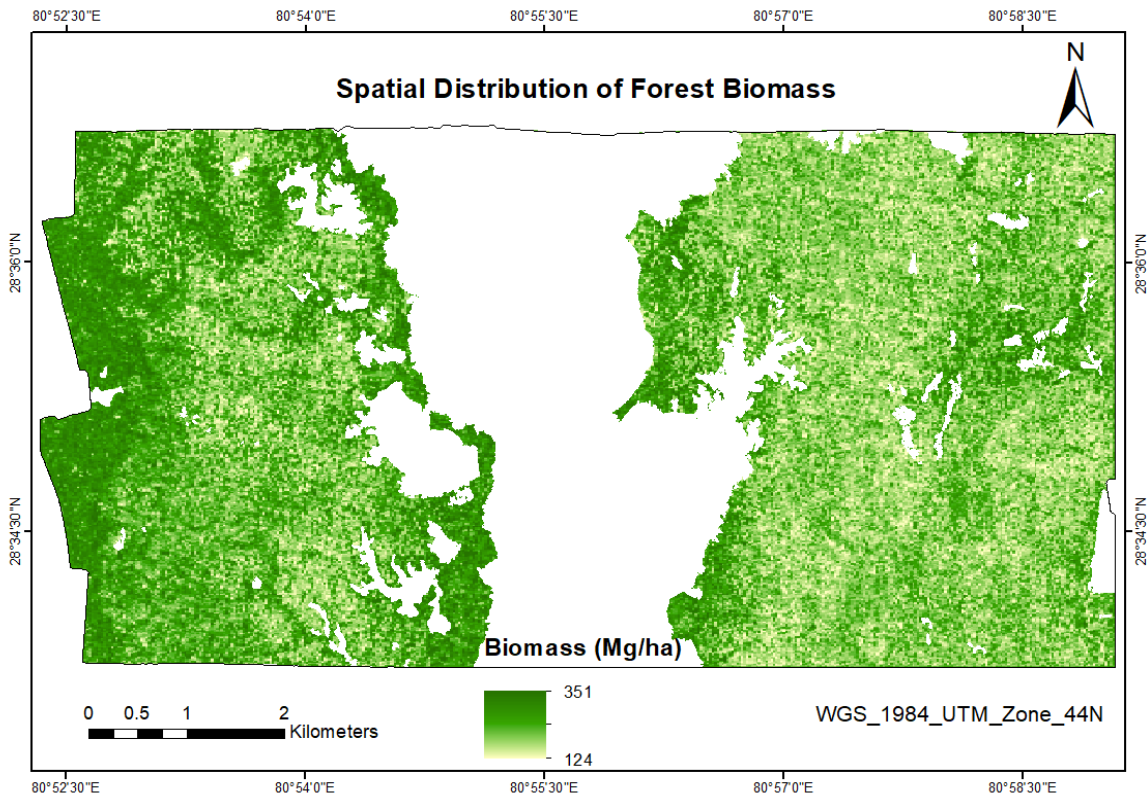


Figure 29: The spatial distribution pattern of forest AGB over the study area

The generated spatial distribution pattern of forest AGB was validated against the 15 independent sample plots using the validation measures namely R^2 , RMSE, $RMSE_{rel}$, and $RMSE_{CV}$. The estimated and validation result measures ($R^2 = 0.72$, $RMSE = 47.71 \text{ Mg ha}^{-1}$, $RMSE_{rel} = 23.41 \%$ and $RMSE_{CV} = 0.23$) of the forest AGB are shown in Table 20 and Figure 30.

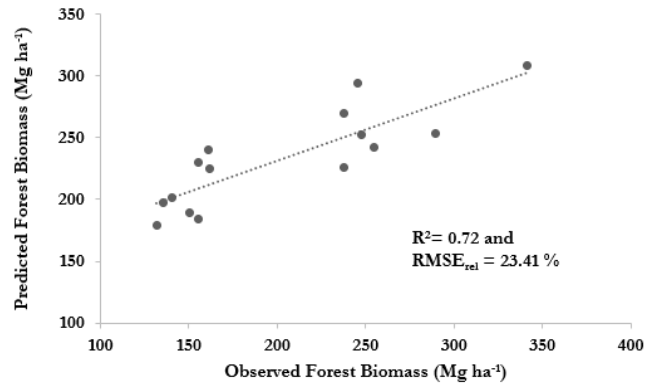


Figure 30: The regression line and accuracy for the validation.

Table 20: The forest AGB prediction result and their validation.

Forest AGB estimates (Mg ha^{-1}) over the study area				Validation measure			
Minimum	Maximum	Mean	Standard deviation	R^2	RMSE (Mg ha^{-1})	$RMSE_{rel}$ (in %)	$RMSE_{CV}$
123.99	350.64	222.54	37.74	0.72	47.71	23.41	0.23

5.9. Uncertainty mapping of the forest AGB

Figure 30 shows uncertainty in the estimation of the forest biomass where green and red colour indicate the presence of minimum and maximum uncertainty. The uncertainty mapping presented that the error ranged from 0 to 34 Mg ha⁻¹. It was also found that the mean and standard deviation of the uncertainty was 13.10 Mg ha⁻¹ and 6.83 Mg ha⁻¹ respectively. It was analysed that edge affected, transitional and dense forest area demonstrated more uncertainty as compared to other areas (Figure 30)

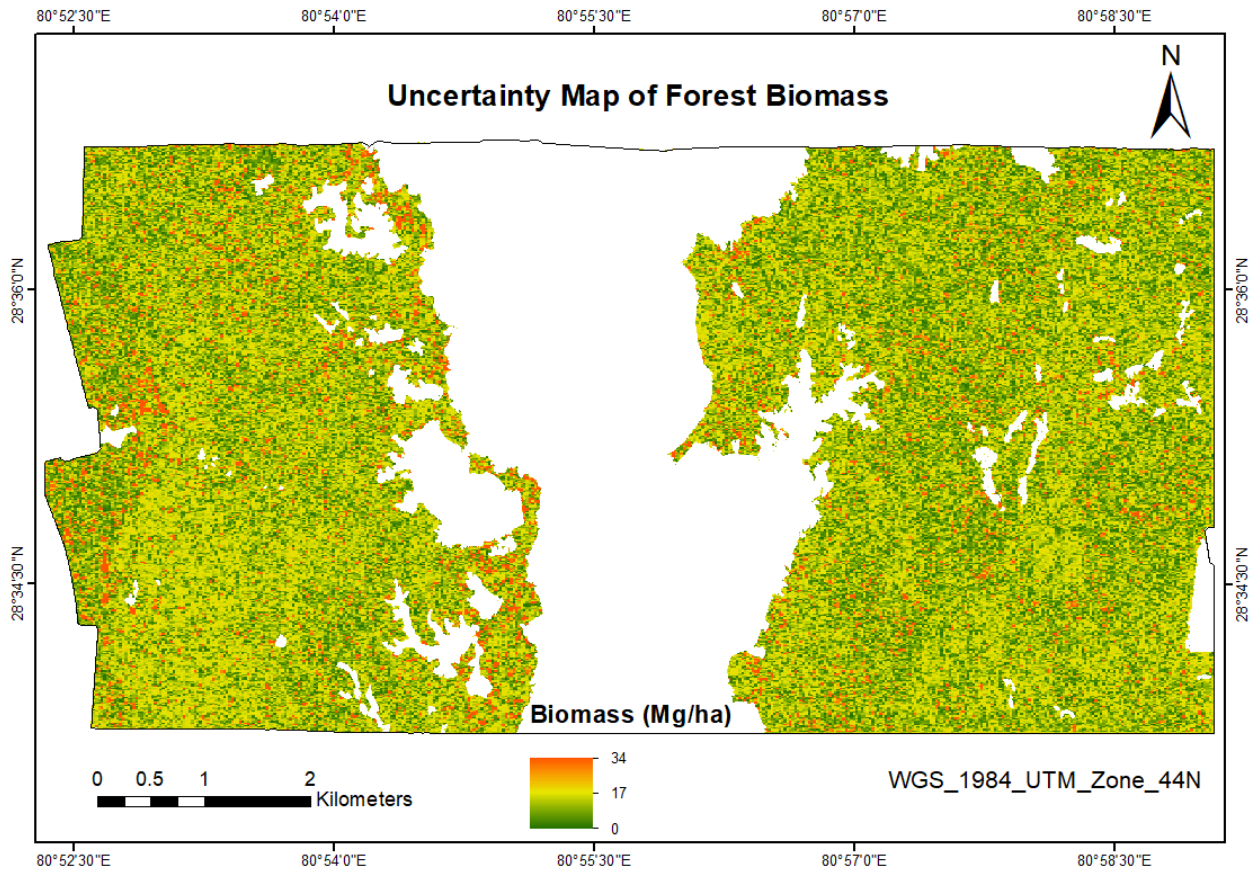


Figure 31: The uncertainty mapping of the forest AGB over the study area.

6. DISCUSSION

This chapter summarises the discussion of the achieved results shown in the previous chapter. The results are also compared with related past studies. A brief overall analysis of this research is presented in the last section.

6.1. Selection of multisensor data and extraction of their parameters

The aim of this research is to check the performance of the RF and SVM regression algorithms for estimating tropical forest AGB in the TAL area of Nepal. The choice of remote sensing data is mainly concerned as per the scope and complexity of the study area to estimate a better forest AGB (Lu et al., 2016). The high spatial resolution imageries, for examples, QuickBird, IKONOS, and RapidEye have not been widely explored as compared to Landsat imagery (Lu, 2006; Lu et al., 2005; Thenkabail et al., 2004). Feng et al. (2017) concluded that RapidEye image could not deliver the good result of forest AGB estimation especially in the case where the forest is too dense and too sparse or small (i.e., the condition where soil/grass may affect in the spectral reflectance). In the dense forests area where forest AGB is too high, the spectral reflectance data cannot address the AGB estimation because of the saturation, integration of LiDAR with the RapidEye data can solve this type of problem (Feng et al., 2017; Lu et al., 2016). In this context, airborne discrete LiDAR data together with RapidEye image used in this research to get the benefit of both data's ability for representing the different surface features. This research also disclosed that multi-sensor data using ML algorithms can provide better AGB estimation than a single sensor as previous study (Dhanda et al., 2017) has presented.

6.2. Comparison of normal CHM and pit-free CHM

The CHM is used to extract the several relevant forest inventory structure parameters such as a crown of the tree, height of the tree and also whole tree delineation. Therefore, a creation of the qualitative and smoothness CHM from LiDAR data is essential for the detection of a tree and their other biophysical measure accurately (Chen et al., 2006; Khosravipour et al., 2014; Yao et al., 2012). It also compared the normal and pit-free CHMs (the methods of CHMs creation is described in section 4.1.2) and depicted in Figure 31. Figure 31 (a) shows the normal CHM where some part with pits (or irregularities in canopy surface elevation) encircled for the demonstration purpose. This types of pits influencing CHM cannot give satisfactory results as compared to pit-free CHM, that is shown in Figure 32 (b) (Khosravipour et al., 2014).

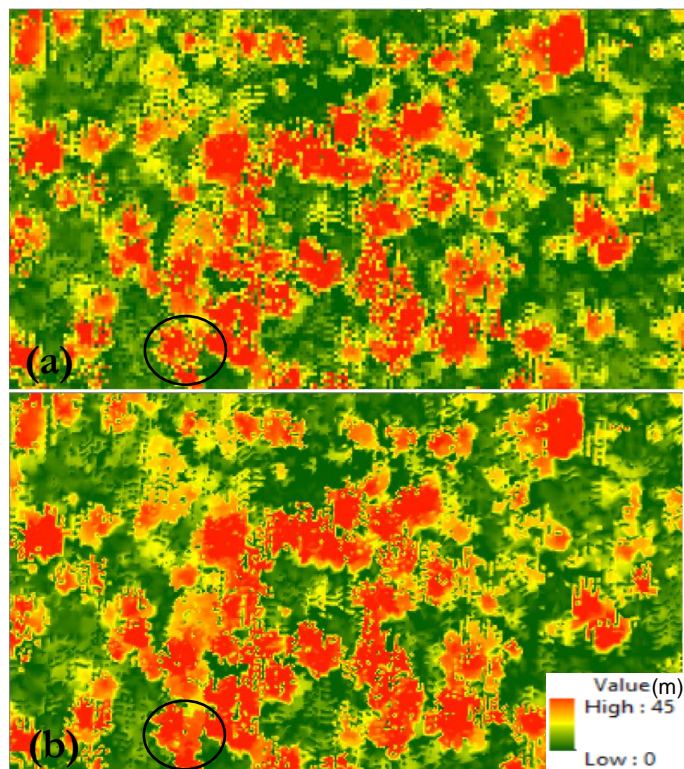


Figure 32: (a) Normal CHM generation using all first returns only
(b) Pit-free CHM generation using all relevant returns.

This research also found that pit-free CHM can provide better tree height (RMSE = 1.07 m, $R^2 = 0.97$) than normal CHM (RMSE = 1.42 m, $R^2 = 0.94$). The finding of this related to past studies, for instances, Gaveau and Hill (2003) found that first return of LiDAR data generate underestimated CHM, this type of CHMs can also reduce the accuracy in recognition of tree, height of tree and its other parameters (Ben-Arie et al., 2009; Zhao et al., 2009b). Further, Shamsoddini et al. (2013) specified that pit may more influence with the measurement of the forest variables like a volume of the tree and their basal area.

6.3. Performance of the RF and SVM ML regression algorithms with AGB prediction models

Both non-parametric algorithms (RF and SVM) can offer better estimation than linear regression models particularly in the case of multisource data (Lu et al., 2016). However, it is hard to decide that one ML algorithm is better than other until they are not evaluated separately, the output depends on several considerations like forest types, data types, topographical structures, biomass prediction models, algorithms and so forth (Kumar & Mutanga, 2017; Lu, 2006, 2007; Lu et al., 2016; Pandit et al., 2018b). It was evaluated the ML regression algorithms on the basis of their performance using AGB prediction models, both algorithms selected the same AGB prediction model (i.e., LiDAR, spectral and textural combined model) as the best performer. Whereas, RF performed better ($R^2 = 0.95$, RMSE = 35.15 Mg ha⁻¹, RMSE_{rel} = 17.25 and RMSE_{CV} = 0.17) than SVM ($R^2 = 0.40$, RMSE = 48.29 Mg ha⁻¹, RMSE_{rel} = 23.70 and RMSE_{CV} = 0.24). This result was in line with related existing studies. For instances, Liu et al. (2017) found satisfactory modelling accuracy ($R^2 = 0.95$ and RMSE = 17.73 Mg ha⁻¹) as compared to the use of support vector regression (SVR) and stepwise regression. Similarly, Wu et al. (2016) summarized that RF performs better with R^2 of 0.63 and RMSE of 26.44 Mg ha⁻¹ than SVR, k-nearest neighbour (kNN) and stepwise linear regression. Pandit et al. (2018b) also revealed that RF provides a good result ($R^2 = 0.87$ and RMSE = 20.5 Mg ha⁻¹) than multiple linear regression. Further, Dang et al. (2019) concluded that RF can capable to estimate the spatial distribution pattern of forest biomass with R^2 of 0.81 and RMSE of 36.67 Mg ha⁻¹.

In addition to the better performance of RF, this research also benefited from its important features such as, ranking the 119 variables of LiDAR, spectral and textural combined model (R^2 of 0.95, RMSE of 35.15 Mg ha⁻¹) and selection of the 20 predictor variables as optimal subset predictor variables (R^2 of 0.93, RMSE of 35.46 Mg ha⁻¹) for biomass estimation. The discussion of the top 20 predictor variables is summarised in the next section. This result is also comparable with existing studies. For example, Pandit et al. (2018b) selected the top 15 important variables with R^2 of 0.95, RMSE of 13.3 Mg ha⁻¹. Also, Dang et al. (2019) showed that the selection of the top 11 variables with least RMSE. Generally, RF has some significant advantageous to use for biomass estimation such as, easy to train, robustness in handling imprecise training data, missing, outliers and unstable dataset, delivers stable predictions, computationally efficient, and capable to block noisy predictor variables (Adam et al., 2014; Cracknell & Reading, 2014; Dube & Mutanga, 2015).

6.4. Analysis of the best predictor variables

The data saturation problem in the optical sensor image is an essential influencing factor for AGB estimation in tropical dense forests (Lu et al., 2016). This problem is reduced using the red-edge band, therefore, red-edge based vegetation indices (VIs) are important variables for measuring biomass (Pandit et al., 2018a; Sibanda et al., 2015; Xie et al., 2018). The red-edge band is capable to handle the minute changes in the canopy, gap and senescence compared to the red/green band (Xie et al., 2018). In this research, some new VIs such as NDVI_{Red&RE}, MSR_{Red&RE} and CI_{Red&RE} jointly with existing indices (Table 6) and textural variables of each band (Table 7) used for the forest AGB estimation. Additionally, some new LiDAR metrics namely, bincentile metrics (Table 8, SN: 5-13), intensity-based height and percentile metrics (Table 8, SN: 29-42), intensity and height based relative height points count and relative height-density (Table 8, SN: 43-52)

metrics were extracted (using LAStools) to test together with commonly used structural metrics (Table 8, SN: 1-4 and 14-28) for the better estimation.

Out of the top 20 variables (Figure 25), 1 red band and 9 textural variables (mainly from NIR, Red-edge and Red) selected as the predictor variables. The NIR band is most important for forest ecology since the plant reflects high energy in this region (Nandy et al., 2017). Therefore, the NIR band has reliably appeared as a topmost order with the texture features. Next, it is found that red-edge band (i.e., important for handling data saturation problem) and the red band have also a strong relationship with the biomass. This finding is related to the existing findings. For examples, Nandy et al. (2017) presented that short-wave infra-red (SWIR) and NIR as uppermost predictor variables out of the top 10 variables. Likewise, Pandit et al., (2018a) found that NIR, red, SWIR and red-edge bands as a top ordered variables for predicting biomass. It also observed that 9 textural variables (i.e., variance, contrast, dissimilarity, homogeneity, second angular momentum and mean) attended as predictor variables. Since the study area of this research is a natural tropical forest (where heterogeneous landscape and complex forest tree structure are present). Lu et al. (2005) concluded that the textural features are more important than the spectral signature for forest AGB estimation in the natural forests where high variability and complex forest stand exist. This output is similar to Dhanda et al. (2017). They also found that texture variables can provide more contribution than spectral variables for biomass prediction.

The height is a major component for biomass estimation (Lu et al., 2016), therefore the rest of predictor variables have appeared from LiDAR metrics (mainly from height metrics -maximum, standard deviation, average, 99th percentile height, bincentiles -10th, 20th, 80th and 90th, height based relative height points count and intensity based relative height point counts). This finding is not completely unique, however, it is in line with several existing studies where LiDAR metrics are found to have a strong relationship with the biomass. For instances, Li et al. (2017) showed that the height metrics (such as mean, skewness, interquartile range), percentile height and other height metrics (variance, standard deviation and range) as major predictors. Further, Feng et al. (2017) summarized that LiDAR produced consistent biomass prediction in the moist tropical region using height metrics (namely, mean, standard deviation, kurtosis, quadratic mean height, skewness) and percentile height (10th, 20th, 30th,90th) metrics with the help of different algorithms. Omar et al. (2015) also demonstrated that CHM was a top predictor variable out of the generated variables such as canopy density, DSM, CHM and intensities from different returns (all, 1st, 2nd, 3rd and last). It is also observed that LiDAR metrics have the highest contribution for the total number of predictor variables than spectral and textural variables for the AGB prediction.

6.5. Accuracy and uncertainty analysis of biomass estimation

In addition to R^2 , RMSE, $RMSE_{rel}$, and $RMSE_{CV}$, the 10-fold cross-validation was used to reduce overfitting problem for biomass estimation using RF regression. The result of validation measures ($R^2 = 0.72$, $RMSE = 47.71 \text{ Mg ha}^{-1}$, $RMSE_{rel} = 23.41 \%$ and $RMSE_{CV} = 0.23$) showed acceptable accuracy using validation dataset (Table 20 and Figure 30). This is a similar way to check the accuracy of biomass estimation in existing studies. For example, Dang et al. (2019) presented the accuracy with an R^2 of 0.81, RMSE of 36.67 Mg ha^{-1} , %RMSE of 19.55 % using 19 validation plots.

The uncertainty analysis is an essential component of biomass study because it assesses the prediction model performance and also indicates major sources of the uncertainty that affect biomass estimation accuracy (Dang et al., 2019; Lu et al., 2012b). However, there are few studies that carried out the uncertainty analysis (Coulston et al., 2016; Dang et al., 2019). In this study, the error range was found less (0 to 34 Mg ha^{-1}) as a comparison to another study (Figure 30). For example, Dang et al. (2019) got uncertainty within the range between 9.87 to 93.27 Mg ha^{-1} using Monte Carlo simulation. It is also revealed that multi-sensor data perform better result than a single sensor using the RF algorithm.

Table 21: A list of the studies using ML algorithms for forest AGB estimation in different climatic zones.

SN	ML algorithms	Remote sensing based dataset	Study area with forest types and climatic zone	Model accuracy		Sources
				R ²	(RMSE (Mg/ha ⁻¹))	
1	ANN	Resourcesat-1 LISS-III	TMDF, subtropical climate, UK, India	0.74	93.41	Nandy et al. (2017)
2		Resourcesat-2	TDDF, Tropical climate, Bundelkhand region of India	0.98		Deb et al. (2017)
3	kNN	IRS P6 LISS-III	TMDF, Timli forest, Subtropical climate UK, India		44.23	Yadav and Nandy (2015)
4		Landsat 7 ETM+	Subtropical monsoon climate, ZP, China	0.21	39.71	Wu et al. (2016)
5	SGB	RapidEye	Plantation forests, subtropical climate, KwaZulu-Natal Province, South Africa	0.61	43.39	Dube et al. (2014)
6		Landsat 7 ETM+	Subtropical monsoon climate, ZP, China	0.55	28.64	Wu et al. (2016)
7	SVM/SVR	Landsat 7 ETM+	Subtropical monsoon climate, ZP, China	0.38	34.61	Wu et al. (2016)
8		ICESat/GLAS and WorldView-2	TMDF, subtropical climate, Doon-valley, UK, India	0.89	13.60	Dhanda et al. (2017)
9		RapidEye and Airborne LiDAR	TMH forests, Tropical climate, TAL area, FWP, Kailali, Nepal	0.40	48.29	This research
10	RF	WorldView-2	Wetland vegetation, subtropical climate, ISWP, KZNP, South Africa		4.41	Mutanga et al. (2012)
11		RapidEye	Plantation forests, subtropical climate, KZNP, South Africa	0.37	59.27	Dube et al. (2014)
12		Landsat 7 ETM+	Subtropical monsoon climate, ZP, China	0.63	26.22	Wu et al. (2016)
13		ICESat/GLAS and WorldView-2	TMDF, subtropical climate, Doon-valley, UK, India	0.84	20.57	Dhanda et al. (2017)
14		SPOT 4 and SPOT 5	Mangrove forests, subtropical climate, Cangio, Ho Chi Minh City, Vietnam	0.73	78.20	Pham and Brabyn (2017)
15		Sentinel-2	PNP, subtropical climate, Nepal	0.81	25.32	Pandit et al. (2018a)
16			YDNP, tropical monsoon climate, Vietnam	0.80	33.69	Dang et al. (2019)
17	RapidEye and Airborne LiDAR	TMH forests, Tropical climate, TAL area, FWP, Kailali, Nepal	0.95	35.15	This research	

Key: TMDF = Tropical Moist Deciduous Forest, TDDF = Tropical Dry Deciduous Forest (sub-humid or semiarid area), TMH = Terai mix hardwood, Linear Imaging Self-scanning Sensor = LISS, SPOT = Satellite for observation of Earth, SGB = Stochastic Gradient Boosting, UK = Uttarakhand, ZP = Zhejiang Province, FWP = Far-western Province, KZNP = KwaZulu-Natal Province, ISWP = ISimangaliso Wetland Park, PNP = Parsa National Park, YDNP = Yok Don National Park

6.6. Overall analysis of the study

The overall approach of biomass estimation provides a non-destructive, resources effectiveness and efficient way for a forest monitoring system in the implementation of climate change policy. The use of multi-sources data (i.e., LiDAR, the optical sensor and in-situ measurement) which is collected simultaneously to reduce biases. Moreover, the integration of high spatial resolution and better structural information is a good strategy to use all information for reducing the saturation problem and to optimise and select the optimal number of their parameters using RF regression for the prediction of AGB. However, it might produce better accuracy than obtained here if the more sample plots (i.e., taking from stratified forests) are used in this study. Further, the use of volume allometric equation due to the unavailability of local biomass allometric equation may create some uncertainty. Similarly, the availability of only five bands of the optical sensor (RapidEye) and sparse density of LiDAR data can potentially cause to limit the accuracy. This weakness and threats should be solved in further research in this area to improve this approach for estimating forest AGB and its monitoring in the aspect of spatiotemporal change at the sub-country level.

7. CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the main findings and makes the conclusions from the overall study on the basis of the result and discussion. As a recommendation, it includes some suggestions to improve the existing approach through the crucial interventions for further studies in this field.

7.1. Conclusions

In this search, airborne LiDAR and RapidEye parameters were integrated and optimized to estimate forest AGB and to evaluate the performance of the ML algorithms (i.e., SVM and RF) and AGB prediction models used in the tropical forests of TAL area in Nepal. The predicted forest AGB spatial distribution map using the best predictors variables was validated using independent in-situ data and mapped its uncertainty with the help of Monte Carlo simulation. In this regard, the following conclusions are drawn to address the proposed research questions under each specific objective.

Specific objective 1: “Extract the different spectral and textural variables from the high-resolution optical satellite data.”

What are the most appropriate spectral and texture variables for forest AGB estimation?

The 27 spectral including 5 band reflectance and 22 VIs (such as NDVI, DVI, SAVI, MSR and so forth) were derived from the reflectance image, where some new VIs (i.e., $NDVI_{Red\&RE}$, $MSR_{Red\&RE}$ and $CI_{Red\&RE}$) were also extracted using red and red-edge band instead of red/green to check the relation with the biomass (Table 6). Additionally, the 40 textural variables (i.e., 8 texture measure for each band of RapidEye image) were extracted using GLCM method and the window size of 5x5 (Section 4.1 and 5.1). All these spectral and textural measures were nominated to form AGB prediction models based on the existing literature.

Specific objective 2: “Extract the LiDAR metrics from airborne LiDAR data.”

What are the suitable LiDAR metrics for forest AGB estimation?

In addition to the spectral and textural variables, 52 LiDAR metrics (such as pit-free CHM, percentiles heights, canopy cover, canopy density, bincentiles, height metrics-minimum, maximum, average, standard deviation, skewness, kurtosis, quadratic average height, different point heights, and density strata) were also extracted based on the height and intensity information of the pulse returns using LAsTools (Section 4.2 and 5.2). Out of 52 metrics (Table 8), some new metrics such as bincentile, percentile and other height metrics (i.e., intensity based), relative height points count and relative height density were also generated to determine the relationship with in-situ AGB (Section 6.4).

What is the difference between the normal CHM and pit-free CHM?

Normal CHM and pit-free CHM were created using all first returns and all relevant returns respectively. The pit-free CHM could provide qualitative and seamless CHM (i.e., without irregularities in height variations) than normal CHM because of the spike-free TIN (Section 4.2.2. and Figure 28). Further, it was also found that the tree heights derived from pit-free CHM was more accurate ($R^2 = 0.97$, $RMSE = 1.07$ m and $RMSE_{rel} = 9.97\%$) than they derived from normal CHM ($R^2 = 0.94$, $RMSE = 1.42$ m and $RMSE_{rel} = 13.27\%$).

Specific objective 3: “Evaluate the efficiency of the SVM and RF for forest AGB estimation.”

Which ML regression algorithms perform the best?

The seven AGB prediction models (namely, (i) LiDAR metrics, (ii) spectral variables, (iii) textural variables, (iv) spectral and textural variables, (v) LiDAR and spectral variables, (vi) LiDAR and textural variables, and (vii) LiDAR, spectral, and textural variables) were tested using both SVM and RF ML regression. Both regression algorithms selected the LiDAR, spectral and textural variables combined model as the best model

with higher accuracy than others. However, the RF regression performed well with R^2 of 0.95, RMSE of 35.15 Mg ha⁻¹, RMSE_{rel} of 17.25 and RMSE_{CV} of 0.17 in compared SVM regression with R^2 of 0.40, RMSE of 48.29 Mg ha⁻¹, RMSE_{rel} of 23.70 and RMSE_{CV} of 0.24.

Initially, the convolutional neural network (CNN) was also proposed to estimate AGB in this study. However, it was not performed due to lack of sufficient data for CNN based modelling.

Specific objective 4: “Optimise the optical and LiDAR data derived variables for forest AGB estimation using the best-performing ML regression algorithm.”

How can the LiDAR, spectral and texture variables be optimised for AGB estimation using the ML regression algorithm?

There were 119 variables in the LiDAR, spectral and textural combined model that used for optimisation (Section 5.7 and Figure 22). They were given the score of importance with the help of RF regression using OBB sample data for ranking purpose in forms of %IncMSE (i.e., measuring the effect in the model if any particular variable is detached) and IncNodePurity (i.e., purity of node in the presence of particular variable) in R.

What is the optimal subset of predictors variables out of the optimised variables?

The 20 predictor variables were nominated as an optimal subset out of the 119 variables using RF regression with R^2 of 0.93 and RMSE of 35.46 Mg ha⁻¹ (Section 5.8 and Figure 23, 24 and 25). In the top 20 predictor variables, 1 red band, 9 texture measure such as variance, contrast, dissimilarity, homogeneity, second angular momentum and mean (from NIR, red-edge, red and blue) and 10 LiDAR metrics such as maximum, standard deviation, average, 99th percentile height, bincentiles (10th, 20th, 80th and 90th), height based relative height points count and intensity based relative height point counts were selected sequentially.

Specific objective 5: “Present the spatial distribution of forest AGB and map its uncertainty over the study area.”

What is the spatial distribution pattern of the predicted forest AGB over the study area?

The forest AGB spatial distribution map (Figure 26) was created using RF regression and the subset of predictor variables. The minimum and maximum forest AGB estimation were found 123.99 Mg ha⁻¹, 350.64 Mg ha⁻¹ respectively with the standard deviation of 37.74 Mg ha⁻¹. Also, the mean forest AGB was estimated 222.54 Mg ha⁻¹ which is slightly higher than the average value (i.e., 202.64 Mg ha⁻¹) determined by DFRS (2014) for the Terai region.

What is the accuracy of the forest AGB distribution over the study area?

The forest AGB distribution over the study area was validated with 15 independent plots with validation measures namely R^2 , RMSE, RMSE_{rel}, and RMSE_{CV} (Equations 4.1-4.4). In the validation result, it was observed that RF regression is capable to estimate forest AGB with an R^2 of 0.72, RMSE of 47.71 Mg ha⁻¹, RMSE_{rel} of 23.41 % and RMSE_{CV} of 0.23.

What is the range of uncertainty of the estimated forest AGB over the study area?

The uncertainty of the forest AGB distribution over the study area was mapped (Figure 30) using Monte Carlo simulation in R. The minimum to maximum range of uncertainty was estimated 0 to 34 Mg ha⁻¹. Also, it was observed that the mean uncertainty with reference to mean predicted AGB was 5.89 %.

Finally, it can be summarized that the integration and optimization of multi-sensor parameters using RF ML regression with in-situ data can able to predict the tropical biomass with satisfactory accuracy. Further, it is also found that spectral bands (i.e., NIR, red and red-edge), texture measure and LiDAR metrics such as bincentiles, relative height points count including other height metrics and percentile heights have a strong relationship with the field measured AGB.

7.2. Recommendations

The RF regression approach for combination and optimisation of LiDAR and high-resolution optical sensor variables with field measured data delivers a non-destructive, reliable, credible, efficient, effective and affordable way for the forest AGB estimation and its spatiotemporal change monitoring. The careful extension and application of this method in large-scale supports to improve the MRV system in Nepal for REDD+ implementation. However, this study suggests the following recommendation to do further research for the betterment and implementation of this approach at the sub-national level.

- It should be better to develop the local allometric equation for calculating field-measured biomass directly.
- High-resolution optical image with more bands and medium point density of LiDAR data should be preferable to reduce the uncertainty.
- The sufficient number of sample plots with stratification on the basis of forest density should be desirable to improve the sample representation of the forests and biomass estimation accuracy.

In Nepal, this research can potentially function as a reference study to improve the approach for the carbon estimation and its periodic monitoring for the National Forests Reference Level under the REDD+ implementation program.

LIST OF REFERENCES

- Adam, E., Mutanga, O., Abdel-Rahman, E. M., & Ismail, R. (2014). Estimating standing biomass in papyrus (*Cyperus papyrus* L.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, *35*(2), 693–714. <https://doi.org/10.1080/01431161.2013.870676>
- Andersson, K., Evans, T. P., & Richards, K. R. (2009). National forest carbon inventories: policy needs and assessment capacity. *Climatic Change*, *93*(1–2), 69–101. <https://doi.org/10.1007/s10584-008-9526-6>
- Baccini, A., Friedl, M. A., Woodcock, C. E., & Warbington, R. (2004). Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, *31*(10), n/a-n/a. <https://doi.org/10.1029/2004GL019782>
- Bali, R., Sarkar, D., Lantz, B., & Lesmeister, C. (2016). Black Box Methods - Neural Networks and Support Vector Machines. In R.: *unleash machine learning techniques* (pp. 535–568). Birmingham: Packt Publishing. Retrieved from https://books.google.co.in/books?id=3ZfcDgAAQBAJ&pg=PA564&lpq=PA564&dq=No+reliable+rule+for+matching+a+kernel+to+a+particular+learning+task+when+using+SVM+algorithm&source=bl&ots=lu5q_ZTi4&sig=ACfU3U2G3HcXxILA51fk7tRuQ81yqqMazg&hl=en&sa=X&ved=2ahUKEwiv9d
- Baral, S. (2011). *Mapping Carbon Stock Using High Resolution Satellite Images In Sub-Tropical Forest Of Nepal*. MSc Thesis. University of Twente, Enschede, The Netherlands. <https://doi.org/10.13140/RG.2.1.3996.9768>
- Baret, F., Guyot, G., & Major, D. J. (1989). TSAVI: A Vegetation Index Which Minimizes Soil Brightness Effects On LAI And APAR Estimation. In *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*, (Vol. 3, pp. 1355–1358). IEEE. <https://doi.org/10.1109/IGARSS.1989.576128>
- 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*. MSc Thesis. University of Twente, Enschede, The Netherlands. Retrieved from http://www.itc.nl/library/papers_2012/msc/nrm/lopezbautista.pdf
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Ben-Arie, J. R., Hay, G. J., Powers, R. P., Castilla, G., & St-Onge, B. (2009). Development of a pit filling algorithm for LiDAR canopy height models. *Computers & Geosciences*, *35*(9), 1940–1949. <https://doi.org/10.1016/j.cageo.2009.02.003>
- Breiman, L. (1994). Bagging Predictors. *Machine Learning*, *24*(2), 123–140. Retrieved from <http://www.cs.utsa.edu/~bylander/cs6243/breiman96bagging.pdf>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 1–32.
- Casas, Á., García, M., Siegel, R. B., Koltunov, A., Ramírez, C., & Ustin, S. (2016). Burned forest characterization at single-tree level with airborne laser scanning for assessing wildlife habitat. *Remote Sensing of Environment*, *175*, 231–241. <https://doi.org/10.1016/j.rse.2015.12.044>
- Cerbu, G. A., Swallow, B. M., & Thompson, D. Y. (2011). Locating REDD : A global survey and analysis of REDD readiness and demonstration activities. *Environmental Science and Policy*, *14*(2), 168–180. <https://doi.org/10.1016/j.envsci.2010.09.007>
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., ... Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190. <https://doi.org/10.1111/gcb.12629>
- Chen, J. M. (1996). Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing*, *22*(3), 229–242. <https://doi.org/10.1080/07038992.1996.10855178>
- Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). *Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data*. Retrieved from [https://nature.berkeley.edu/biometlab/pdf/Chen et al 2006 PERS.pdf](https://nature.berkeley.edu/biometlab/pdf/Chen%20et%20al%202006%20PERS.pdf)
- Chen Qu. (2013). *LiDAR Remote Sensing of Vegetation Biomass*. *Remote Sensing of Natural Resources* (Vol. 20135777). Boca Raton: FL: CRC Press and Taylor & Francis Group. <https://doi.org/10.1201/b15159>

- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297. Retrieved from http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf
- Coulston, J. W., Blinn, C. E., Thomas, V. A., & Wynne, R. H. (2016). Approximating Prediction Uncertainty for Random Forest Regression Models. *Photogrammetric Engineering & Remote Sensing*, 82(3), 189–197. <https://doi.org/10.14358/PERS.82.3.189>
- Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63(January), 22–33. <https://doi.org/10.1016/j.cageo.2013.10.008>
- Dang, A. T. N., Nandy, S., Srinet, R., Luong, N. V., Ghosh, S., & Senthil Kumar, A. (2019). Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam. *Ecological Informatics*, 50(July 2018), 24–32. <https://doi.org/10.1016/j.ecoinf.2018.12.010>
- Dang, N. anh. (2012). *Error Propagation in Carbon Estimation Using The Combination of Airborne LiDAR and Very High Resolution GEO-EYE Imagery in Ludhikhola Watershed, Gorkha, Nepal*. MSc Thesis. University of Twente, Enschede, The Netherlands.
- DDC. (2015). *Periodic District Development Plan*. Dhangadhi, Kailali. Retrieved from www.sdincnepal.org
- Deb, D., Singh, J. P., Deb, S., Datta, D., Ghosh, A., & Chaurasia, R. S. (2017). An alternative approach for estimating above ground biomass using Resourcesat-2 satellite data and artificial neural network in Bundelkhand region of India. *Environmental Monitoring and Assessment*, 189(11), 576. <https://doi.org/10.1007/s10661-017-6307-6>
- DFO. (2018). *Annual Progress Report*. Dhangadhi, Kailali.
- DFRS. (1999). *Forests Resources of Nepal*. Kathmandu.
- DFRS. (2014). *TERAI FORESTS of NEPAL*. Kathmandu. Retrieved from http://frtc.gov.np/downloadfile/The-TeraiForestsofNepal_Press-copy-final_1502865012.pdf
- DFRS. (2015). *STATE OF NEPAL'S FORESTS*. Kathmandu. Retrieved from www.dfrs.gov.np
- Dhanda, P., Nandy, S., Kushwaha, S., Ghosh, S., Murthy, Y. K., & Dadhwal, V. (2017). Optimizing spaceborne LiDAR and very high resolution optical sensor parameters for biomass estimation at ICESat/GLAS footprint level using regression algorithms. *Progress in Physical Geography*, 41(3), 247–267. <https://doi.org/10.1177/0309133317693443>
- Domingo, D., Lamelas-Gracia, M. T., Montealegre-Gracia, A. L., & de la Riva-Fernández, J. (2017). Comparison of regression models to estimate biomass losses and CO₂ emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *European Journal of Remote Sensing*, 50(1), 384–396. <https://doi.org/10.1080/22797254.2017.1336067>
- Dube, T., & Mutanga, O. (2015). Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 12–32. <https://doi.org/10.1016/j.isprsjprs.2015.06.002>
- Dube, T., Mutanga, O., Elhadi, A., & Ismail, R. (2014). Intra-and-Inter Species Biomass Prediction in a Plantation Forest: Testing the Utility of High Spatial Resolution Spaceborne Multispectral RapidEye Sensor and Advanced Machine Learning Algorithms. *Sensors*, 14(8), 15348–15370. <https://doi.org/10.3390/s140815348>
- FAO. (2009). *The future of forests in Asia and the Pacific : outlook for 2020 : 16-18 October 2007, Chiang Mai, Thailand*. Bangkok. Retrieved from <http://www.fao.org/docrep/011/i0627e/I0627E00.htm#Contents>
- FAO. (2010). *Global Forest Resources Assessment 2010*. Rome: FAO. [https://doi.org/ISBN 978-92-5-106654-6](https://doi.org/ISBN%20978-92-5-106654-6)
- FAO. (2015). *FAO assessment of forests and carbon stocks, 1990-2015*. Rome. Retrieved from <http://www.fao.org/forestry/fra/67090/en/>
- Feng, Y., Lu, D., Chen, Q., Keller, M., Moran, E., Dos-Santos, M. N., ... Batistella, M. (2017). Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. *International Journal of Digital Earth*, 10(10), 996–1016. <https://doi.org/10.1080/17538947.2017.1301581>
- García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., & Riquelme, J. C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, 167, 24–31. <https://doi.org/10.1016/j.neucom.2014.09.091>
- Gaveau, D. L. A., & Hill, R. A. (2003). *Quantifying canopy height underestimation by laser pulse penetration in small-*

- footprint airborne laser scanning data*. Retrieved from <https://www.tandfonline.com/doi/pdf/10.5589/m03-023?needAccess=true>
- Genuer, R., Poggi, J., & Tuleau-malot, C. (2010). Variable selection using Random Forests. *Pattern Recognition Letters*, 31(14), 2225–2236.
- Ghosh, S. M., & Behera, M. D. (2018). Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. *Applied Geography*, 96, 29–40. <https://doi.org/10.1016/j.apgeog.2018.05.011>
- Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2(4), 045023. <https://doi.org/10.1088/1748-9326/2/4/045023>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Gitelson, A. A. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32(8), L08403. <https://doi.org/10.1029/2005GL022688>
- Gitelson, A. A., & Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, 22(5), 689–692. [https://doi.org/10.1016/S0273-1177\(97\)01133-2](https://doi.org/10.1016/S0273-1177(97)01133-2)
- Gitelson, A., & Merzlyak, M. N. (1994). *Journal of plant physiology*. *Journal of Plant Physiology* (Vol. 143). Urban & Fischer. Retrieved from <https://nebraska.pure.elsevier.com/en/publications/spectral-reflectance-changes-associated-with-autumn-senescence-of>
- Gleason, C. J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, 125, 80–91. <https://doi.org/10.1016/j.rse.2012.07.006>
- Goodale, C. L., Apps, M. J., Birdsey, R. A., Field, C. B., Heath, L. S., Houghton, R. A., ... Shvidenko, A. Z. (2002). Forest carbon sinks in the Northern Hemisphere. *Ecological Applications*, 12(3), 891–899. <https://doi.org/10.1890/1051-0761>
- Haboudane, D. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. <https://doi.org/10.1016/j.rse.2003.12.013>
- Hajar, N. Z., Mohd Shukri, W. W., Danial, M. M., Shahid, M. M., & Nizam, M. A. (2015). *Forest Biomass Monitoring for Redd+ Manual* (1st ed.). Kepong, Selangor, Malaysia: Forest Research Institute Malaysia. Retrieved from http://redd.ffpri.affrc.go.jp/events/seminars/_img/_20150211/Manual_Biomass.pdf
- Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*, 225(1–3), 378–390. <https://doi.org/10.1016/j.foreco.2006.01.014>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- He, Q.-S., Cao, C.-X., Chen, E.-X., Sun, G.-Q., Ling, F.-L., Pang, Y., ... Li, X.-W. (2012). Forest stand biomass estimation using ALOS PALSAR data based on LiDAR-derived prior knowledge in the Qilian Mountain, western China. *International Journal of Remote Sensing*, 33(3), 710–729. <https://doi.org/10.1080/01431161.2011.577829>
- Henry, M., Besnard, A., Asante, W. A., Eshun, J., Adu-Bredu, S., Valentini, R., ... Saint-André, L. (2010). Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. *Forest Ecology and Management*, 260, 1375–1388. <https://doi.org/10.1016/j.foreco.2010.07.040>
- Houghton, R. A. (2005). Aboveground Forest Biomass and the Global Carbon Balance. *Global Change Biology*, 11(6), 945–958. <https://doi.org/10.1111/j.1365-2486.2005.00955.x>
- Houghton, R. A., Hall, F., & Goetz, S. J. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research: Biogeosciences*, 114(G2), n/a-n/a. <https://doi.org/10.1029/2009JG000935>
- Hussin, Y. A., Gilani, H., van Leeuwen, L., Murthy, M. S. R., Shah, R., Baral, S., ... Qamer, F. M. (2014). Evaluation of object-based image analysis techniques on very high-resolution satellite image for biomass estimation in a watershed of hilly forest of Nepal. *Applied Geomatics*, 6(1), 59–68. <https://doi.org/10.1007/s12518-014-0126-z>
- IPCC. (2007). *Climate Change 2007: Synthesis Report*. Geneva. Retrieved from https://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_full_report.pdf
- Isenburg, M. (2016). Generating Spike-Free Digital Surface Models from LiDAR. Retrieved January 21,

- 2019, from <https://rapidlasso.com/2016/02/03/generating-spike-free-digital-surface-models-from-lidar/>
- Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2003). National-scale biomass estimators for United States tree species. *Forest Science*, *49*(1), 12–35. <https://doi.org/10.1093/forestsience/49.1.12>
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, *50*(4), 663–666. <https://doi.org/10.2307/1936256>
- Karna, Y. K., Hussin, Y. A., Gilani, H., Bronsveld, M. C., Murthy, M. S. R., Qamer, F. M., ... Baniya, C. B. (2015). Integration of WorldView-2 and airborne LiDAR data for tree species level carbon stock mapping in Kayar Khola watershed, Nepal. *International Journal of Applied Earth Observation and Geoinformation*, *38*, 280–291. <https://doi.org/10.1016/j.jag.2015.01.011>
- Kauranne, T., Joshi, A., Gautam, B., Manandhar, U., Nepal, S., Peuhkurinen, J., ... Leppänen, V. (2017). LiDAR-Assisted Multi-Source Program (LAMP) for Measuring Above Ground Biomass and Forest Carbon. *Remote Sensing*, *9*(2), 154. <https://doi.org/10.3390/rs9020154>
- Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T., & Hussin, Y. A. (2013). Development of an algorithm to generate a Lidar pit-free canopy height model. *SilviLaser*, 125–128. Retrieved from <http://www.riegl.com/media-events/projects/airborne-scanning/project/generating-pit-free-canopy-height-models-from-lidar/>
- Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T., & Hussin, Y. A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. *Photogrammetric Engineering & Remote Sensing*, *80*(9), 863–872. <https://doi.org/10.14358/PERS.80.9.863>
- Kim, Y., Yang, Z., Cohen, W. B., Pflugmacher, D., Lauver, C. L., & Vankat, J. L. (2009). Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sensing of Environment*, *113*(11), 2499–2510. <https://doi.org/10.1016/j.rse.2009.07.010>
- Kindermann, G., McCallum, I., Fritz, S., & Obersteiner, M. (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, *42*(3), 387–396. <https://doi.org/10.14214/sf.244>
- Kumar, L., & Mutanga, O. (2017). Remote Sensing of Above-Ground Biomass. *Remote Sensing*, *9*(9), 935. <https://doi.org/10.3390/rs9090935>
- Kumar, L., Sinha, P., Taylor, S., & Alqurashi, A. F. (2015). Review of the use of remote sensing for biomass estimation to support renewable energy generation. *Journal of Applied Remote Sensing*, *9*(1), 097696. <https://doi.org/10.1117/1.JRS.9.097696>
- Kumar, R., Nandy, S., Agarwal, R., & Kushwaha, S. P. S. (2014). Forest cover dynamics analysis and prediction modeling using logistic regression model. *Ecological Indicators*, *45*, 444–455. <https://doi.org/10.1016/j.ecolind.2014.05.003>
- Kushwaha, S. P. S., & Nandy, S. (2012). Species diversity and community structure in sal (*Shorea robusta*) forests of two different rainfall regimes in West Bengal, India. *Biodiversity and Conservation*, *21*(5), 1215–1228. <https://doi.org/10.1007/s10531-012-0264-8>
- Kushwaha, S. P. S., Nandy, S., & Gupta, M. (2014). Growing stock and woody biomass assessment in Asola-Bhatti Wildlife Sanctuary, Delhi, India. *Environmental Monitoring and Assessment*, *186*(9), 5911–5920. <https://doi.org/10.1007/s10661-014-3828-0>
- Lehtonen, A., Mäkipää, R., Heikkinen, J., Sievänen, R., & Liski, J. (2004). Biomass expansion factors (BEFs) for Scots pine, Norway spruce and birch according to stand age for boreal forests. *Forest Ecology and Management*, *188*(1–3), 211–224. <https://doi.org/10.1016/j.foreco.2003.07.008>
- Li, A., Dhakal, S., Glenn, N., Spaete, L., Shinneman, D., Pilliod, D., ... McIlroy, S. (2017). Lidar Aboveground Vegetation Biomass Estimates in Shrublands: Prediction, Uncertainties and Application to Coarser Scales. *Remote Sensing*, *9*(9), 903. <https://doi.org/10.3390/rs9090903>
- Liu, K., Wang, J., Zeng, W., & Song, J. (2017). Comparison and Evaluation of Three Methods for Estimating Forest above Ground Biomass Using TM and GLAS Data. *Remote Sensing*, *9*(4), 341. <https://doi.org/10.3390/rs9040341>
- Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, *27*(7), 1297–1328. <https://doi.org/10.1080/01431160500486732>
- Lu, D. (2007). International Journal of Remote Sensing The potential and challenge of remote sensing-based biomass estimation The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, *27*(7), 1297–1328. <https://doi.org/10.1080/01431160500486732>
- Lu, D., Batistella, M., & Moran, E. (2005). Satellite Estimation of Aboveground Biomass and Impacts of

- Forest Stand Structure. *Photogrammetric Engineering & Remote Sensing*, 71(8), 967–974.
<https://doi.org/10.14358/PERS.71.8.967>
- Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, 9(1), 63–105. <https://doi.org/10.1080/17538947.2014.990526>
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., ... Saah, D. (2012a). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012(1), 1–16.
<https://doi.org/10.1155/2012/436537>
- Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., ... Saah, D. (2012b). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research*, 2012(April), 1–16.
<https://doi.org/10.1155/2012/436537>
- Lu, D., & Lu, D. (2006). The potential and challenge of remote sensing - based biomass estimation, 27(7), 1297–1328. <https://doi.org/10.1080/01431160500486732>
- Maharjan, S. (2012). *Estimation and Mapping Above Ground Woody Carbon Stocks Using Lidar Data And Digital Camera Imagery In The Hilly Forest Of Gorkha, Nepal*. MSc Thesis. University of Twente, Enschede, The Netherlands.
- Mbaabu, P. R., Hussin, Y. A., Weir, M., & Gilani, H. (2014). Quantification of carbon stock to understand two different forest management regimes in Kayar Khola watershed, Chitwan, Nepal. *Journal of the Indian Society of Remote Sensing*, 42(4), 745–754. <https://doi.org/10.1007/s12524-014-0379-3>
- McRoberts, R. E., Næsset, E., & Gobakken, T. (2013). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment*, 128, 268–275.
<https://doi.org/10.1016/j.rse.2012.10.007>
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., Solovchenko, A. E., & Pogosyan, S. I. (2003). Application of Reflectance Spectroscopy for Analysis of Higher Plant Pigments. *Russian Journal of Plant Physiology*, 50(5), 704–710. <https://doi.org/10.1023/A:1025608728405>
- Mitchard, E. T. A., Saatchi, S. S., White, L. J. T., Abernethy, K. A., Jeffery, K. J., Lewis, S. L., ... Meir, P. (2012). Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences*, 9(1), 179–191.
<https://doi.org/10.5194/bg-9-179-2012>
- MoFE. (2018). *Nepal National REDD+ Strategy*. Kathmandu. Retrieved from
http://redd.gov.np/public/upload/office-id-2/file/REDD__Strategy_Nepal_2018.pdf
- MoFSC. (1989). *Master Plan for the Forestry Sector (Main Report)*. Kathmandu.
- MoFSC. (2015). *Strategy and Action Plan 2015-2025 Terai Arc Landscape, Nepal*. Kathmandu.
- Mohd Zaki, N. A., & Abd Latif, Z. (2017). Carbon sinks and tropical forest biomass estimation: a review on role of remote sensing in aboveground-biomass modelling. *Geocarto International*, 32(7), 701–716.
<https://doi.org/10.1080/10106049.2016.1178814>
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259.
<https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Mutanga, O., Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18, 399–406. <https://doi.org/10.1016/j.jag.2012.03.012>
- Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112(6), 3079–3090.
<https://doi.org/10.1016/j.rse.2008.03.004>
- Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S. P. S., Kumar, A. S., & Dadhwal, V. K. (2017). Neural network-based modelling for forest biomass assessment. *Carbon Management*, 8(4), 305–317.
<https://doi.org/10.1080/17583004.2017.1357402>
- NASA. (2018). Global Climate Change. Retrieved November 26, 2018, from
<https://climate.nasa.gov/causes/>
- Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., ... Tommaselli, A. (2017). Individual Tree Detection and Classification with UAV-Based Photogrammetric Point Clouds and Hyperspectral Imaging. *Remote Sensing*, 9(3), 185. <https://doi.org/10.3390/rs9030185>
- Omar, H., O., N. A., A.D, N. D., F., M. A., & K, A. R. (2015). AIRBORNE LIDAR FOR ESTIMATING ABOVEGROUND BIOMASS IN DIPTEROCARP FORESTS OF MALAYSIA.

Asian Conference on Remote Sensing, 1(February 2016), 1–10.

- Pandit, S., Tsuyuki, S., & Dube, T. (2018a). Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data. *Remote Sensing*, 10(4), 601. <https://doi.org/10.3390/rs10040601>
- Pandit, S., Tsuyuki, S., & Dube, T. (2018b). Landscape-Scale Aboveground Biomass Estimation in Buffer Zone Community Forests of Central Nepal: Coupling In Situ Measurements with Landsat 8 Satellite Data. *Remote Sensing*, 10(11), 1848. <https://doi.org/10.3390/rs10111848>
- PCTMCDB. (2015). Introduction of Chure. Retrieved February 1, 2019, from <http://chureboard.gov.np/en/>
- Pham, L. T. H., & Brabyn, L. (2017). Monitoring mangrove biomass change in Vietnam using SPOT images and an object-based approach combined with machine learning algorithms. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, 86–97. <https://doi.org/10.1016/j.isprsjprs.2017.03.013>
- Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3–4), 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>
- Rodríguez-Veiga, P., Saatchi, S., Tansey, K., & Balzter, H. (2016). Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. *Remote Sensing of Environment*, 183, 265–281. <https://doi.org/10.1016/j.rse.2016.06.004>
- Roujean, J.-L., & Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51(3), 375–384. [https://doi.org/10.1016/0034-4257\(94\)00114-3](https://doi.org/10.1016/0034-4257(94)00114-3)
- Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. In *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A* (pp. 1–9). Texas. Retrieved from <https://ntrs.nasa.gov/search.jsp?R=19740022614>
- Saini, R., & Ghosh, S. K. (2017). Ensemble classifiers in remote sensing: A review. In *2017 International Conference on Computing, Communication and Automation (ICCCA)* (Vol. 2017–Janua, pp. 1148–1152). IEEE. <https://doi.org/10.1109/CCAA.2017.8229969>
- Segura, M., & Kanninen, M. (2005). Allometric models for tree volume and total aboveground biomass in a tropical humid forest in Costa Rica. *Biotropica*, 37(1), 2–8. <https://doi.org/10.1111/j.1744-7429.2005.02027.x>
- Shamsoddini, A., Turner, R., & Trinder, J. C. (2013). Improving lidar-based forest structure mapping with crown-level pit removal. *Journal of Spatial Science*, 58(1), 29–51. <https://doi.org/10.1080/14498596.2012.759092>
- Sharma, E. R., & Pukkala, T. (1990). *Volume Equations and Biomass Prediction of Forest Trees of Nepal* (Publication No. 47). Kathmandu.
- Shrestha, S. K. (2011). *Carbon Stock Estimation Using Very High Resolution satellite Imagery and Individual Crown Segmentation*. MSc Thesis. University of Twente, Enschede, The Netherlands. Retrieved from http://www.itc.nl/library/papers_2011/msc/nrm/shrestha.pdf
- Sibanda, M., Mutanga, O., & Rouget, M. (2015). Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. <https://doi.org/10.1016/j.isprsjprs.2015.10.005>
- Sinha, S., Jeganathan, C., Sharma, L. K., & Nathawat, M. S. (2015). A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology*, 12(5), 1779–1792. <https://doi.org/10.1007/s13762-015-0750-0>
- Suresh, M., Kiran Chand, T. R., Fararoda, R., Jha, C. S., & Dadhwal, V. K. (2014). Forest above ground biomass estimation and forest/non-forest classification for Odisha, India, using L-band Synthetic Aperture Radar (SAR) data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-8(8), 651–658. <https://doi.org/10.5194/isprsjprs-XL-8-651-2014>
- Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., van der Meer, B., & Enclona, E. (2004). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing*, 25(23), 5447–5472. <https://doi.org/10.1080/01431160412331291279>
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- UNFCCC. (2008). *Kyoto Protocol Reference Manual*. Bonn: United Nations Framework Convention on Climate Change All rights reserved This. Retrieved from https://unfccc.int/resource/docs/publications/08_unfccc_kp_ref_manual.pdf

- UNFCCC. (2009). *Report of the Conference of the Parties on its fifteenth session*. Copenhagen: United Nations United Nations Framework Convention on Climate Change. Retrieved from <https://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf>
- Van Der Meer, F., Bakker, W., Scholte, K., Skidmore, A., De Jong, S., Clevers, J., & Epema, G. (2000). VEGETATION INDICES, ABOVE GROUND BIOMASS ESTIMATES AND THE RED EDGE FROM MERIS. *International Archives of Photogrammetry and Remote Sensing*, XXXIII(Part B7), 1580–1587. Retrieved from http://www.isprs.org/proceedings/XXXIII/congress/part7/1580_XXXIII-part7.pdf
- Vosselman, G., & Maas, H.-G. (2010). *Airborne and terrestrial laser scanning* (1st ed.). Dunbeath: Whittles Publishing.
- Walker, W. S., Baccini, A., Nepstad, M., Horning, N., Knight, D., Braun, E., ... Bausch, A. (2011). *Field Guide for Forest Biomass and Carbon Estimation* (1st ed.). Falmouth: Woods Hole Research Center. Retrieved from www.whrc.org
- Wing, B. M., Ritchie, M. W., Boston, K., Cohen, W. B., & Olsen, M. J. (2015). Individual snag detection using neighborhood attribute filtered airborne lidar data. *Remote Sensing of Environment*, 163, 165–179. <https://doi.org/10.1016/j.rse.2015.03.013>
- Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148(8–9), 1230–1241. <https://doi.org/10.1016/j.agrformet.2008.03.005>
- Wu, C., Shen, H., Shen, A., Deng, J., Gan, M., Zhu, J., ... Wang, K. (2016). Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. *Journal of Applied Remote Sensing*, 10(3), 035010. <https://doi.org/10.1117/1.JRS.10.035010>
- Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., ... Ye, H. (2018). Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval, 11(5), 1482–1493.
- Yadav, B. K. V., & Nandy, S. (2015). Mapping aboveground woody biomass using forest inventory, remote sensing and geostatistical techniques. *Environmental Monitoring and Assessment*, 187(5), 308. <https://doi.org/10.1007/s10661-015-4551-1>
- Yao, W., Krzystek, P., & Heurich, M. (2012). Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. <https://doi.org/10.1016/j.rse.2012.03.027>
- Zhao, K., Popescu, S., & Nelson, R. (2009a). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, 113(1), 182–196. <https://doi.org/10.1016/j.rse.2008.09.009>
- Zhao, K., Popescu, S., & Nelson, R. (2009b). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, 113(1), 182–196. <https://doi.org/10.1016/j.rse.2008.09.009>
- Zianis, D., Muukkonen, P., Mäkipää, R., & Mencuccini, M. (2005). Biomass and stem volume equations for tree species in Europe. *Silva Fennica Monographs*, 4, 1–63. <https://doi.org/citeulike-article-id:11858948>

APPENDICES

Appendix 1: Specification of RapidEye system

Characteristics	Information												
Spectral bands	Capable of capturing any of the following spectral bands: <table border="1"> <thead> <tr> <th>Name</th> <th>Spectral Bands (nm)</th> </tr> </thead> <tbody> <tr> <td>Blue</td> <td>440-510</td> </tr> <tr> <td>Green</td> <td>520-590</td> </tr> <tr> <td>Red</td> <td>630-685</td> </tr> <tr> <td>Red Edge</td> <td>690-730</td> </tr> <tr> <td>NIR</td> <td>760-850</td> </tr> </tbody> </table>	Name	Spectral Bands (nm)	Blue	440-510	Green	520-590	Red	630-685	Red Edge	690-730	NIR	760-850
Name	Spectral Bands (nm)												
Blue	440-510												
Green	520-590												
Red	630-685												
Red Edge	690-730												
NIR	760-850												
Satellites number	5												
Spacecraft lifetime	7 years												
Orbit altitude	630 km in Sun-synchronous orbit												
Equator crossing time	11:00 am (approximately)												
Sensor Type	Multi-spectral push broom imager												
Ground sampling distance (nadir)	6.5 meter												
Pixel size	5 meter												
Swath Width	77 km												
On board data storage	1500 km of image data per orbit												
Revisit time	Daily (off-nadir)/ 5.5 days (at nadir)												
Image capture capacity	4 million sq km/day												
Dynamic Range	Up to 12 bit												
Image angle	Less than 20 degree off-nadir												
Tile size	25 km *25 km												
Image acquisition date	03-01-2010 to 30-04-2010 and 11-06-2010 to 04-24-2011												
Projected datum	WGS84 UTM												

Appendix 2: Used commands script of quality checking, data preparation, and LiDAR metrics extraction using airborne LiDAR data

Commands script for the quality checking

- 1) lasinfo -cpu64 -v -lof file_list.11216.txt -cores 12 -otxt -odir "F:\LiDAR_Data_Work\Las_Info_Report" -odix "_info" -cd -repair_bb -repair_counters -histo z 5 -histo intensity 64
- 2) lasview -v -lof file_list.14248.txt -cores 12
- 3) lasboundary -v -lof file_list.11860.txt -merged -use_bb -overview -labels -odir "F:\LiDAR_Data_Work\BF_Lasboundary_Merged" -o "BF_boundary.shp"
- 4) lasgrid -cpu64 -v -lof file_list.9316.txt -merged -point_density -false -set_min_max 1 2 -odir "F:\LiDAR_Data_Work\Qlas_grid" -o "density_ppm.tif"
- 5) lasduplicate -cpu64 -lof file_list.19032.txt -olaz
- 6) lasoverlap -lof file_list.11736.txt -merged -faf -elevation -lowest -opng

Data preparation and creating Pit-free CHM

- 7) lastile -cpu64 -v -lof file_list.9500.txt -o "tile.laz" -tile_size 1000 -buffer 25 -flag_as_withheld -cores 11 -odir "F:\LiDAR_Data_Work\Tiles_BlockF" -olaz
- 8) lasnoise -cpu64 -v -lof file_list.15232.txt -cores 12 -step_xy 4 -step_z 1 -odir "F:\LiDAR_Data_Work\BF_Tiles_Denoised" -olaz

- 9) lasground -cpu64 -v -lof file_list.1768.txt -cores 12 -wilderness -odir "F:\LiDAR_Data_Work\BF_Tiles_Denoised_Lasground" -odix "_G" -olaz
- 10) lasheight -cpu64 -lof file_list.12348.txt -cores 12 -replace_z -drop_above 45 -odir "F:\LiDAR_Data_Work\BF_Lasheight" -olaz
- 11) lastile -cpu64 -v -lof file_list.11252.txt -cores 12 -remove_buffer -odir "F:\LiDAR_Data_Work\Final_Tiles_Denoised_Lasground" -olaz
- 12) lasboundary -v -lof file_list.15572.txt -cores 12 -use_bb -overview -labels -odir "F:\LiDAR_Data_Work\BF_Lasboundary" -o "Tiles_boundary.shp"
- 13) blast2dem -v -lof file_list.9824.txt -merged -keep_classification 2 -elevation -use_tile_bb -odir "F:\LiDAR_Data_Work\DEM_BF" -o "DTM.tif"
- 14) blast2dem -v -lof file_list.9824.txt -merged -keep_classification 2 -step 0.5 -elevation -odir "F:\LiDAR_Data_Work\DEM_BF" -o "DTM.tif"
- 15) blast2dem -v -lof file_list.9824.txt -merged -keep_classification 2 -step 0.5 -elevation -hillshade -odir "F:\LiDAR_Data_Work\DTM_BF\DTM_Hillshade" -o "DTM.tif" -utm 44north
- 16) blast2dem -i (for pit-free CHM):
 - Height_normalized (-lof file_list.9824.txt)
 - keep_first
 - drop_z_below 2
 - step 0.5
 - kill 1.3
 - odir "F:\LiDAR_Data_Work\Pit_Free_CHM" -odix "_2" -obill
 (Similarly, partial CHMs are created from 5-40 m in regular interval of 5 m)
 - lasgrid
 - lof file_list.9824.txt
 - merged
 - strp 0.5
 - elevation
 - highest
 - o "Pit_free_CHM.tif"

Other LiDAR metrics extraction (using following command scripts with help of lascanopy model of LAStools)

- 17) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -cov -cover_cutoff 8 -height_cutoff 1.3 -odir "F:\Lascanopy_Metrics\Cannopy_Cover" -oasc
- 18) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -cov -gap -cover_cutoff 8 -height_cutoff 1.3 -odir "F:\Lascanopy_Metrics\Cannopy_Cover_Gap" -oasc
- 19) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -dns -cover_cutoff 8 -height_cutoff 1.3 -odir "F:\Lascanopy_Metrics\Cannopy_Density" -oasc
- 20) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -min -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_Minimum" -oasc
- 21) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -max -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_Maximum" -oasc
- 22) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -avg -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_Average" -oasc
- 23) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -std -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_StandardDeviation" -oasc
- 24) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -ske -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_Skweness" -oasc
- 25) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -kur -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_Kurtosis" -oasc

- 26) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -qav -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_QuadraticAverage" -oasc
- 27) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -all -odir "F:\Lascanopy_Metrics\Height_Metrics\Ht_All" -oasc
- 28) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 1 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P01" -oasc
- 29) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 5 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P05" -oasc
- 30) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 10 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P10" -oasc
- 31) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 25 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P25" -oasc
- 32) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 50 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P50" -oasc
- 33) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 75 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P75" -oasc
- 34) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 90 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P90" -oasc
- 35) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 95 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P95" -oasc
- 36) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -p 99 -odir "F:\Lascanopy_Metrics\Ht_Percentiles\Ht_P99" -oasc
- 37) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 10 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B10" -oasc
- 38) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 20 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B20" -oasc
- 39) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 30 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B30" -oasc
- 40) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 40 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B40" -oasc
- 41) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 50 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B50" -oasc
- 42) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 60 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B60" -oasc
- 43) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 70 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B70" -oasc
- 44) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 80 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B80" -oasc
- 45) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -b 90 -odir "F:\Lascanopy_Metrics\Ht_Bincentiles\Ht_B90" -oasc
- 46) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -c 0.5 15 -odir "F:\Lascanopy_Metrics\RHt_Strata_Counts\RHt_Point_Counts_0.5-15" -oasc
- 47) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -c 15 30 -odir "F:\Lascanopy_Metrics\RHt_Strata_Counts\RHt_Point_Counts_15-30" -oasc
- 48) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -c 30 45 -odir "F:\Lascanopy_Metrics\RHt_Strata_Counts\RHt_Point_Counts_30-45" -oasc
- 49) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -d 0.5 15 -odir "F:\Lascanopy_Metrics\RHt_Strata_Densities\RHt_Point_Densities_0.5-15" -oasc
- 50) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -d 15 30 -odir "F:\Lascanopy_Metrics\RHt_Strata_Densities\RHt_Point_Densities_15-30" -oasc

- 51) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -d 30 45 -
odir "F:\Lascanopy_Metrics\RHt_Strata_Densities\RHt_Point_Densities_30-45" -oasc
- 52) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_min -
odir "F:\Lascanopy_Metrics\Int_Metrics\Int_Minimum" -oasc
- 53) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_max -
odir "F:\Lascanopy_Metrics\Int_Metrics\Int_Maximum" -oasc
- 54) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_avg -odir
"F:\Lascanopy_Metrics\Int_Metrics\Int_Average" -oasc
- 55) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_std -odir
"F:\Lascanopy_Metrics\Int_Metrics\Int_StandardDeviation" -oasc
- 56) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_ske -odir
"F:\Lascanopy_Metrics\Int_Metrics\Int_Skweness" -oasc
- 57) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_kur -odir
"F:\Lascanopy_Metrics\Int_Metrics\Int_Kurtosis" -oasc
- 58) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_qav -
odir "F:\Lascanopy_Metrics\Int_Metrics\Int_Quadratic_Average" -oasc
- 59) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 1 -odir
"F:\Lascanopy_Metrics\Int_Percentiles\Int_P01" -oasc
- 60) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 5 -odir
"F:\Lascanopy_Metrics\Int_Percentiles\Int_P05" -oasc
- 61) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 10 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P10" -oasc
- 62) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 25 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P25" -oasc
- 63) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 50 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P50" -oasc
- 64) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 75 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P75" -oasc
- 65) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 90 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P90" -oasc
- 66) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 95 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P95" -oasc
- 67) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_p 99 -
odir "F:\Lascanopy_Metrics\Int_Percentiles\Int_P99" -oasc
- 68) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_c 0.5 15
-odir "F:\Lascanopy_Metrics\Int_Strata_Counts\RHt_Int_Counts_0.5-15" -oasc
- 69) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_c 15 30 -
odir "F:\Lascanopy_Metrics\Int_Strata_Counts\RHt_Int_Counts_15-30" -oasc
- 70) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_c 30 45 -
odir "F:\Lascanopy_Metrics\Int_Strata_Counts\RHt_Int_Counts_30-45" -oasc
- 71) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_d 0.5 15
-odir "F:\Lascanopy_Metrics\Int_Strata_Densities\RHt_Int_Densities_0.5-15" -oasc
- 72) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_d 15 30 -
odir "F:\Lascanopy_Metrics\Int_Strata_Densities\RHt_Int_Densities_15-30" -oasc
- 73) lascanopy -v -lof file_list.12476.txt -drop_classification 7 -step 25 -height_cutoff 1.3 -int_d 30 45 -
odir "F:\Lascanopy_Metrics\Int_Strata_Densities\RHt_Int_Densities_30-45" -oasc

Appendix 3: The generated lasinfo report for the basic information and quality checking of airborne LiDAR point cloud data

Lasinfo (180919) report for 66 merged files	
Reporting all LAS header entries:	
File signature	'LASF'
File source ID	0
Global_encoding	0
Project ID GUID data 1-4	00000000-0000-0000-0000-000000000000
Version major.minor	1.2
System identifier	''
Generating software	'TerraScan'
File creation day/year	314/2012
Header size	227
Offset to point data	229
Number var. length records	0
Point data format	3
Point data record length	34
Number of point records	70354696
Number of points by return	50747390 19607306 0 0 0
Scale factor x y z	0.01 0.01 0.01
Offset x y z	0 0 0
Min x y z	487500.38 3159501.32 -3.77
The header is followed by 2 user-defined bytes	
Reporting minimum and maximum for all LAS point record entries ...	
X	48750038 49850000
Y	315950132 316508538
Z	-377 4402
Intensity	0 255
Return_number	1 2
Number_of_returns	1 4
Edge_of_flight_line	0 0
Scan_direction_flag	0 0
Classification	1 2
Scan_angle_rank	0 0
User_data	0 0
Point_source_ID	18 1530
GPS_time	-15436977.235369 -15434911.299490
WARNING: range violates GPS week time specified by global encoding bit 0	
Color	R 65280 65280 G 65280 65280 B 65280 65280
Number of first returns	50747390
Number of intermediate returns	2906347
Number of last returns	53575728
Number of single returns	36874769
Covered area in square units/kilounits	54957428/54.96

Point density	all returns 1.28 last only 0.97 (per square units)
Spacing	all returns 0.88 last only 1.01 (in units)
Overview over number of returns of given pulse	36874769 29297684 4112574 69669 0 0 0
Histogram of classification of point	33532732 unclassified (1) 36821964 ground (2)

Appendix 4: List of the species and their model coefficient.

SN	Genus	Species	Local name	Comments	a	b	c
1	Abies	<i>Abies pindrow</i>	Gobre salla	n.a.	-2.4453	1.722	1.0757
2	Acacia	<i>Acacia catechu</i>	Khair	n.a.	-2.3256	1.6476	1.0552
3	Adina	<i>Adina cordifolia</i>	Haldu	n.a.	-2.5626	1.8598	0.8783
4	Albizzia	spp	Sisris	n.a.	-2.4284	1.7609	0.9662
5	Alnus	<i>Alnus nepalensis</i>	Utis	n.a.	-2.7761	1.9006	0.9428
6	Anogeissus	<i>Anogeissus latifolia</i>	Banjhi	n.a.	-2.272	1.7499	0.9174
7	Bombax	<i>Bombax malabaricum</i>	n.a.	n.a.	-2.3865	1.7414	1.0063
8	Cedrela	<i>Cedrela toona</i>	Toon	n.a.	-2.1832	1.8679	0.7569
9	Dalbergia	<i>Dalbergia sissoo</i>	Sissoo	n.a.	-2.1959	1.6567	0.9899
10	Eugenia	<i>Eugenia jambolana</i>	Jamun	n.a.	-2.5693	1.8816	0.8498
11	Hymanodictyon	<i>Hymanodictyon excelsum</i>	Bhurkul	n.a.	-2.585	1.9437	0.7902
12	Lagerstroemia	<i>Lagerstroemia parviflora</i>	Botdhairo	n.a.	-2.3411	1.7246	0.9702
13	Michelia	<i>Michelia champaca</i>	Champ	n.a.	-2.0152	1.8555	0.763
14	Pinus	<i>Pinus roxburghii</i>	Chir Pine	n.a.	-2.977	1.9235	1.0019
15	Pinus	<i>Pinus wallichiana</i>	Blue Pine	n.a.	-2.8195	1.725	1.1623
16	Quercus	spp	n.a.	n.a.	-2.36	1.968	0.7469
17	Schima	<i>Schima wallichii</i>	Chilaune	n.a.	-2.7385	1.8155	1.0072
18	Shorea	<i>Shorea robusta</i>	Sal	n.a.	-2.4554	1.9026	0.8352
19	Terminalia	<i>Terminalia tomentosa</i>	Asna	n.a.	-2.4616	1.8497	0.88
20	Trewia	<i>Trewia nudiflora</i>	Gutel	n.a.	-2.4585	1.8043	0.922
21	Tsuga	spp	n.a.	Hemlock	-2.5293	1.7815	1.0369
22	Misc. 1	n.a.	n.a.	Miscellaneous species in Terai region	-2.3993	1.7836	0.9546
23	Misc. 2	n.a.	n.a.	Miscellaneous species in Siwaliks region	-2.3204	1.8507	0.8223

Appendix 5: List of the species and their wood density

Genus	Species	Local name	Density, kg/m ³
Abies	<i>Abies pindrow</i>	Gobre salla	480
Abies	<i>Abies spectabilis</i>	Bunga salla	480
Abies	spp	n.a.	480
Acacia	<i>Acacia catechu</i>	Khair	960
Acer	spp	n.a.	640
Adina	<i>Adina cordifolia</i>	Haldu	670
Albizzia	spp	Sisris	673
Alnus	<i>Alnus nepalensis</i>	Utis	390
Anogeissus	<i>Anogeissus latifolia</i>	Banjhi	900
Betula	<i>Betula utilis</i>	Bhojpatra	700
Bombax	<i>Bombax malabaricum</i>	n.a.	368
Castanopsis	spp	n.a.	740
Cedrela	<i>Cedrela toona</i>	Toon	480
Cedrus	<i>Cedrus deodora</i>	Dyar salla	560
Cupressus	<i>Cupressus torulosa</i>	Raisalla	600
Dalbergia	<i>Dalbergia sissoo</i>	Sissoo	780
Daphniphyllum	<i>Daphniphyllum himalense</i>	Rakta Chandan	640
Diospyros	<i>Diospyros spp.</i>	n.a.	840
Eugenia	<i>Eugenia jambolana</i>	Jamun	770
Hymanodictyon	<i>Hymanodictyon excelsum</i>	Bhurkul	513
Juniperus	<i>Juniperus indica</i>	Dhupi	500
Lagerstroemia	<i>Lagerstroemia parviflora</i>	Botdhairo	850
Larix	<i>Larix griffithiana</i>	Himali salla	510
Litsea	spp	n.a.	610
Michelia	<i>Michelia champaca</i>	Champ	497
Myrica	<i>Myrica esculenta</i>	Kafal	750
Pinus	<i>Pinus roxburghii</i>	Chir Pine	650
Pinus	<i>Pinus wallichiana</i>	Blue Pine	480
Quercus	<i>Quercus floribunda</i>	n.a.	970
Quercus	<i>Quercus lamellosa</i>	Thulo Phalant	940
Quercus	<i>Quercus lanata</i>	Banjh	880
Quercus	<i>Quercus leucotrichophora</i>	Tikhe Banjh	1020
Quercus	<i>Quercus semecarpifolia</i>	Kharsu	860
Quercus	spp	n.a.	860
Rhododendron	<i>Rhododendron arboreum</i>	Lali gurans	640
Schima	<i>Schima wallichii</i>	Chilaune	690
Shorea	<i>Shorea robusta</i>	Sal	880
Syzygium	spp	n.a.	770
Terminalia	<i>Terminalia tomentosa</i>	Asna	950
Trewia	<i>Trewia nudiflora</i>	Gutel	452
Tsuga	<i>Tsuga dumosa</i>	Hemlock	450
Other	n.a.	n.a.	680

Appendix 6: The generated spectral variables including vegetation indices and band reflectance.

