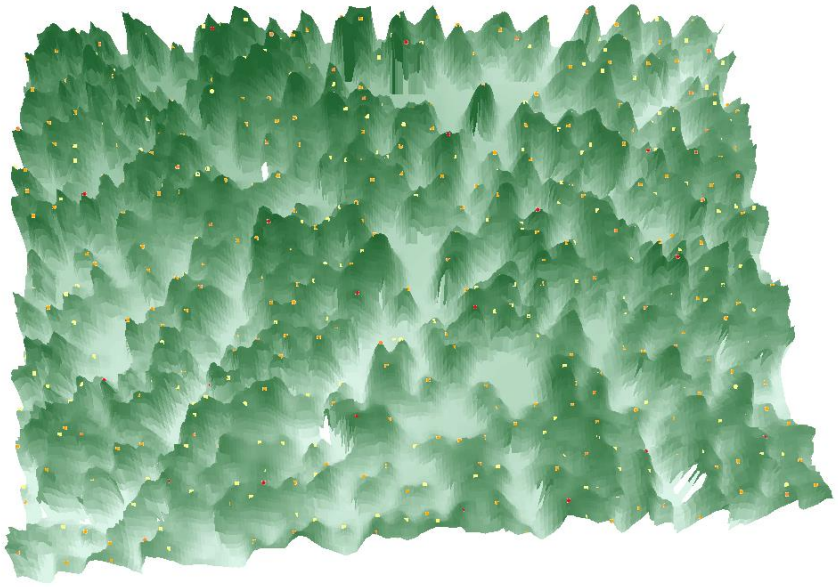
Lund University GEM thesis series nr 15

A multi-scale based method for estimating coniferous forest aboveground biomass using low density airborne LiDAR data



Jialong Duanmu

2016

Department of Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

LUND UNIVERSITY

UNIVERSITY OF TWENTE.
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

A multi-scale based method for estimating coniferous forest aboveground biomass using low density airborne LiDAR data

by

Jialong Duanmu

Thesis submitted to the department of Physical Geography and Ecosystem Science, Lund University, in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation for Environmental Modelling and Management

Thesis assessment Board

First Supervisor: Dr. Abdulghani Hasan (Lund University)
Co-supervisors: Professor. Petter Pilesjö (Lund University)

Exam committee:
Dr. Per-Ola Olsson (Lund University)
Dr. Harry Lankreijer (Lund University)

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed there in remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Course title: Geo-information Science and Earth Observation for Environmental Modelling and Management (GEM)

Level: Master of Science (MSc)

Course duration: January 2016 until June 2016

Consortium partners:

The GEM master program is a cooperation of departments at 5 different universities:
University of Twente, ITC (The Netherlands)
University of Lund (Sweden)
University of Southampton (UK)
University of Warsaw (Poland)
University of Iceland (Iceland)

Cover image:
(CHM model with local maximum height points from high density LiDar data)

Abstract

Forest biomass acts as an important indicator of carbon resources in terrestrial system. Estimation of forest biomass enables a straightforward measurement of carbon storage and provides initial values for process-based carbon cycle models to simulate carbon dynamics. LiDAR (Light Detection and Ranging) remote sensing is increasingly used to estimate forest biomass because of its ability to detect the structure of forest. However, it is still not adequately studied from a viewpoint of multi-scale.

This study is a new attempt for the application of multi-scale theory in forest aboveground biomass (AGB) estimation based on low density LiDAR data (less than 1 point/m^2). The study area is located in Krycklan catchment which is approximately 50 km northwest of Umeå, Sweden. A method based on local maximum height point identification and downscaling calibration is provided. By implementing local maximum elevation extraction and visualization of aerial images, an algorithm directly based on point cloud data is designed. This algorithm retains more details of the LiDAR data and therefore provides better results. Two calibration look-up tables are provided to approximate the forest AGB derived from low density LiDAR data to the forest AGB derived from high density LiDAR data (more than 1 point/m^2). The error of downscaling calibration in the test sample plot is of 0.28%, which proved validity of the method applied.

Key words: LiDAR, remote sensing, aboveground biomass (AGB), multi-scale, downscaling, canopy height model (CHM), individual tree identification, coniferous forest

Popularized Summary

Forest biomass acts as an important indicator of carbon resources in terrestrial system. Estimation of forest biomass enables a straightforward measurement of carbon storage and provides initial values for process-based carbon cycle models to simulate carbon dynamics. Recently, LiDAR (Light Detection and Ranging) remote sensing, as a surveying technology measuring distance by illuminating a target with laser light, is increasingly used to estimate forest biomass because of its ability to detect the structure of forest. However, it is still not adequately studied from a viewpoint of multi-scale.

Briefly, scale describes the resolution and extent in which the data is shown. Multi-scale studies aim at analyzing the relationship between data in different resolution and/or different extent. Multi-scale transform includes downscaling and upscaling. Downscaling is to push down the scale from coarser spatial and temporal resolution into more detailed information with finer spatial and temporal resolution while upscaling is just the opposite. By implementing downscaling and upscaling, multi-scale studies provide possibilities to understand the behavior of variables while changing scale. For forest biomass estimation based on LiDAR data, most of the researches from a viewpoint of multi-scale are about extent. However, the multi-scale based research about resolution is rarely attempted.

This study is an attempt for the application of multi-scale theory in forest aboveground biomass (AGB) estimation based on low density LiDAR data (less than 1 point/m^2). The study area is located in Krycklan catchment which is approximately 50 km northwest of Umeå, Sweden. A method based on local maximum height point identification and downscaling calibration is provided. By implementing local maximum elevation extraction and visualization of aerial images, an algorithm directly based on point cloud data is designed. This algorithm retains more details of the LiDAR data and therefore provides better results. Two calibration look-up tables are founded from a viewpoint of downscaling which is widely applied in geomorphology study. By inferring the result extracted from high density LiDAR data (more than 1 point/m^2) with

low density LiDAR data, the forest parameter estimation accuracy based on low density LiDAR data is improved.

The error of downscaling calibration in the test sample plot is of 0.28%, which proved validity of the method applied. Furthermore, the calibration look-up tables can be used directly in the further researches of the study area in the same situation.

Key words: LiDAR, remote sensing, aboveground biomass (AGB), multi-scale, downscaling, canopy height model (CHM), individual tree identification, coniferous forest

Table of Contents

List of figures

List of tables

List of abbreviations:

AGB – Aboveground Biomass

CHM – Canopy Height Model

DEM – Digital Elevation Model

DSM – Digital Surface Model

DTM – Digital Terrian Model

GPS – Global Positioning System

IWS – Inverse Watershed Segmentation

LiDAR – Light Detection and Ranging

LM – Local Maxima

VWF – variable window filter

Acknowledgements

Firstly, I would like to express my special thanks of gratitude to my supervisor, Dr. Abdulghani Hasan, for this continuous support of my study. He has always been helpful and enthusiastic. His guidance helped me in all the time of research and writing of this thesis.
I would also like to thank my co-supervisor Petter Pilesjö. Without his participation and help, the study may not have a chance to be started.
A special thank you goes to Yanzi Yan, who has sustained me through the hardest time.
Finally, I am thankful to the people at Lund University, for everyone who has shared their time and idea with me.
Thank you very much, everyone!

1. Introduction

1.1 background

Forest biomass acts as an important indicator of carbon resources in terrestrial system (Le Toan et al. 2011). Estimation of forest biomass enables a straightforward measurement of carbon storage and provides initial values for process-based carbon cycle models to simulate carbon dynamics (Sessa and Dolman, 2008).

There have been different approaches applied for forest biomass estimation. Traditionally, field measurements are the most accurate methods for estimation. However, these methods are usually labor intensive and time consuming, and also cannot provide the temporal and spatial distribution of biomass at large scales (Brown, 2002).

Through correlations between spectral information detected by remote sensing and forest biomass, forest biomass can be estimated with large spatial and temporal coverage. However, due to the complexity of canopy characteristics, it remains a challenge to establish such correlations. Recently, LiDAR (Light Detection and Ranging) remote sensing is increasingly used to estimate forest aboveground biomass (AGB) because of its advantage of detecting the structure of forest. However, it is still not adequately studied from a viewpoint of multi-scale.

1.1.1 LiDAR based forest biomass estimation from a viewpoint of multi-scale

Conceptually, scale represents the 'window of perception', the filter, or measuring tool with which a system is viewed and quantified (Hay et al., 2001). An important characteristic of scale lies in the distinction between grain and extent. For raster data, grain is equivalent to the spatial resolution of the pixels composing a raster data, while extent is the total area that a raster data covers. Briefly, scale describes the resolution and extent in which the data is shown. It is used in the context of space (geographic scale), time (temporal scale) and many other dimensions of research (Goodchild and

Quattrochi, 1997). Multi-scale studies aim at analyzing the relationship between data in different resolution and/or different extent. As a scale changes, so do the associated patterns and processes of reality. Multi-scale transform includes downscaling and upscaling. Downscaling is to push down the scale from coarser spatial and temporal resolution into more detailed information with finer spatial and temporal resolution while upscaling is just the opposite. By implementing downscaling and upscaling, multi-scale studies provide possibilities to understand the behavior of variables while changing scale.

Like topographic map, LiDAR data in each scale has its own meaning to achieve different goals. Balanced by size, processing speed, accuracy and precision, LiDAR processing at each specific resolution has its irreplaceable significance. By weighing the pros and cons, decisions of spatial resolution selection are made before researches. Through multi-scale analysis, researches in different scales are united as a system but not isolated.

Multi-scale analysis based on LiDAR measurements are of great importance in studies of biology, geomorphology, ecology etc. Multi-scale analysis using LiDAR measurements of canopy height is performed well to study the spatial scale of habitat selection of riparian birds (Seavy et al., 2009; Hyde, 2006), or marine species (Pittman and Brown, 2011; Zawada and Brock, 2009). Lidar data is also used to characterize and monitor forest successional stages at different scales (Falkowski et al., 2009). Multi-scale pattern analysis of LiDAR DEM (digital elevation model) could delineate the morphometric characteristic of landform elements (Drăguţ and Blaschke, 2011). In addition, Lidar is suggested as premier instrument for mapping biomass across broad spatial scales (Clark et al., 2011; Asner et al., 2012). (Zhao et al., 2009) proposed two scale-invariant models for estimation of forest biomass at a range of scales from individual tree, plot, stand, local or even up to regional levels. Multi-scale drivers of spatial variation in forest carbon density are studied with LiDAR and an individual-based landscape level (Seidl et al., 2012).

For forest biomass estimation based on LiDAR data, most of the researches from

a viewpoint of multi-scale are about extent (Cohen et al., 2013; Englhart et al., 2011; Du et al., 2014). However, the researches in different resolutions are still isolated in most cases. The multi-scale based research about resolution is rarely attempted.

1.1.2 The overview of forest aboveground biomass estimation method based on airborne LiDAR data

Forest aboveground biomass estimation based on airborne LiDAR data has been successfully attempted by many researchers. For forest biomass estimation based on airborne LiDAR data, tree height is one of the most appropriate forest parameter that is able to predict biomass estimate with an accuracy of approximately 85% (Papathanassiou et al., 2005). After the identification of individual trees, tree height of each individual tree is extracted. With the tree height-biomass model, the aboveground biomass of each individual tree is estimated. By summing up the biomass of all the individual trees in the study area, the forest aboveground biomass is estimated. For boreal coniferous forest, as the tree height-biomass model is already built by former researchers (Shendryk et al., 2014), the individual tree identification, tree height extraction and the calibration of the extracted parameters becomes particularly important. However, the most commonly used processing methods still have their limitations.

1.1.2.1 Individual tree identification

In the previous works, two software algorithms has been most commonly used, TreeVAW and Inverse Watershed Segmentation (IWS).

TreeVaw operates on a canopy height model (CHM) using a variable window filter (VWF) that varies its search window size (Falkowski et al., 2008; Popescu and Wynne, 2004), otherwise known as a convolution kernel (Jensen, 1996), determines a tree location based on elevation data by passing a local maxima (LM) filter over the CHM The highest elevation value is taken to indicate the tree apex based on the assumption that surrounding pixels are assumed to represent laser hits of the same tree crown (Kini and Popescu, 2004). When the filter determines a LM value, a tree's x and y coordinate

location is identified and then the crown diameter is determined based on the allometric relationship to height (Popescu, 2004; Kini and Popescu. 2004).

IWS is a combination of inversion and watershed segmentation. In the CHM, Each tree in the forest was represented by a single local maximum and the pixels around of the maximum should be assigned to the maximum (treetop) they most probably belong to. After inversion, it is just like a hydrologic drainage basin. Watershed segmentation is an algorithm designed for delineating the basins. It finds local minima in a grayscale image and tries to assign each pixel of the image to a local minimum. All pixels belonging to the same minimum were labeled with a unique value, forming a hydrologic drainage basin. Following inversion, a watershed segmentation algorithm separates the CHM into the equivalent of individual hydrologic drainage basins. This algorithm is the most common method applied to determining locations of individual tree crowns and raster crown diameter and height values (Ziegle et al., 2000; Edson and Wing, 2011).

Although both algorithms have been implemented by the former researchers, they still seems to have limitations:

1. Huge amount of field measurement is required to calculate the parameters used by the algorithms. The main idea of these two algorithms are almost the same. Briefly, for identifying an individual tree, a pixel with the maximum height in the local area is extracted as the top of a tree. To acquire the area of this 'local area', TreeVAW requires a width-height model of the tree structure while IWS requires a threshold describing the minimum size of canopy. Both the width–height model and the minimum canopy threshold needs field measurements of hundreds of trees to make them precise and reliable.

2. A lot of details are lost while converting the LiDAR data to raster data. Both algorithms are based on raster data. For both of these two algorithms, the LiDAR data has to be converted to raster data while implementing pre-processing. By resampling and interpolation, some of the data is lost, some of the data is created arbitrarily. Furthermore, by implementing some filtering algorithm (e.g. smoothing), the reliability of the data decreased again which leads to great error in process.

1.1.2.2 Tree height extraction

In the previous research, tree height is the elevation extracted by the coordinate of the each identified tree in CHM. Although the airborne LiDAR based measurement of elevation is with high accuracy, it shows a trend that there is always an underestimation of the tree height in the previous research, especially for the tree height extraction based on low density LiDAR data (less than 1 point/m^2).

1.1.2.3 Calibration of the extracted parameters

Due to the high accuracy of airborne LiDAR measurement, the calibration of the extracted data is rarely implemented. Based on high density airborne LiDAR data (more than 1 point/m^2), the biomass estimation can reach an accuracy of approximately 85% (Papathanassiou et al., 2005). However, based on low density LiDAR data, both individual tree identification and tree height extraction shows a far greater error which make the calibration necessary. The most commonly used calibration method is to build a relationship to approximate the extracted parameters to the field data. However, it also has its limitation. Field measurement is always restricted by labor, research fund and accessibility. For instance, while estimating the forest biomass a few years earlier, a field measurement seems not that reliable since the forest would grow in the gap years which makes the field data not that representative for the truth.

1.2 Aims and hypothesis

This study is an attempt for the application of multi-scale theory in forest AGB estimation based on low density LiDAR data. A method based on local maximum height point identification and downscaling calibration is provided.

To overcome the limitations of the most commonly used software algorithms, an algorithm based on local maximum height point extraction is provided. In this algorithm, individual tree identification and tree height extraction are directly based on original LiDAR point cloud data which retains more details of the raw data. The dependence of field data based canopy structure model in the previous algorithms is avoided by the following calibration of individual tree identification based on visualization of aerial image.

A theoretical model is founded to analyze the underestimation of the tree height, which influences the way tree height calibration is implemented.

Calibrations of the tree parameters are implemented by building relationships between data in different scales. In order to implement the calibration of individual tree identification, the relationship between the result from aerial image visualization and the result from LiDAR data processing is built. In order to implement the calibration of tree height extraction, the relationship between the result from high density LiDAR data processing and the result from low density LiDAR data processing is built. The calibration is therefore implemented by inferring a better result from the result of low density LiDAR data processing with the mentioned relationships. By implementing the calibration from a viewpoint of multi-scale, field measurement is avoided, which makes the forest AGB estimation of the forest in the past available.

All the calibration look-up tables derived from the processing can be used directly by the researchers in the future study to make the forest AGB estimation based on low density LiDAR data in the same situation much more accurate and flexible.

The aims in this study were：

To develop an algorithm based on local maximum height point extraction for individual tree identification and tree height extraction using low resolution LiDAR data.

To develop a new method from the viewpoint of downscaling for the calibration of individual tree identification and tree height extraction.

To establish a theoretical model to analyze the causes of the underestimation while implementing the tree height extraction based on airborne LiDAR data.

To create calibration lookup tables that can be used by the further researchers while dealing with the forest AGB estimation under similar conditions.

The hypotheses addressed in this study were:

The relationships between the extracted forest inventory parameters based on different density LiDAR data can be built to infer the result of forest AGB estimation based on high density LiDAR data (more than 1 point/m^2) with the result based on low

density LiDAR data (less than 1 point/m^2) with an acceptable accuracy.

Algorithm based on local maximum height point identification directly can be used to estimate the coniferous forest aboveground biomass using low density airborne LiDAR data (less than 1 point/m^2).

Downscaling calibration can be used to approximate the forest aboveground biomass estimated from low density LiDAR data (less than 1 point/m^2) to the forest aboveground biomass estimated from high density LiDAR data (more than 1 point/m^2).

2. Materials and methods

2.1 study area

The Krycklan catchment is approximately 50 km northwest of Umeå, Sweden. The climate is characterized by long winters and short summers. As measured from 1980 to 2008, the mean annual precipitation and temperature are 612 mm, and +1.7°C, respectively, with an average runoff of 312 mm. Snow covers the ground for 168 days on average, from the end of October to the beginning of May (1980–2007) (Laudon et al., 2011).

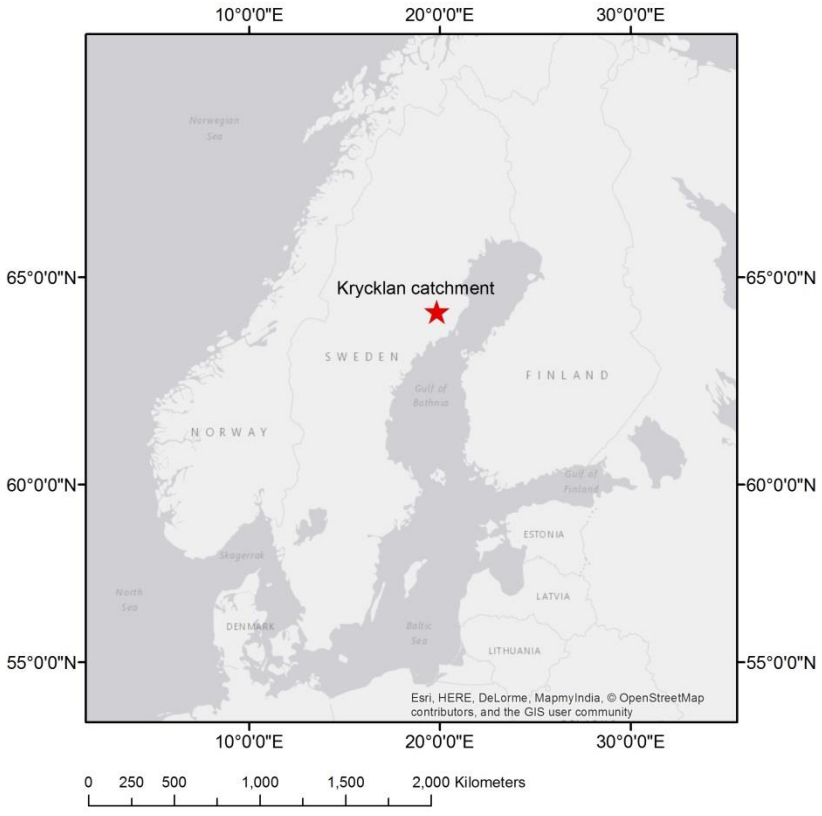


Figure 1 Map of the location of the study area

Over 87% of the Krycklan Catchment is covered by forests, mainly Scots pine (*Pinus sylvestris*), spruce (*Picea abies*), and birch (*Betula spp*.) (Peralta-Tapia et al.,

2015). The upper part of Krycklan catchment is mainly forested by conifer. Norway spruce is the dominant tree species in low-lying areas while Scots pine is the dominant tree species in upslope areas interspersed by mires and lakes. In lower part of the catchment, although the conifer is still the dominant tree species, deciduous trees becomes more common along the streams.

Considering different landscapes of forest and distributions of tree height, five areas of coniferous forests without significant external disturbance during the period from 2008 to 2010 (e.g. thinning, wild fire) are considered as the study area, Area 1 to 5 (Figure 1). Generally, trees with height great than 10 m are mostly seen in Area 2. In Area 3, forest, with a wide bare ground inside it, are of relative equal height distribution. There is a lake inside Area 4, which varies the landscape of forest in Area 4. Low trees are quite significantly distributed in the Area 5, while high trees are around low vegetation.

In this study, Area 1 is chosen to be the test area, while Area 2 to 5 are chosen to be the training areas. After segmentation and evaluation, 4 sample plots from the test area and 23 sample plots from the training areas are selected to be used in this study.

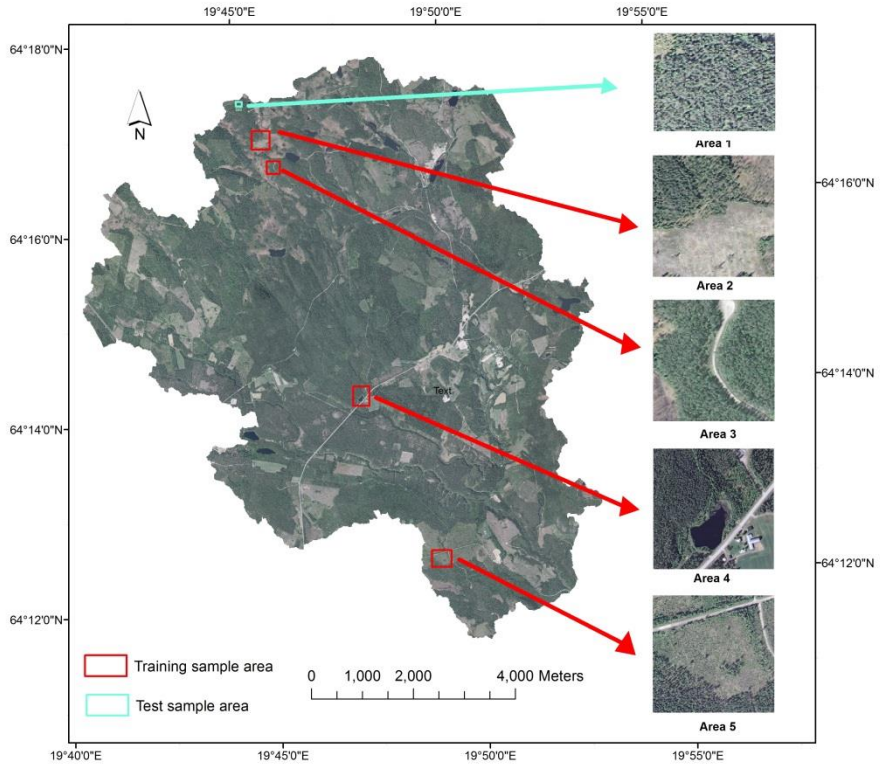


Figure 2 Map of the study area

(The background map is from Swedish National Land Survey Dnr: I2014/00579)

2.2 Data

In this study, four types of data are used for implementing forest biomass estimation, namely, high density LiDAR data, low density LiDAR data, DEMs (Digital Elevation Models) and aerial photographs. High density LiDAR data and aerial photographs of the Krycklan catchment as part of the BioSAR 2008 campain (Hajnsek et al., 2009) was performed on 5-6 August 2008, with the TopEye MKII system (S/N 425) that is mounted on a helicopter flying at an altitude of 500 m above ground level for the main strips and 250 m above ground level for the cross strips (Persson and Fransson, 2014). Approximately 70 km^2 were covered using an average density of approximately 5 returns per square meter in the main strips and 15 return per square meter in the cross strips. The aerial photographs are with a ground resolution of about

0.09 m. The coordinate system of both laser point data and aerial photographs used was RT90, but the data were transformed to SWEREF 99 TM in order to match the other data sources.

The low density LiDAR data is from the Swedish National Land Survey. A project has been undertaken by Swedish National Land Survey since 2009, with the goal of scanning entire Sweden in order to generate an accurate digital terrain model (DTM) which, as only minor changes in the terrain are expected, should remain stable for the foreseeable future (Owemyr and Lundgren. 2010). This airborne laser scanning is with at least 0.5 returns per square meter and the average height error in all terrain types (coniferous forest, grassland, etc.) is about 0.25 m. The coordinate system is SWEREF 99 TM. The low density laser data used in this paper were collected in October 2010.

A bare-ground DEM with 2 m resolution with an average elevation error of 0.5 m, was generated from the ground elevation returns of the low density LiDAR signals using triangulated irregular network interpolation (Ågren et al., 2015). The coordinate system is SWEREF 99 TM.

2.3 Method

This section is divided into two parts. The first part is how to implement the forest AGB estimation using low resolution LiDAR data with this algorithm and how to improve the estimation with the given calibration look-up tables. The second part is how the calibration look-up tables are derived.

2.3.1 Forest aboveground biomass estimation using low density LiDAR data

2.3.1.1 Data preprocessing

In this study, all the correction for the data has been implemented by the data provider.

Each point in the point cloud has a first return which represents the highest point of the object in the certain coordinate. Before processing the data, all the LiDAR data and DEM data needs to be converted into the same coordinate system in order to make

the processing easier. The pre-processed LiDAR should be saved as ASCII files with 3 columns which contains the data of each point: the X-coordinates, the Y-coordinates, and the elevation of the first return. The pre-processed DEM data should be ASCII files.

2.3.1.2 CHM model extraction

Each point in the point cloud has a first return which represents the highest point of the object in the certain coordinate. All the first return points constitutes the surface of the land cover, which is called digital surface model (DSM). With elevation from DEM data subtracted, data about the distance between the surface and the terrain can be extracted. In the forest, this data is always regarded as the canopy height model (CHM) (Naesset, 1997).

In this study, CHM is processed and stored in the form of point cloud data. The ground elevation of each point in the LiDAR point cloud data is retrieved from the DEM data based on its X-Y coordinate. By subtraction of the two elevation, the height of each point would be extracted and stored in a matrix with its X-Y coordinates. For detecting trees, all the ground points should be removed from the CHM. In this study, all the points with a height less than 2m would be ignored as ground points.

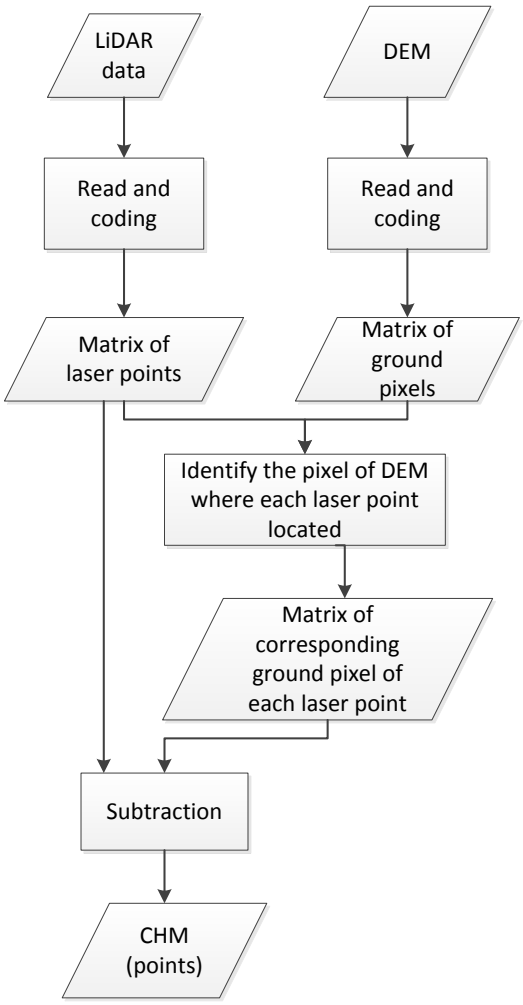


Figure 3 Procedures for CHM extraction

2.3.1.3 Calculating the local maximum level

Each point in the CHM has a height value which represents its height from ground surface. If a point has a height value greater than any other points nearby, we would be inclined to infer that it might be the most appropriate point to present the real top of the tree it falls on. In order to identify the individual trees in the CHM data, a concept of local maximum level is involved in this study. Local maximum level of a specific point represents the radius of the maximum local circular region in which area the point has the greatest height value. The rounded down radius (with the unit of meter) of the area is recorded as an attribute called local maximum level. For all the points with a local maximum level greater than 15, it is almost sure that they would represent the top of a tree and do not need a much more detailed classification. To simplify the classification,

for the point which has a local maximum level greater than 15, the local maximum level is set as 15 arbitrarily to make them in the same class.

In this study, local maximum level is the basis of identifying an individual tree. To get this value, each point in the LiDAR data would be tested in order to extract the horizontal distance between itself and the nearest point which has a greater height value.

As shown in Figure 4 each point in the LiDAR data would be processed for its local maximum level. In the first step, all the points with an absolute difference (along X-Axis and Y-Axis respectively) less than 15m between itself and the current tested point would be retrieved. These points will then be compared with the current tested point to see whether its height value is greater than that of the current tested point. The nearest higher point would be updated until all the other points in the Lidar data are compared. The distance between the current tested point and the nearest higher point will be rounded down and set as the local maximum level. The local maximum level would be set as 15 if it is greater than 15. All the points with a local maximum level no less than 2 would be filtered out as the local maximum height points. The height value of these points will be regarded as the tree height of the corresponding trees if the trees are correctly identified.

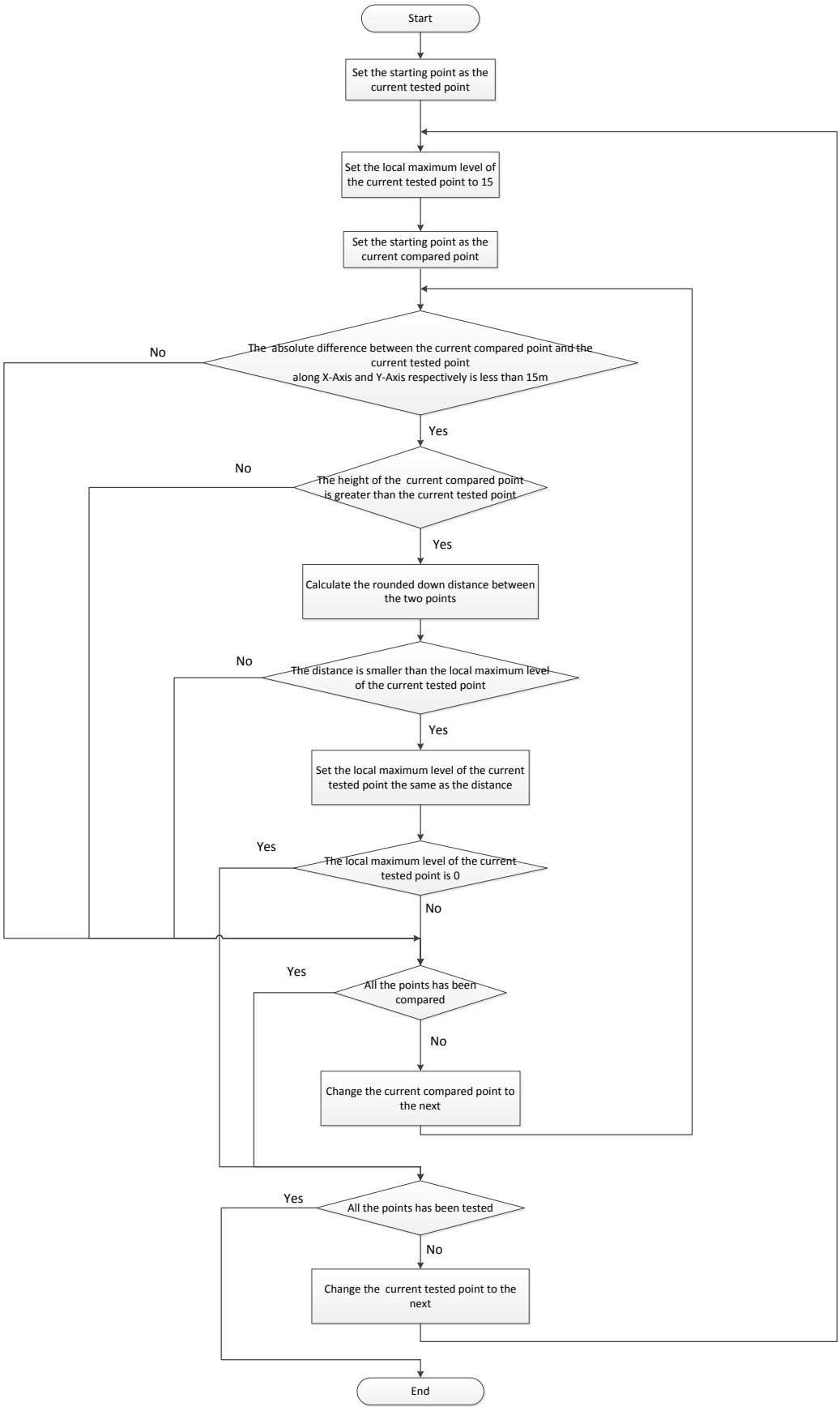


Figure 4 Procedures for local maximum level extraction

2.3.1.4 Looking up local maximum point representativeness value in the look-up table

Due to the lack of data precision and the limitation of identification technique, single tree identification is always a technical barrier (Edison and Wing, 2011). Obviously, the number of the trees in practice is not as same as the number of the local maximum height points. What can be concluded from the processing is these local maximum height points represents these trees. To quantify this representativeness, a concept of local maximum point representativeness is involved to implement the calibration. The local maximum point representativeness describes how many similar trees a local maximum height point represents in average. For instance, a point with a local maximum point representativeness value of 0.9 means this point represents 0.9 trees in average with the same height. In this step, with the local maximum level and the height value of a point, the local maximum point representativeness value is retrieved in the look-up table for further use. What the table looks like and how the table is obtained will be precisely elaborated in Section 2.3.2.1.

2.3.1.5 Tree height calibration

Tree height extraction usually causes great error. In this step, the tree height calibration coefficients of each local maximum height point are retrieved from the look-up table with its height and local maximum level. How the tree height calibration look-up table is built will be precisely elaborated in Section 2.3.2.2.

2.3.1.6 Final Biomass estimation

For forest biomass estimation based on airborne LiDAR data, tree height is one of the most appropriate forest parameter that is able to predict biomass estimate with an accuracy of approximately 85% (Papathanassiou et al., 2005). The mostly typical model for describing the relationship between tree height and AGB was described is a power function of the form $y = a \times x^c$ (Mette et al., 2006), where y denotes the biomass, x denotes the tree height, a and c are constants. The dominant species in the study area is Norway spruce and Scot pine. As we are estimating the AGB of a forest in 2008, field measurement is not possible which makes a detailed classification between different coniferous species not available. In this study, all the coniferous forest biomass

estimation will be calculated with the Norway spruce tree height-biomass model (Shendryk et al., 2014):

$$AGB_{spruce} = 0.1183 \times H^{2.528} \quad (1)$$

Where H denotes the height of tree.

Calibrated by the local maximum point representative value and tree height calibration value, the final formula of forest AGB estimation is:

$$AGB = \sum 0.1183 \times C_R \times (H \times C_H)^{2.528} \quad (2)$$

Where C_R denotes the local maximum point representativeness, C_H the tree height calibration coefficient and H the tree height.

2.3.2 Derivation of the calibration look-up tables

2.3.2.1 Calibration of individual tree identification

Since the aim of this study is forest AGB estimation which sums the estimated biomass of all the trees in the forest, the precise position of the misidentified trees is not that essential. The emphasis of the identification calibration is to take all the misidentified trees into consideration, and make a correction of the final estimated biomass to make sure the biomass of the misidentified trees is taken into account reasonably.

The concept of local maximum point representativeness is involved to enhance the calibration. The Local maximum point representativeness value describes the number of trees a local maximum height point represents. For instance, “the local maximum point representativeness of a local maximum height point is 1.38” means that a point with a specific height and a specific local maximum level represents there should be 1.38 trees with the similar height by the average.

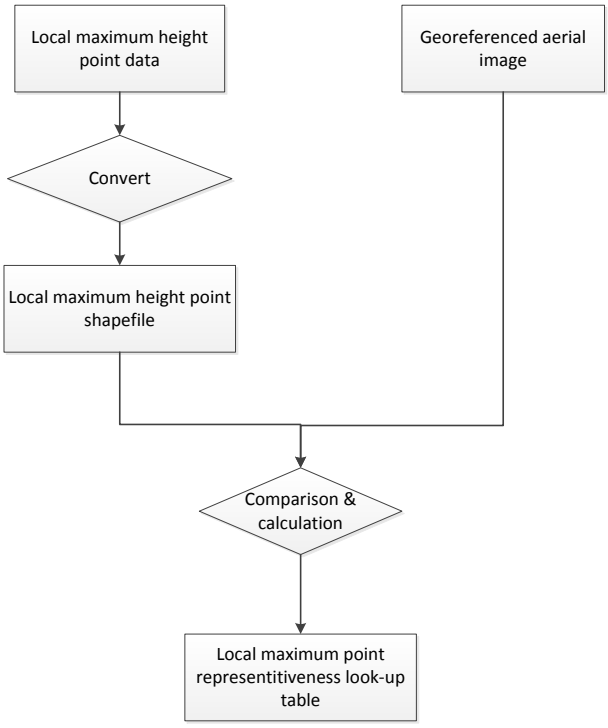


Figure 5 Procedures for calculation of local maximum point representativeness

Figure 5 is the flowchart of how local maximum point representativeness is calculated. The shapefile of the local maximum height points would be stacked on the geo-referenced aerial image for a visualized comparison.

All the points which doesn't present the top of a tree will be detected by visualization and ignored. As the tree height in the artificial forest usually presents a trend of homogeneity, all the trees which are not identified would be approximately regarded as a tree with the same height and same local maximum level as the nearest local maximum height point. 8 height levels are set in this research for a further classification. For each height level and each local maximum level, the number of the local maximum height points and the number of corresponding trees detected by visualization would be counted. The local maximum point representativeness would be calculated with the formula:

$$C_R = \frac{N_{local\_maximum\_height\_points}}{N_{corresponding\_trees\_detected\_by\_visualization}} \quad (3)$$

Where C_R denotes the representativeness, N means numbers.

A look-up table for the representativeness would be produced then. The representativeness would be displayed by different height level and different local maximum level in the look-up table.

2.3.2.2 Calibration of tree height extraction

Limited by time, labor and research fund, it is not possible to calibrate the tree height results to fit the tree height in practice directly in a large-scale with enough samples. However, by building a relationship between the tree height data extracted from high density LiDAR data and the tree height data extracted from low density LiDAR data, far larger amount of data could be used as samples to produce a full-scale tree height calibration look-up table.

In this section, the error of tree height extraction using LiDAR data is analyzed theoretically. Supported by the theoretical analysis, a method based on downscaling theory is designed. The relationship between the tree height data extracted from high density LiDAR data and the tree height data extracted from low density LiDAR data is built by weighted-piecewise pairing. A simplified piecewise linear regression is implemented then to provide a calibration look-up table from which the coefficient for tree height calibration can be retrieved by matching the height value of the local maximum height point.

a. Theoretical model analysis

This section is about the relationship between the relative error of the tree height estimation and the laser point density which supports the method of tree height calibration.

To simplify the research, a theoretical model is built. The shape of a coniferous tree is assumed as a cone approximatively.

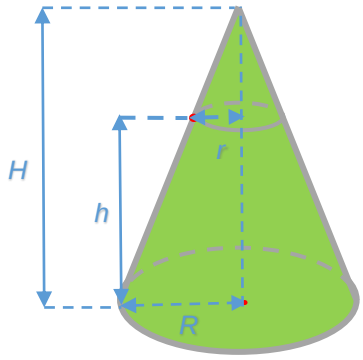


Figure 6 Tree height extraction model

In Figure 6, the red dot denotes a laser point on the tree. *R* denotes the radius of the cone base while *r* denotes the radius of the circle paralleling to the cone base where the laser point is on the boundary. H denotes the tree height while *h* is the vertical distance between the point and the ground plain. For each specific laser point on the canopy, there is:

$$\frac{H-h}{H}=\frac{r}{R} \tag{4}$$

While implementing a tree height estimation based on LiDAR data, a lot of laser points falls on the boundary of the tree. The point with the greatest height value is regarded as the representation of the tree top. The greatest *h* value is then regarded as the estimated tree height. *h* is always smaller than *H*, that is why there is always an underestimation while extracting tree height from LiDAR data no matter how high the density of the laser points is. The relative error of the estimation is expressed as:

$$\delta=\frac{H-h_{max}}{H} \tag{5}$$

where δ denotes the relative error, h_{max} denotes the maximum *h* value.

According to equation (4), since *R* and *H* is constant for a specific tree, the bigger *h* is, the smaller *r* is. The point with the biggest *h* value has the smallest *r* value, therefore:

$$\delta=\frac{H-h_{max}}{H}=\frac{r_{min}}{R} \tag{6}$$

where r_{min} denotes the smallest *r* value among that of all the laser points.

Assuming that there are *n* laser points on the cone, and r_i is *r* measured from point *i*. It's easy to derive the distribution function of r_i which is shown in Equation (7) and Equation (8),

$$P(r_i \le x) = \frac{x^2}{R^2}, \quad x \in [0, R] \tag{7}$$

$$P(r_i > x) = 1 - \frac{x^2}{R^2}, \quad x \in [0, R] \tag{8}$$

where *x* is less than or equal to *R* but not less than 0.

As all the *r* values are mutually independent, the probability of *n* points with their *r* value greater than *x* is calculated as:

$$P(r_i > x; i = 1, 2, \ldots, n) = \left(P(r_i > x)\right)^n = \left(1 - \frac{x^2}{R^2}\right)^n, \tag{9}$$

The probability of the minimum r_i $(i = 1, 2, \ldots, n)$ with value less than *x* could be derived by Equation (10):

$$P(r_{min} \le x) = 1 - \left(1 - \frac{x^2}{R^2}\right)^n \tag{10}$$

where $\min r_i$ is the minimum value of *r* from *n* points.

The probability density function of r_{min} is shown below:

$$f(x) = \frac{dP}{dx} \tag{11}$$

According to Equation(6):

$$E(\delta) = \frac{E(r_{min})}{R} \tag{12}$$

where $E(r_{min})$ denotes the expectation of r_{min} and $E(\delta)$ denotes the expectation of δ, Thus $E(\delta)$ could be calculated as Equation (12) – (15):

$$E(r_{min}) = \int_0^R x f(x)\, dx = \int_0^R x\, dP\Big|_0^R - \int_0^R P\, dx = \int_0^R \left(1 - \frac{x^2}{R^2}\right)^n dx \tag{12}$$

$$\int_0^R \left(1-\frac{x^2}{R^2}\right)^n dx = \int_0^R \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{R^{2k}} dx = \sum_{k=0}^{n} (-1)^k \frac{C_n^k x^{2k+1}}{(2k+1)R^{2k}}\Bigg|_0^R = R\sum_{k=0}^{n} \frac{(-1)^k C_n^k}{2k+1} \tag{13}$$

$$E(r_{min}) = \int_0^R \left(1-\frac{x^2}{R^2}\right)^n dx = \frac{2n}{2n+1}\int_0^R \left(1-\frac{x^2}{R^2}\right)^{n-1} dx \tag{14}$$

$$E(\delta) = \frac{E(r_{min})}{R} = \frac{1}{R}\int_0^R \left(1-\frac{x^2}{R^2}\right)^n dx = \frac{1}{R}\frac{2n}{2n+1}\int_0^R \left(1-\frac{x^2}{R^2}\right)^{n-1} dx \tag{15}$$

From Equation (15), the relationship between *E(δ)* and *n* is built.

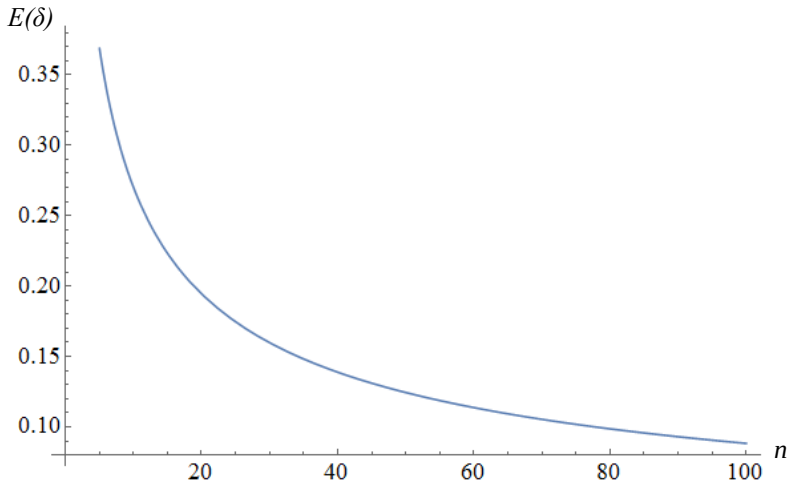


Figure 7 Distribution of *E(δ)* and *n* where *(E(δ)* denotes expectation of the relative error of the tree height extraction based on LiDAR data while *n* denotes the number of the laser points on the tree

Figure 7 shows the relationship between *E(δ)* and n, where *E(δ)* denotes expectation of the relative error of the tree height extraction based on LiDAR data, *n* denotes the number of the laser points on the tree. It can be easily seen that with *n* increasing, *E(δ)* has a decreasing property close to 0. It means that the estimated tree height is getting closer to the truth when the number of the laser points on the tree increases. It also suggests that estimated tree height with low density LiDAR data will be closer to that with high density LiDAR data when it turns to a bigger tree with more laser points on the canopy.

Furthermore, since the relative difference between the tree height extracted from the different density LiDAR data decreases monotonically with the size of the tree

increasing, the processing method based on establishing piecewise linear correlations between the tree height extracted from low density LiDAR data and high density LiDAR data could be appropriate for tree height calibration.

b. Weighted-piecewise pairing

Before implementing the regression, the extracted tree height data from high density LiDAR data and low density LiDAR data should be matched in pairs. However, as the number of the detected trees from different LiDAR data might not be the same, it is inappropriate to build the relationship by matching individual trees in pairs.

To solve this problem, a piecewise pairing weighted by local maximum point representativeness is implemented, because no matter how accurate the individual tree identification is, the proportion of the trees in each height level will not present much heterogeneity.

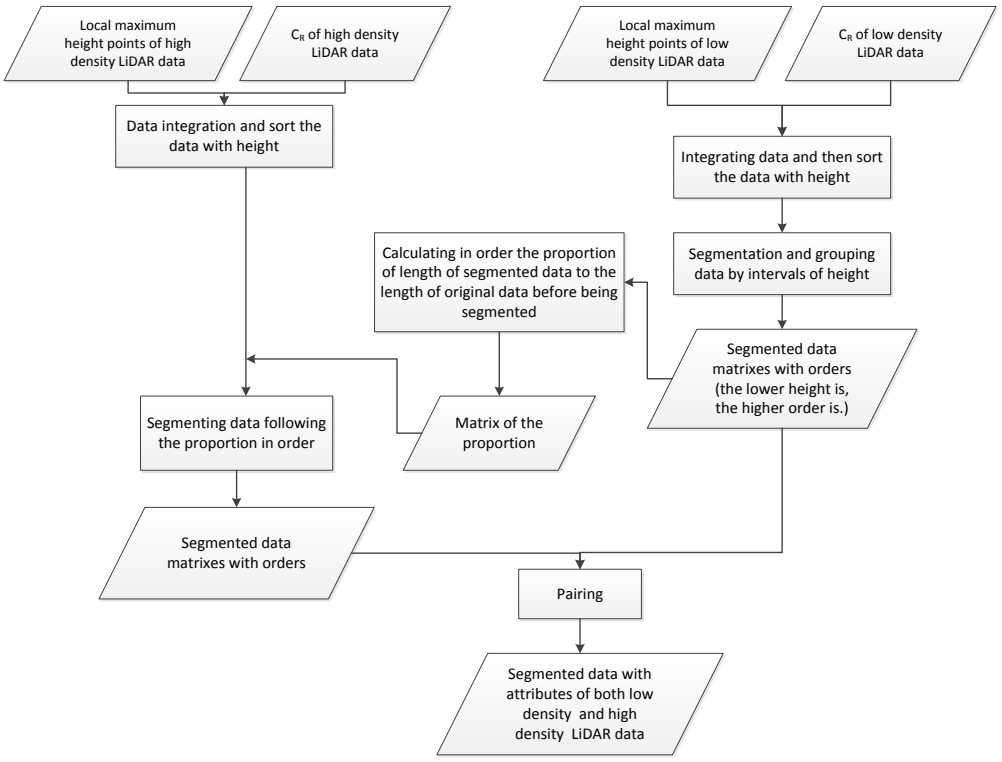


Figure 8 Piecewise pairing process of low density and high density LiDAR data

As shown in figure 8, first, the tree height data extracted from low density LiDAR data is ordered and segmented by the height value. The number of trees (N) of each segmentation weighted by local maximum point representativeness will be calculated

with the following formula.

$$N = \sum H \times C_R \tag{16}$$

Where C_R means the local maximum point representativeness and H the tree height.

The proportion of the number of trees for each segmentation data is also calculated and recorded for further use.

To match the segmentations, the local maximum height point data extracted from high density LiDAR data is also ordered by the height value. The tree number weighted by local maximum point representativeness is then segmented by proportion of each tree height levels in the data derived from low density LiDAR data. A pairing between two sets of tree height data is implemented by the segmentation of height value.

c. Piecewise linear regression

In order to simplify the calibration, the tree height average of each segmentation weighted by local maximum point representativeness is calculated. A subtraction is implemented between each pair of segmentation. After that, a look-up table for tree height calibration is built. The tree height calibration coefficient can be retrieved from the look-up table with the height value of the local maximum height point derived from low density LiDAR data.

3. Results

In this section, the final calibration look-up tables and the forest AGB estimation results will be presented. Both the result and some details about processing will be displayed to make a better explanation. The accuracy assessment will be shown at last to evaluate the effectiveness of the calibration for both the forest inventory parameter and the biomass estimation.

3.1 The look-up table of local maximum point representativeness

Due to the time consuming manual identification of individual trees, only one sample plot is chosen as the training sample plot from each of the four training areas (Area 2 to 5). These four chosen training sample plots constitute the training data of this look-up table.

As a reference to show the position and number of trees in practice, the aerial images of the sample plots are well georeferenced by registration at first. The local maximum height points and the corresponding trees detected by visualization are then classified by height and local maximum level. All the local maximum height points which doesn't present the top of a tree will be ignored. All the trees which are not identified would be approximately regarded as a tree with the same height and the same local maximum level as the nearest local maximum height point.

The comparison between the distribution of the local maximum height points and the distribution of the trees corresponding trees detected by visualization is shown in Table 1.

Table 1 Comparison of local maximum height points with corresponding trees in practice in training sample plots (The fraction in the table is a form to compare the number of local maximum height points with specific local maximum level and height with the number of corresponding trees in practice.

Height \ Local maximum Level	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4 – 6 m	14/24	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
6 – 8 m	37/28	7/3	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
8 – 10 m	58/42	14/21	7/2	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
10 – 12 m	81/79	38/25	16/13	2/0	4/2	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
12 – 14 m	127/101	93/102	47/53	17/17	10/11	2/1	3/0	0/0	1/0	1/0	0/0	0/0	0/0	0/0
14 – 16 m	137/145	111/115	65/66	40/48	48/45	25/22	10/15	10/13	4/6	1/1	1/1	0/0	0/0	3/6
16 – 18 m	34/45	29/34	31/17	29/18	12/5	12/10	10/7	12/9	5/3	7/7	6/6	7/7	3/4	26/25
> 18 m	11/10	6/7	5/5	2/2	1/1	3/4	0/1	1/1	2/2	2/2	0/0	1/0	0/0	11/5

The local maximum point representativeness is calculated by the formula. The look-up table of local maximum point representativeness is shown in Table 2. To avoid the bias caused by the lack of samples, the calibration is implemented only for the local maximum height points with a sample size larger than 50.

Table 2 Look-up table of local maximum point representativeness

Height \ Local maximum Level	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4 – 6 m	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6 – 8 m	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8 – 10 m	1.381	1	1	1	1	1	1	1	1	1	1	1	1	1
10 – 12 m	1.026	1	1	1	1	1	1	1	1	1	1	1	1	1
12 – 14 m	1.257	0.912	1	1	1	1	1	1	1	1	1	1	1	1
14 – 16 m	0.945	0.965	0.985	1	1	1	1	1	1	1	1	1	1	1
16 – 18 m	1	1	1	1	1	1	1	1	1	1	1	1	1	1
> 18 m	1	1	1	1	1	1	1	1	1	1	1	1	1	1

3.2 The look-up table of tree height calibration

All of sample plots from the training areas (Area 2 to 5) are considered as the training sample plots for creating the look-up table of tree height calibration. The

sample plots in Area 1 is considered as the test sample plots to evaluate the effectiveness of the calibration.

The identified trees are classified by height. The sum of each class weighted by local maximum point representativeness is shown in Table 3.

Table 3 Weighted sum of local maximum height points of low density laser points of training sample plots

Height	Weighted sum
4 – 6 m	578
6 – 8 m	715
8 – 10 m	1030
10 – 12 m	1154
12 – 14 m	1840
14 – 16 m	2383
16 – 18 m	2185
> 18 m	1069

Based on the proportion of each class and the local maximum point representativeness look –up table of high density LiDAR data, a classification of local maximum height points derived from high density LiDAR data is implemented then. By calculating the number of trees in each corresponding class, the sum of each class weighted by local maximum point representativeness is shown in Table 4.

Table 4 Weighted sum of local maximum height points of high density laser points of training sample plots

Height	Weighted sum
4 – 6 m	548
6 – 8 m	678
8 – 10 m	977
10 – 12 m	1095
12 – 14 m	1737
14 – 16 m	2260
16 – 18 m	2073
> 18 m	1014

The tree height calibration coefficient (C_H) is calculated by the Formula.17:

$$C_H = \frac{\sum C_{R\_low} \times H_{low}}{C_{R\_high} \times H_{high}} \tag{17}$$

Where C_{R_low} and C_{R_high} is the local maximum point representativeness of

local maximum height points from low density laser points and high density laser points, respectively, while H_{low} and H_{high} is the height of local maximum height points from low density and high density laser points, respectively. The look-up table of the tree height calibration coefficient is shown in Table.5

Table 5 Calibration coefficient of tree height

Height	Calibration coefficient
4 – 6 m	1.035
6 – 8 m	1.060
8 – 10 m	1.039
10 – 12 m	1.024
12 – 14 m	1.006
14 – 16 m	0.994
16 – 18 m	0.990
> 18 m	0.993

3.3 Forest AGB estimation

All the sample plots in Area 1 constitute of the test sample plots which are used to evaluate the forest AGB estimation.

The final forest AGB is calculated by the formula.1 (Shendryk et al., 2014) as mentioned earlier. The estimated biomass of the test sample plots from low density laser points is 3.926kg/m^2.

3.4 Accuracy assessment

3.4.1 The accuracy assessment of the local maximum point representativeness look-up table

Due to the time consuming manual identification of individual trees, only one of the four plots in Area 1 is chosen as the test sample plot to evaluate the effectiveness of the calibration. The local maximum height points with a local maximum level of 2 or higher are distributed in Figure 9.

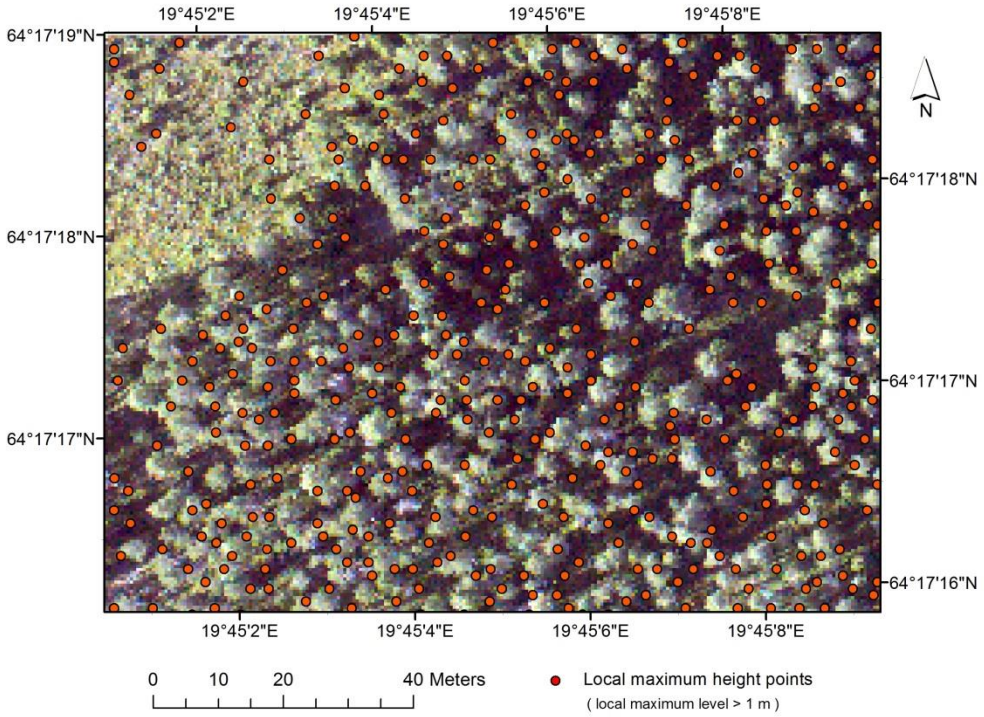


Figure 9 Local maximum points derived from low density LiDAR data of the test sample plot

The tree number distribution weighted by the local maximum point representativeness is shown in Table.6.

Table 6 Weighted sum of local maximum height points of low density laser points of the test sample plot

Height \ Local maximum Level	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4 – 6 m	2	0	0	0	1	0	1	0	0	0	0	0	1	0
6 – 8 m	5	4	2	0	0	0	0	0	0	0	0	0	0	1
8 – 10 m	10.14	6	3	0	1	0	0	0	0	0	0	0	0	0
10 – 12 m	14.63	14	5	0	2	0	0	0	0	0	0	0	0	0
12 – 14 m	25.45	32.90	19	13	2	1	0	0	0	0	0	0	0	0
14 – 16 m	22.23	43.51	27.42	13	14	7	5	3	3	5	1	2	0	1
16 – 18 m	3	6	6	6	1	0	4	3	2	1	2	0	2	12
> 18 m	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Sum	362.28													

The tree number distribution in practice is investigated by combining the local maximum height point data and visualization of the geo-referenced aerial image (Figure 10). In Figure 10, Red dots are local maximum points from low density LiDAR data, while yellow circles denote crowns of trees that are visualized from aerial photograph. Green lines denote links between crowns and supposed tops of trees. Although there's

significant offset of the tree tops after geo-referencing, it is still possible to identify individual trees by visualization. The tree number distribution in practice is shown in Table.7.

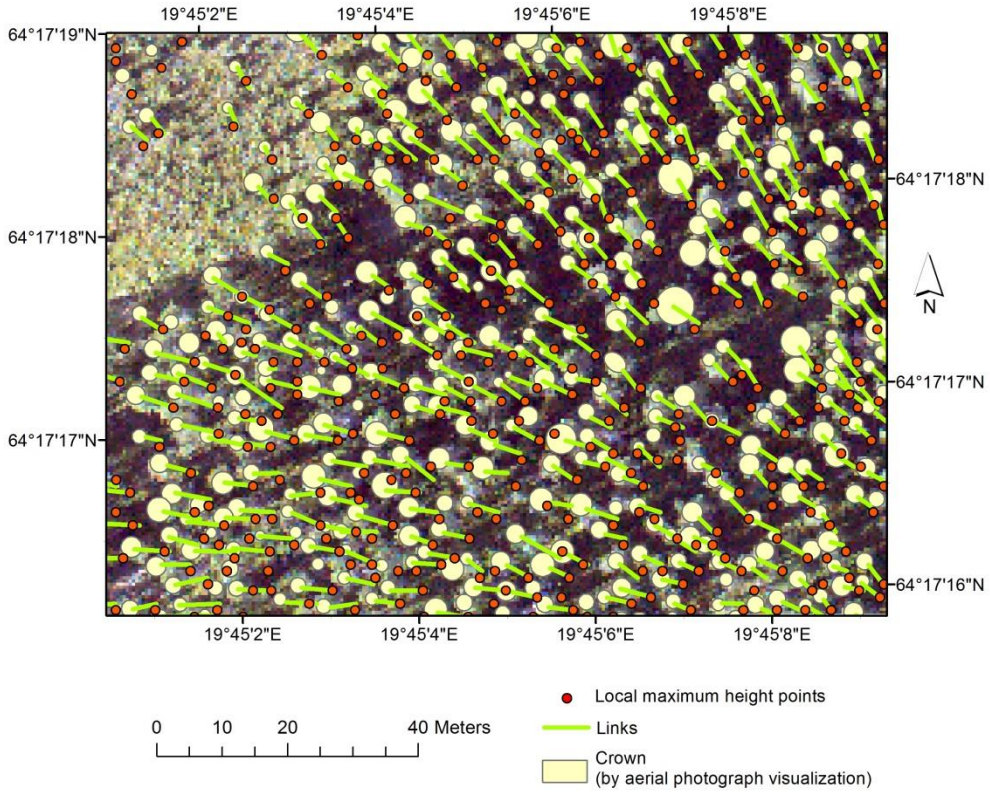


Figure 10 Visualization of corresponding true trees of each local maximum height point of the test sample plot

Table 7 Distribution of the numbers of trees in practice of the test sample plot

Tree \ Local maximum Level	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4 – 6 m	2	0	0	0	1	0	1	0	0	0	0	0	1	0
6 – 8 m	5	4	2	0	0	0	0	0	0	0	0	0	0	1
8 – 10 m	14	6	3	0	1	0	0	0	0	0	0	0	0	0
10 – 12 m	15	14	5	0	2	0	0	0	0	0	0	0	0	0
12 – 14 m	32	30	19	13	2	1	0	0	0	0	0	0	0	0
14 – 16 m	21	42	27	13	14	7	5	3	3	5	1	2	0	1
16 – 18 m	3	6	6	6	1	0	4	3	2	1	2	0	2	12
> 18 m	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Sum	367													

Combining the tree number in each class of the table, a regression is made to verify the effectiveness of the calibration. The accuracy of the calibrated number of trees in

the test sample plot is 98.71%.

3.4.2 The accuracy assessment of the tree height calibration look-up table

In this section, all the sample plots in Area 1 constitute the test sample plot to evaluate the effectiveness of the calibration. The calibrated tree height of the test sample plot which is derived from the low density LiDAR data will be compared with the tree height extracted from high density LiDAR data. The tree height average weighted by local maximum point representativeness will be analyzed as the basis to verify the effectiveness of the tree height calibration.

Calibrated by the tree height calibration look-up table and the local maximum point representativeness table, the weighted average tree height of each height class is calculated by the formula:

$$Avg_{weitgehd} \quad \overline{\overline{H}} \frac{\sum H \times C_R}{\sum C_R} \tag{18}$$

Where C_R means the local maximum point representativeness and H the tree height.The weighted average tree height of the corresponding classes extracted from the high density LiDAR data is also calculated. The weighted average in each class and the error compared with that derived from high density LiDAR data is shown in Table 8.

Table 8 Errors statistics of Weighted average tree height

(The errors statistics of weighted average tree height is made between calibrated low density LiDAR data and high density LiDAR data. “Error” in the table is calculated by subtracting weighted average tree height derived from high density LiDAR data from that from calibrated low density LiDAR data. “Fractional error” means the rate of “Error” to weighted average tree height from high density LiDAR data.)

Height	Calibrated low	High	Error	Fractional error
4 – 6 m	5.31	5.16	0.15	2.91%
6 – 8 m	7.43	7.36	0.07	0.95%
8 – 10 m	9.34	9.35	-0.01	-0.11%
10 – 12 m	11.45	11.39	0.06	0.53%
12 – 14 m	13.26	13.22	0.04	0.30%
14 – 16 m	14.95	14.86	0.10	0.67%
16 – 18 m	16.56	16.53	0.02	0.12%
> 18 m	19.07	19.10	-0.03	-0.16%

In order to evaluate the calibration, a regression model is made to verify the

effectiveness of the calibration. From the regression model we could see the slope of the fitted line is close to 1, the offset is close to 0 and the R2 is 0.9999, also close to 1, which illustrates that the calibration performs well on the test sample plots.

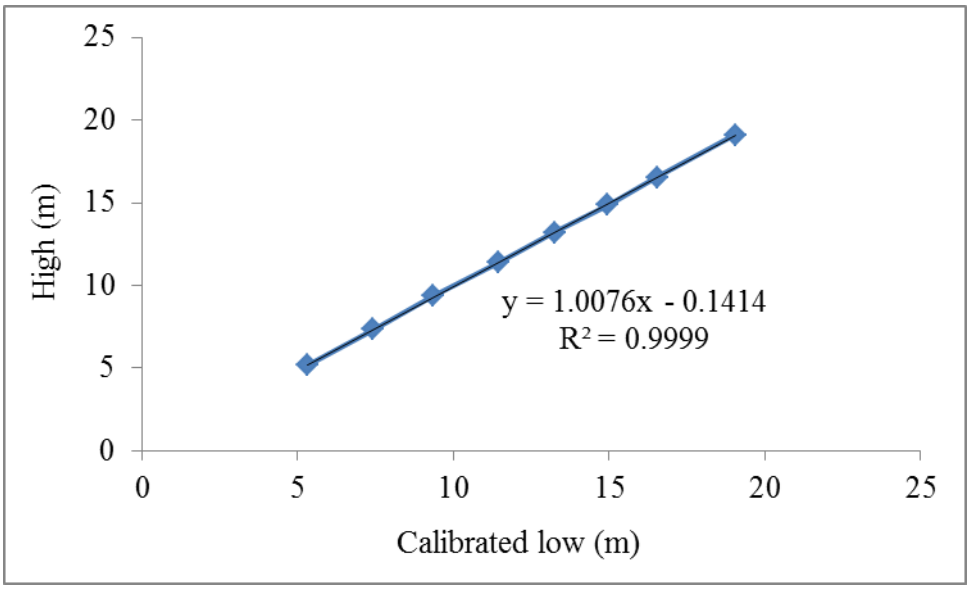


Figure 11 Evaluation of calibrated weighted average tree height derived from low density LiDAR ("High", y axis, denotes the weighted average tree height calculated from high density LiDAR data, while "Calibrated Low" the weighted average tree height from calibrated low density LiDAR data)

3.4.3 The accuracy assessment of forest biomass estimation

The biomass estimated from high density LiDAR data in the test sample plot is 3.915kg/m^2, while from calibrated low density LiDAR data it is 3.926kg/m^2. It is shown that the error of downscaling calibration is of 0.28%.

4. Discussion

4.1 Evaluation of the method

Comparing the calibrated number of trees with the tree numbers derived from visualization, the accuracy of the calibrated number of trees in the test sample plot is 98.71%. Comparing the calibrated tree height derived from low density LiDAR data with the tree height derived from high density LiDAR data, the fractional error of the weighted average tree height in each height level is from 2.91% to -0.16%. Comparing the calibrated biomass estimated from low density LiDAR data with the biomass estimated from high density LiDAR data, the error in the test sample plot is 0.28%. The result shows that the method works well on the test sample plot. The relationships between the extracted forest inventory parameters based on different density LiDAR data can be built to infer the result of forest AGB estimation based on high density LiDAR data with the result based on low density LiDAR data.

However, what has to be considered is that only a few classes which contains more data than the others in the height level-local maximum level table is calibrated. Due to the limitation of labor and cost, the method is evaluated in only one test sample plot. Whether it works well in other plots or other forest sites or not still needs to be verified by further study. It is suggested that more samples in different forest sites should be taken in the future to improve the application scope of the method.

4.2 The combination of point-cloud based identification and downscaling calibration

Both the most commonly used forest AGB estimation algorithms, local maximum algorithm with variable-size window (usually processed by TreeVAW software) and IWS, are all raster data based algorithms. By converting LiDAR data into raster data and pre-processing (e.g. smoothing, interpolation), the reliability of data reduces. A lot of details get lost, a lot of details are presumed arbitrarily. The elevation of the points changes which caused the error of tree height extraction. Some nearby trees are even

fused as one before identification. Compared with the raster based algorithms, point-cloud based identification remains far more details and better accuracy of the LiDAR data, which leads to a higher effectiveness of identification and tree height extraction. However, more details brings more chances of misidentification, for which, a precision guided calibration is essential. Compared with the calibration based on field data, downscaling calibration contains larger amount of data with more diversity. A precisely classified calibration table can be made to provide different calibration strategy for different situation which fits the point-cloud based identification perfectly. This study is a simplified attempt of the combination of the point-cloud based identification and downscaling calibration. Further research about using this combination for LiDAR data based forest inventory is highly recommended.

4.3 Validation method: Aerial image based visualization vs field work

In the planning stage of this study, both aerial image based visualization and field work were considered for validating the individual tree identification. Field work with GPS (Global Positioning System) is the most commonly used method for positioning. The precision of GPS is always a technical barrier in forest inventory. Shaded by the forest canopy, the accuracy of the GPS cannot be guaranteed, especially in the closed forest where the precise tree position is needed to validate the identification. As is attempted by (Shendryk et al., 2014) in the boreal forest of Skogaryd, Sweden, the readouts of GPS accuracy varies from 4 m to 6 m, which cannot support the validation of individual tree identification. Furthermore, as what is estimated in this study is the biomass of a forest in 2008, the field measurement would also endure an effect of time difference. In this study, although aerial image requires geo-matrix correction and presents worse accuracy of positioning, it presents much more reliable topological relationships of the ground features which effectively improves the identification of feature clusters like forest, and provides an eligible result for validation.

However, the resolution of the aerial image is not high enough to distinguish small trees, which limited the range of application. In this study, all the trees lower than 4m is not investigated limited by the precision of validation. Better validation method is

recommended in the further study.

4.4 The influence of the time difference among different datasets

Limited by the availability of data, data with time difference from each other has been used to validate the method in this study. Some of them is from 2008 while some of them is from 2010, which may have influenced the result of this study.

According to the theoretical model mentioned in Section 2.3.2.2, there should always be an underestimation of tree height extraction theoretically. Compared with the high density LiDAR data, low density LiDAR data should perform worse while extracting the tree height due to the lack of point density. However, in the tree height calibration look-up table, the calibration coefficients become less than 1 when dealing with local maximum height points with a height value greater than 14m. This means the extracted height value of the local maximum height points derived from low density LiDAR data has a trend to be greater than that derived from high density LiDAR data in this interval. This phenomenon might be mainly caused by the time difference between the datasets. During this gap, the trees grew higher which might lead to an offset of the underestimated tree height while comparing the extracted tree heights in different point density scale. The individual tree identification might also be affected to become more approximate since trees becomes bigger and easier to identify, which may improve the accuracy of the downscaling calibration. It is strongly suggested that datasets with little time difference can be also tested in the further study to evaluate this influence.

Due to the time difference, the calibration look-up tables in this study are only reliable for calibrating the result derived from low density LiDAR data in 2010 to approximate the result derived from high density LiDAR data in 2008. However, for calibrating the data in different situation, the method of building the look-up table still works.

4.5 About classification

In this study, the classification between Norway spruce and Scot Pine is not

implemented. While implementing the biomass estimation, all the coniferous tree s are considered as Norway spruce. In fact, there are still differences between the height-biomass model of Norway spruce and Scot pine. Although it doesn't influence the accuracy assessment of downscaling calibration, it reduced the accuracy of the final biomass estimation. As this study focuses more on tree top identification and tree height calibration. How to implement the classification of the forest species and how to enhance the biomass estimation with the classified forest is not studied in this thesis. In the further research, better forest classification could be done to improve the method to improve the accuracy and fit much more complex situation. The identification algorithm and biomass estimation model of different species could be combined to expand the scope of application. With better classification, the biomass-height model of each specific species can be involved to enhance the accuracy of estimation.

4.6 About the processing speed

As is mentioned in Section 2.3.1.3, in order to narrow the range of searching, a for loop is nested in another one, which increased the computational complexity of this algorithm to $O(n^2)$. The speed of this algorithm is therefore not fast. Better data indexing could be implemented in the future to simplify the algorithm.

4.7 About downscaling in LiDAR processing

This study attempts to develop a calibration method for LiDAR data from a viewpoint of downscaling. The main idea is building the relationship between different scales but not calibrate the result to fit the field measured data in practice directly. Like topographic map, the LiDAR data in each scale has its own significance to achieve different goals. Mostly, the researchers have to make a balance of size, processing speed, precision and accuracy. By weighing the pros and cons of different scale, different researches are done. To make all the researches united as a system but not isolated, it is suggested that a multi-scale view should be built in the future. This thesis can be regarded as a start of this attempt. In the further research, relationship between other different scales could be built, for instance, between SAR and airborne LiDAR data,

between airborne LiDAR data and terrestrial LiDAR data, or even between some processed scales that fits the standards built in purpose.

5. Conclusion

This study provided a method for coniferous forest aboveground biomass estimation using low density LiDAR data (less than 1 point/m^2). An algorithm based on local maximum point identification is developed, which processes point-cloud data directly without converting the data to raster data. A calibration method is designed from a viewpoint of downscaling. Two calibration look-up tables are provided to approximate the forest AGB derived from low density LiDAR data to the forest AGB derived from high density LiDAR data (more than 1 point/m^2).

The result shows that both the tree number calibration look-up table and the tree height calibration look-up table work well in the test sample plot as expected. The calibrated forest aboveground biomass derived from low density LiDAR data is highly close to the forest aboveground biomass derived from high density LiDAR data.

With respect to the first hypothesis, the relationships between the extracted forest inventory parameters based on different density LiDAR data can be built to infer the result of forest AGB estimation based on high density LiDAR data (more than 1 point/m^2) with the result based on low density LiDAR data (less than 1 point/m^2) with an acceptable accuracy. With respect to the second hypothesis, algorithm based on local maximum height point identification directly can be used to estimate the coniferous forest aboveground biomass using low density airborne LiDAR data (less than 1 point/m^2).With respect to the third hypothesis, downscaling calibration can be used to approximate the forest aboveground biomass estimated from low density LiDAR data (less than 1 point/m^2) to the forest aboveground biomass estimated from high density LiDAR data (more than 1 point/m^2).

The combination of point-cloud based identification and downscaling calibration is convinced to be a nice match. Point-cloud based identification remains most details of the LiDAR data, which enhanced the identification and requires more precisely guided calibration as well. With the great amount of samples provided by downscaling analysis, a precision guided calibration becomes possible. It is highly recommended

that this combination could be used by other researchers in the future.

A lot of further research could be done to improve the method. Species classification and biomass models for more species could be involved to expand the application scope of the method. Better data index could be designed to improve the processing speed.

Above all, this study is an attempt to estimate the forest AGB with LiDAR data from a viewpoint of downscaling. Balanced by size, processing speed, accuracy and precision, the LiDAR processing in each scale has its irreplaceable significance. To make all the researches united as a system but not isolated, it is strongly suggested that a multi-scale view for LiDAR data processing should be built in the future.

Reference

Ågren, A. M., W. Lidberg, and E. Ring. 2015. Mapping temporal dynamics in a forest stream network—Implications for riparian forest management. *Forests* 6. Multidisciplinary Digital Publishing Institute: 2982–3001. DOI:10.3390/f6092982

Asner, G. P., J. Mascaro, H. C. Muller-Landau, G. Vieilledent, R. Vaudry, M. Rasamoelina, J. S. Hall, and M. van Breugel. 2012. A universal airborne LiDAR approach for tropical forest carbon mapping. Oecologia 168. Springer: 1147–1160. DOI:10.1007/s00442-011-2165-z

Brown, S. 2002. Measuring carbon in forests: current status and future challenges. Environmental pollution 116. Elsevier: 363–372. DOI:10.1016/S0269-7491(01)00212-3

Clark, M. L., D. A. Roberts, J. J. Ewel, and D. B. Clark. 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sensing of Environment 115. Elsevier: 2931–2942. DOI:10.1016/j.rse.2010.08.029

Cohen, R., J. Kaino, J. A. Okello, J. O. Bosire, J. G. Kairo, M. Huxham, and M. Mencuccini. 2013. Propagating uncertainty to estimates of above-ground biomass for Kenyan mangroves: A scaling procedure from tree to landscape level. Forest Ecology and Management 310. Elsevier: 968–982. DOI:10.1016/j.foreco.2013.09.047

Drăguț, L., and T. Blaschke. 2006. Automated classification of landform elements using object-based image analysis. Geomorphology 81. Elsevier: 330–344. DOI:10.1016/j.geomorph.2006.04.013

Du, L., T. Zhou, Z. Zou, X. Zhao, K. Huang, and H. Wu. 2014. Mapping forest biomass using remote sensing and national forest inventory in China. *Forests* 5. Multidisciplinary Digital Publishing Institute: 1267–1283. DOI:10.3390/f5061267

Edson, C., and M. G. Wing. 2011. Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements. Remote Sensing 3. Molecular Diversity Preservation International: 2494–2528. DOI:10.3390/rs3112494

Englhart, S., V. Keuck, and F. Siegert. 2011. Aboveground biomass retrieval in tropical forests—The potential of combined X-and L-band SAR data use. Remote sensing of environment 115. Elsevier: 1260–1271. DOI:10.1016/j.rse.2011.01.008

Falkowski, M. J., A. M. S. Smith, P. E. Gessler, A. T. Hudak, L. A. Vierling, and J. S. Evans. 2008. The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. Canadian Journal of Remote Sensing 34. Taylor & Francis: S338–S350. measurement algorithms using lidar data. DOI:10.5589/m08-055

Falkowski, M. J., J. S. Evans, S. Martinuzzi, P. E. Gessler, and A. T. Hudak. 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. Remote Sensing of Environment 113. Elsevier: 946–956. DOI:10.1016/j.rse.2009.01.003

Goodchild, M. F., and D. A. Quattrochi. 1997. Scale, multiscaling, remote sensing, and GIS. CRC Press.

Hajnsek, I., R. Scheiber, L. Ulander, A. Gustavsson, G. Sandberg, S. Tebaldini, A. M. Guarnieri, F. Rocca, et al. 2009. Technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2007 experiment. ESA, Paris, France, Tech. Rep 22052.

Hay, G. J., D. J. Marceau, P. Dube, and A. Bouchard. 2001. A multiscale framework for landscape analysis: object-specific analysis and upscaling. Landscape Ecology 16. Springer: 471–490.

DOI:10.1023/A:1013101931793

Jensen, J. R. 2005. Introductory digital image processing 3rd edition. In Upper saddle river: Prentice hall.

Kini, A. U., and S. C. Popescu. 2004. TreeVaW: a versatile tool for analyzing forest canopy LIDAR data: A preview with an eye towards future. Kansas city, Mossouri, ASPRS.

Laudon, H., M. Berggren, A. Ågren, I. Buffam, K. Bishop, T. Grabs, M. Jansson, and S. Köhler. 2011. Patterns and dynamics of dissolved organic carbon (DOC) in boreal streams: the role of processes, connectivity, and scaling. Ecosystems 14. Springer: 880–893. DOI:10.1007/s10021-011-9452-8

Le Toan, T., S. Quegan, M. W. J. Davidson, H. Balzter, P. Paillou, K. Papathanassiou, S. Plummer, F. Rocca, et al. 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote sensing of environment 115. Elsevier: 2850–2860. DOI:10.1016/j.rse.2011.03.020

Mette, T., and others. 2007. Forest biomass estimation from polarimetric SAR interferometry. Technische Universität München.

Naesset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing 52. Elsevier: 49–56. DOI:10.1016/S0924-2716(97)83000-6

Owemyr, P., and J. Lundgren. 2010. Noggrannhetskontroll av laserdata för ny nationell höjdmodell. (In Swedish, English summary)

Papathanassiou, K. P., T. Mette, I. Hajnsek, and A. Moreira. (2005): Polarimetric SAR Interferometry: Potential and Limitations for Biomass Estimation. Retrieved May 23, 2016, from http://www.eorc.jaxa.jp/ALOS/kyoto/feb2005/pdf/9-KC6_Papathanassiou.pdf

Peralta-Tapia, A., R. A. Sponseller, A. Ågren, D. Tetzlaff, C. Soulsby, and H. Laudon. 2015. Scale-dependent groundwater contributions influence patterns of winter baseflow stream chemistry in boreal catchments. Journal of Geophysical Research: Biogeosciences 120. Wiley Online Library: 847–858. DOI:10.1002/2014JG002878

Persson, H., and J. E. S. Fransson. 2014. Forest variable estimation using radargrammetric processing of TerraSAR-X images in boreal forests. Remote Sensing 6. Multidisciplinary Digital Publishing Institute: 2084–2107. DOI:10.3390/rs6032084

Pittman, S. J., and K. A. Brown. 2011. Multi-scale approach for predicting fish species distributions across coral reef seascapes. PloS one 6. Public Library of Science: e20583. DOI:10.1371/journal.pone.0020583

Popescu, S. C., and R. H. Wynne. 2004. Seeing the trees in the forest. Photogrammetric Engineering & Remote Sensing 70. American Society for Photogrammetry and Remote Sensing: 589–604. DOI:10.14358/PERS.70.5.589

Seavy, N. E., J. H. Viers, and J. K. Wood. 2009. Riparian bird response to vegetation structure: a multiscale analysis using LiDAR measurements of canopy height. Ecological Applications 19. Wiley Online Library: 1848–1857. DOI:10.1890/08-1124.1

Seidl, R., T. A. Spies, W. Rammer, E. A. Steel, R. J. Pabst, and K. Olsen. 2012. Multi-scale drivers of spatial variation in old-growth forest carbon density disentangled with Lidar and an individual-based landscape model. Ecosystems 15. Springer: 1321–1335. DOI:10.1007/s10021-012-9587-2

Sessa, R., and H. Dolman. 2008. Terrestrial Essential Climate Variables for Climate Change Assessment, Mitigation and Adaptation. Retrieved May 23, 2016, from

http://www.fao.org/gtos/doc/pub52.pdf

Shendryk, I., M. Hellström, L. Klemedtsson, and N. Kljun. 2014. Low-density LiDAR and optical imagery for biomass estimation over boreal forest in Sweden. Forests 5. Multidisciplinary Digital Publishing Institute: 992–1010. DOI:10.3390/f5050992

Zawada, D. G., and J. C. Brock. 2009. A multiscale analysis of coral reef topographic complexity using lidar-derived bathymetry. Journal of Coastal Research. Coastal Education and Research Foundation: 6–15. DOI:10.2112/SI53-002.1

Zhao, K., S. Popescu, and R. Nelson. 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. Remote Sensing of Environment 113. Elsevier: 182–196. DOI:10.1016/j.rse.2008.09.009

Ziegler, M., H. Konrad, J. Hofrichter, A. Wimmer, G. S. Ruppert, M. Schardt, and J. M. Hyyppae. 2000. Assessment of forest attributes and single-tree segmentation by means of laser scanning. In AeroSense 2000, 73–84. DOI:10.1117/12.397780

Appendices

Department of Physical Geography and Ecosystem Science, Lund University

Lund University GEM thesis series are master theses written by students of the international master program on Geo-information Science and Earth Observation for Environmental Modelling and Management (GEM). The program is a cooperation of EU universities in Iceland, the Netherlands, Poland, Sweden and UK, as well a partner university in Australia. In this series only master thesis are included of students that performed their project at Lund University. Other theses of this program are available from the ITC, the Netherlands (www.gem-msc.org or www.itc.nl).

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 2013. The complete list and electronic versions are also electronic available at the LUP student papers (https://lup.lub.lu.se/student-papers/search/) and through the Geo-library (www.geobib.lu.se).

1 Soheila Youneszadeh Jalili (2013) The effect of land use on land surface temperature in the Netherlands
2 Oskar Löfgren (2013) Using Worldview-2 satellite imagery to detect indicators of high species diversity in grasslands
3 Yang Zhou (2013) Inter-annual memory effects between Soil Moisture and NDVI in the Sahel
4 Efren Lopez Blanco (2014) Assessing the potential of embedding vegetation dynamics into a fire behaviour model: LPJ-GUESS-FARSITE
5 Anna Movsisyan (2014) Climate change impact on water and temperature conditions of forest soils: A case study related to the Swedish forestry sector
6 Liliana Carolina Castillo Villamor (2015) Technical assessment of GeoSUR and comparison with INSPIRE experience in the context of an environmental vulnerability analysis using GeoSUR data
7 Hossein Maazallahi (2015) Switching to the "Golden Age of Natural Gas" with a Focus on Shale Gas Exploitation: A Possible Bridge to Mitigate Climate Change?
8 Mohan Dev Joshi (2015) Impacts of Climate Change on *Abies spectabilis*: An approach integrating Maxent Model (MAXent) and Dynamic Vegetation Model (LPJ-GUESS)
9 Altaaf Mechiche-Alami (2015) Modelling future wheat yields in Spain with LPJ-GUESS and assessing the impacts of earlier planting dates
10 Koffi Unwana Saturday (2015) Petroleum activities, wetland utilization and livelihood changes in Southern Akwa Ibom State, Nigeria: 2003-2015
11 José Ignacio Díaz González (2016) Multi-objective optimisation algorithms for GIS-based multi-criteria decision analysis: an application for evacuation planning
12 Gunjan Sharma (2016) Land surface phenology as an indicator of performance of conservation policies like Natura2000
13 Chao Yang (2016) A Comparison of Four Methods of Diseases Mapping

14 Xinyi Dai (2016) Dam site selection using an integrated method of AHP and GIS for decision making support in Bortala, Northwest China
15 Jialong Duanmu (2016) A multi-scale based method for estimating coniferous forest aboveground biomass using low density airborne LiDAR data