ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

FOREST PARAMETER ESTIMATION FROM AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS AREAS

Fang Fang

Department of Urban and Environmental Engineering
(Environmental Science and Engineering)

Graduate school of UNIST

2014

FOREST PARAMETER ESTIMATION FROM AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS AREAS

Fang Fang

Department of Urban and Environmental Engineering
(Environmental Science and Engineering)

Graduate school of UNIST

2014

ULSAN NATIONAL INSTITUTE OF
SCIENCE AND TECHNOLOGY

FOREST PARAMETER ESTIMATION FROM AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS AREAS

A thesis
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Fang Fang

06.09.2014

Approved by

Advisor
Jungho Im

FOREST PARAMETER ESTIMATION FROM AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS AREAS

Fang Fang

This certifies that the thesis of Fang Fang is approved.

06.09.2014

Jungho Im
Assistant Professor of Urban and Environmental Engineering
Ulsan National Institute of Science and Technology
Thesis Committee Chairman

Myong-In Lee
Associate Professor of Urban and Environmental Engineering
Ulsan National Institute of Science and Technology
Thesis Committee Member

Dong Keun Yoon
Assistant Professor of Urban and Environmental Engineering
Ulsan National Institute of Science and Technology
Thesis Committee Member

FOREST PARAMETER ESTIMATION FROM AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS AREAS

Fang Fang

ABSTRACT

During the past decade, the procedure for quantification of forest parameters using LiDAR data has been rapidly improved. Among various forest parameters, biomass is the paramount in understanding the potentials productivity of forests. Various methods have been developed to estimate biomass at both plot and individual tree levels. In order to quantify biomass at the individual tree level, tree crown delineation must be conducted, which is sometimes challenging especially for multi-layer dense forests in rugged mountainous areas. In this study, Light Detection and Ranging (LiDAR) data were used to delineate tree crowns and estimate biomass in a mountainous forest. Firstly, a novel algorithm was proposed to identify individual tree crowns using the concept of live crown ratios based solely on LiDAR data. Then, above ground biomass (AGB) was estimated using machine learning approaches based on tree crowns delineated in the previous step. LiDAR-derived metrics related to forest parameters such as tree height and crown areas as well as topographic characteristics extracted based on the delineated tree crowns were used to estimate AGB. Three machine learning models—random forest, Cubist, and support vector regression—were evaluated for AGB estimation and relative importance of input variables was examined.

Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Forests, or wooded area, are the place with a high density of trees. All the trees in the forest vary significantly in size, species, structure and living conditions. Nearly 10 percent of the earth's surface is covered by tree forest. More importantly, forests provide the living space and materials for many organisms; in addition, forest offers oxygen for human beings. Thus, forests act as one of the most important roles of the biosphere. In this case, numerous forest studies are attempting to fully understand the forests' impact on ecosystems, biodiversity, and the environment. Among these forest studies, biomass and above ground biomass (AGB) has drawn lots of attention since biomass estimation is necessary for the forests management and works as a common criteria for evaluating the quality and productivity of forests. AGB can be derived from in situ or remote sensing data. Compared with the former, remote sensing data can reach every section of the ground. In this study, we are trying to estimate forest biomass from LiDAR (Light detection and ranging) data. LiDAR data are optimum for biomass estimation since it provides physical structure and characteristic of canopy surface with high spatial resolution. This research was based on LiDAR derived images to delineate individual tree and biomass estimation. Various methods can be built to derive the regression relationship between biomass and other variable metrics like diameter at breast and canopy height based on LiDAR data. Recently, besides the regression using the LiDAR derived height metrics, there are increasing methods focusing on individual tree level biomass estimation and machine learning. Small scale or individual tree level based biomass estimation requires high quality LiDAR data and several individual tree based forest parameters. In this study biomass estimation is based on individual tree level.

In order to improve the accuracy of the biomass result, the first step is individual tree crown delineation. Section 2 described this topic in detail and introduced a novel method. Since all the algorithms for crown delineation have their own strength and weakness. Various studies are created and improved the methods to detect individual trees' location and proper crown boundary. Individual tree crown delineation has successfully combined digital image processing techniques and forest concepts together. We also aim to create a novel algorithm based on traditional methods for image processing through adding a new threshold and concept related to tree structure. Results were promising and of high accuracy, which gives the basis for biomass estimation in section 3. In this case, section 3 introduced the methods and results for biomass estimation based on the output from individual tree crown delineation. Within these tree crowns, individual tree height and other kinds of variables are extracted for biomass estimation. Three kinds of machine learning algorithms: random forest, support vector machine and Cubit were adopted in section 3 to predict biomass. The variables selection took both tree structure and terrain condition into consideration. Section 3 aims at finding out the most relevant variables for estimating biomass and built assumption to determine the variables in advance. The assumption is that tree structure especially the ratio of crown height to tree height, and the terrain condition should have greater importance than other commonly used variables. The result showed that our assumption was validated and the three machine learning approaches were compared based on their performance.

In section 2, a novel individual tree crown delineation method was introduced. In section 3, individual tree biomass estimation based on machine learning was presented. Conclusion was derived summarizing the two sections in section 4, and outlines future possible research.

2. A NOVEL TREE CROWN DELINEATION METHOD BASED ON AIRBORNE LIDAR DATA IN RUGGED MOUNTAINOUS FORESTS

Abstract

Individual tree crown delineation works as the basic unit for further studies of like forest biomass estimation. Canopy Height Model (CHM) from LiDAR data can be used for individual tree crown delineation. However, the complexity of forest structure such as the branches and clusters makes it more challenging to improve the accuracy of tree crown delineation. What is more, the boundary of tree crowns especially in LiDAR data should be clearly defined. This study aimed to develop an optimum approach for tree crown delineation in mixed forests with multiple stories based on an improved watershed algorithm and compared with region growing algorithm. Different from previous studies, the novel approach was focused on the fake tree crowns exclusion based on morphological analysis and adaptive tree boundary detection based on rule-based improved watershed segmentation. A new criterion was presented for locating the exact boundary of each tree crown. These proposed approaches were evaluated by using LiDAR data which collected over forest area in Gangwon, S. Korea in September, 2013. The automated method correctly delineated 87 percent of the tree crowns in the average based on the accuracy assessment.

2.1 Introduction

Forest ecosystems interact with the atmosphere, soil, and water systems in the Earth's surface and play an important role as a repository of terrestrial biological diversity. In order to manage forests in a sustainable way in changing climate conditions, it is necessary to monitor key forest parameters such as basal areas, biomass, and carbon stocks from local to global scales. Although traditional field surveying can accurately measure such forest parameters, it is time consuming and labor intensive, and sometimes access to certain areas is simply impossible. Extracting forest parameters from remote sensing data is becoming more necessary and efficient other than field survey. Remote sensing data can be used to estimate such forest characteristics at the tree or stand levels (Moffiet, Mengersen, Witte, King, & Denham, 2005; Kwak, Lee, Lee, Biging, & Gong, 2007; Chambers et al., 2007; Jung, Kwark, Park, Lee, & Yoo, 2011; Straub & Koch, 2011; Yao et al., 2011; Koukoulas & Blackburn, 2005; Yang et al., 2013). A basic unit for monitoring forest parameters and structure is individual tree crowns. For example, mapping of forest communities can base in individual tree crowns (Bunting, Lucas, Jones, & R. Bean, 2010). In this case various studies have focused on individual tree crown detection. Many studies have proposed algorithms to detect and delineate tree crowns from remote sensing data (Colgan, Baldeck, Féret, & Asner et al., 2012; Hu et al., 2014; Leckie et al., 2005; Bunting & Lucas, 2006; Hirschmugl, Ofner, Raggam, & Schardt, 2007; Horváth, Jermyn, Kato, & Zerubia, 2009, Vastaranta et al., 2012). Six individual tree crown detection algorithms are compared and evaluated using image data set (Larsen et al., 2009). In addition, Ke and Quackenbush (2011) described the existing algorithm and improvements for individual tree crown detection and delineation. For example, the existing and commonly applied individual tree crown delineation methodology such as valley following (Gougeon, 1995), edge detection (Popescu et al., 2003), watershed segmentation (Jing et al., 2012, Hu et al., 2014), and 3D modeling (Gong et al., 2002). Because of the complexity of the forest with various tree species inside, these methods show promising result at the same time some issues remained. For instance, various species inside will lead to different size of trees. Some small trees near the large trees are easily omitted. Density area also makes problem since tree tops are too close to be distinguished. The crown boundaries and between crown valleys are also hard to be detected exactly both win the high density and low density area.

Among all the remote sensing data, LiDAR derived data of high resolution has been widely used for forest related studies (Ene, Næ sset & Gobakken 2012; Kaartinen et al., 2012). In particular, LiDAR remote sensing is useful for characterizing 3D structure of individual trees and forest stands (Castillo-Núñez et al., 2011; Véga & Durrieu, 2011; Hu, Li, Jing, & Judah, 2014; Reitberger, Schnörr, Krzystek, & Stilla, 2009; Li, Hu, & Noland, 2013; Brandtberg, A. Warner, E. Landenberger, & B. McGraw, 2003). Individual tree crowns have been delineated using high resolution and high point density LiDAR data (Hu et al., 2014; Zhang, Quackenbush, Im, & Zhang, 2012; Clark, Roberts, & Clark, 2005). For LiDAR derived data, a number of methods have been developed for crown delineation. Most of the methods are trying to detect the tree top and crown boundary based on the height information from LiDAR data. Canopy height models (CHM), typically calculated using digital surface models (DSM) and digital terrain (ground) models (DTM), have been often used for delineating tree crowns. Two algorithms for crown delineation are most common in LiDAR data. First, watershed segmentation tries to invert CHM and then find the local minima as the starting point for reaching the boundary of tree crowns. Valley of the inverted CHM image can be treated as basins. The crown boundary using watershed segmentation is created by filling up the basin and dams are created between two adjacent areas (Jing et al., 2012, Hu et al., 2014). Marker controlled watershed segmentation has been further applied after marker selection in LiDAR derived images (Wang, Gong, & Biging, 2004; Chen, Baldocchi, Gong, & Kelly, 2006). Second, region growing is another algorithm commonly used for individual tree crown delineation in LiDAR data. This algorithm starts from the local maxima (i.e., seeds) from CHM and by changing the relationship between the seeds and surrounding pixels to determine whether the surrounding pixels are included in the corresponding crown. Improvements can be developed for the seed selection process and threshold identification by examining the surrounding pixels (Colgan et al., 2012). In conclusion, the key issues for crown delineation are: tree top detection and crown boundary detection (Ke & Quackenbush, 2011). For tree top detection, finding an optimum scale for searching local maxima is the main problem. For example, Jing et al. (2012) proposed a multi-scale segmentation method for individual tree crown delineation. Multi-scale selection was applied based on erosion and dilation operation, and then boundary-refined watershed segmentation results were derived. The final segmented image was integrated with different tree crown sizes., Several approaches have been developed to improve tree boundary by controlling the shape of crowns, including threshold-based methods (Hu et al., 2014) and angle-based shape refinement (Pouliot et al., 2002).

While crown segmentation methods have been improved in many ways, there are still several problems that should be further investigated to get better crown delineation results. First, some fake tree crowns appear even after the selection of optimum scale for segmentation and tree top detection. Tree top detection is closely related to segmentation scale. Typically, a segmentation scale is chosen manually through a trial and error approach. While a large segmentation scale may omit some small tree tops, a smaller scale typically leads to over segmentation since too many local maxima are detected as tree tops. More recently, a multi-scale approach based on morphological opening operations works better for high spatial resolution imagery (Jing et al., 2012). Morphological analysis, like opening operation helps to remove the background smaller than dominant size of crown and determine the best scale for image segmentation. Pouliot (2002) found that there was a strong positive relationship between crown size and tree height. Future studies should focus on searching local maxima to effectively avoid omission and over-detection of tree tops.

Secondly, tree boundary searching for tree crown delineation still needs more effort and concentration. Some researchers have already tried to make crown boundary more circular using angle removal and split of too large segments. However, with the assumption that tree crowns should be crown liked shape, or nearly circular shape, angle removal was based on the degree threshold of the crown boundary, and such removal may destroy the actual boundary of crowns. More precise definition for crown boundary should be given. Although the shape of crown outline is important, the revised crown outlines should not only take circularity into consideration. The boundary detection should base on the structure of tree. A novel approach for boundary improvement is necessary to search for the edge of the tree crowns based on tree structure automatically.

The third issue is that most of existing algorithms for tree crown delineation require many user-specified parameters to run. In that case, the result of the algorithm is time-consuming and needs more trials. More study should attempt to minimize the number of manually determined parameters.

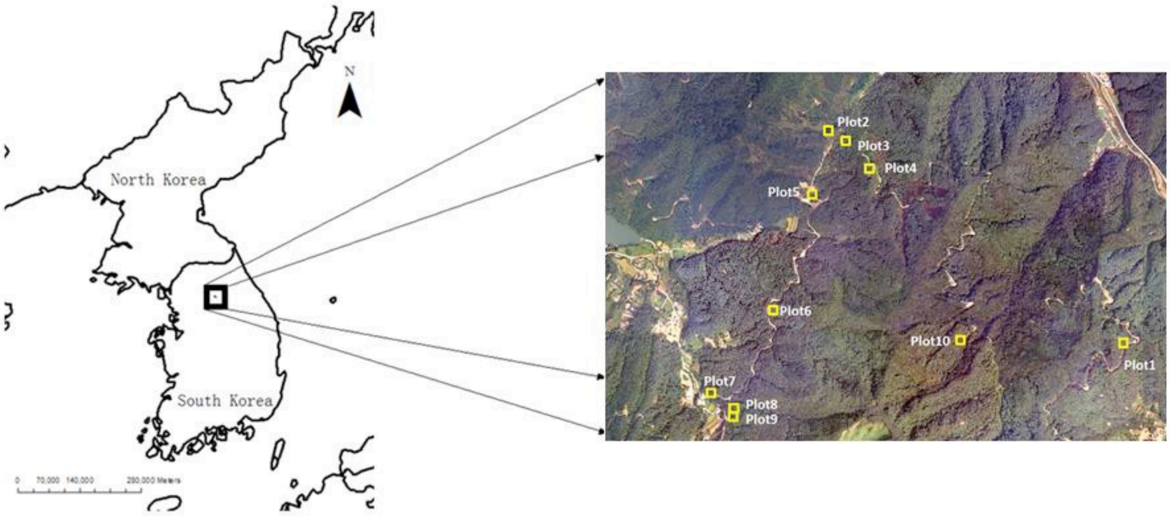
This study aims to propose novel tree crown delineation methods that can be applied to rugged mountainous forests using airborne LiDAR data. The objectives of this study were to 1) develop a novel tree crown delineation approach using a live crown ratio concept, 2) assess the novel method to delineate tree crowns from LiDAR data in a rugged mountainous area , and 3) compare its performance with region-based segmentation approaches.

2.2 Study area and data

The study area is Maehwa Mountain, located within Gangwon province, South Korea showed in Fig 2.1. Ten plots were selected with the diameter of 22.8m and the location of each plot was showed in Fig 2.1. Some species in each plot are homogenous and some are heterogeneous. The number of trees varies from plot to plot. The recorded variables are coordinates, DBH, tree height, and so on. The study area contains various sizes of trees, shrubs and grasses. The main tree species in the study plot are *Pinus densiflora* and *Larix kaempferi.* Ground data were collected during August, 2013.

This study used discrete multiple-return LiDAR from an airborne CESSNA 208 CARAVAN at the altitude of 2300 meters. The camera was DMC 01-125 with the lens of 120mm. The original data were digital elevation model (DEM) and digital surface model (DSM). Both were unfiltered and discrete images of the whole mountainous area. Image cropping was applied as the first step to extract the ten plots and image smoothing was applied in Matlab R2012a.

For the ten plots, the conditions of the tress are various. For example, the trees in plot 3 are much taller than the others. However plot 3 has the least amount of the tree. On the other hand, plot 1,2,4,7 have nearly 30 trees inside with high density.



a. b.

Fig 2.1 (a) The study area and (b) Sample plots in the study area. Each plot is with the diameter of 22.8m.

For individual tree crown delineation, the reference crowns were generated using two kinds of data. One is the data from ground survey, the other one is the data from LiDAR. The data from ground survey provides the accurate location of each tree. Reference tree crowns were generated in ArcMap 10 with the smoothed CHM and aerial photo. Crowns were delineated by experienced researchers based on the location from survey. All the references also reconfirmed with the 3D map of CHM to find the boundaries. Totally 198 trees were delineated as the reference boundary for crowns. The delineation tried to match with the location from field data in order to get more accurate reference crowns.

2.3 Methods

2.3.1 Improved watershed segmentation process using crown ratio

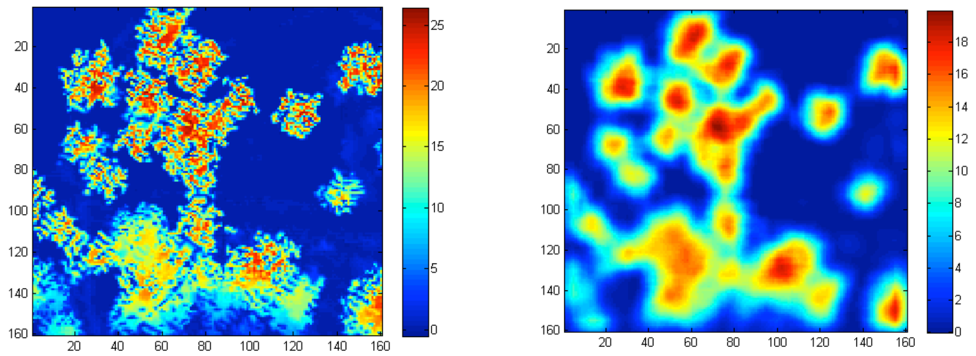
A novel approach for tree crown delineation improvement was described in this paper. In order to prove the effectiveness and advantages of this algorithm, region growing segmentation which used as comparison also applied in this research. The tree crown delineation proposed in this study consists of four steps: 1) image filtering using the Gaussian function with a scale selection process; 2) automated process to distinguish trees from the background; 3) traditional watershed segmentation aiming to find the original boundary of each segment using a specific window size for each plot; 4) determination of true boundary through shrinking from the result of using the concept of crown ratio. This method was compared with an improved region growing method. Each step of the proposed method was explained in detail below.

(1) Prepossessing and image filtering

In this study, 10 plots were used for tree crown delineation. Digital surface models (DSM) and digital elevation models (DEM) were produced from the original LiDAR point cloud data. Canopy height models (CHM) were derived based on the difference of DSM and DEM. All pixels in CHM are considered to represent height information of materials on the surface. Trees can be detected according to the height information. Each study plot was circular with the diameter of 22.8 meters. For tree crown delineation, all the plots were extracted from the CHM as squares with the side of 40meters with buffers. Reference data for tree crown delineation were derived manually through the overlapped area from 2 professional researchers using 3D surface of the smoothed CHM at various scales.

To delineate tree crowns, each plot was first smoothed with Gaussian filtering as prepossessing for image segmentation. A window size for Gaussian filtering was determined by the number of trees in the plot. While more trees in a plot lead to a relative small window size, fewer trees in the plot result in a larger size. An assumption here is that within the same area of the plot, if there are more trees in the plot each tree area is smaller than that in the plot with fewer trees. In this study, three kinds of window sizes were used as a small (no larger than 5*5), middle (6*6 and 7*7) and large sizes (9*9 and 10*10).

Gaussian filtering was applied after the scale selection. This process was widely used for tree crown delineation since it serves as the smoothing algorithm with weighted average values using neighbors of each pixel. For CHM in this research, the original images contain many small branches, peaks and valleys. Gaussian filtering can successfully smooth the small peaks and valleys as the low pass filter. After choosing the proper window size, a two-dimension Gaussian filtering with the window size was created and sigma, which determines the smooth degree, was set at 0.7 pixels. Since tree crowns were considered as the similar shape with the Gaussian filter kernel, the object similar to the size of the kernel will be enhanced and smaller part peaks and valleys will be removed. Examples of CHM images before and after Gaussian filtering are presenting in Fig 2.2.



a. **b.**

Fig 2.2 Original CHM and filtered CHM image.

(2) Background removal

In order to enhance the difference between tree crowns with the background including the shrubs, areas with the height less than a certain value are regarded as bare ground and set at 0 meter after the filtering. This can effectively remove small portions of the plot like shrubs. Shrubs and blobs are removed which can obviously reduce the amount of useless segments and calculation for the watershed algorithm. Comparisons of watershed segmentation before and after the background removal using plot 5 are shown in Fig 2.3. The figure in the left shows more segments than the right one since some of the peaks of shrubs are also regarded as tree tops. Instead of setting the threshold manually(Hu et al., 2014), automated Otsu's threshold selection for background removal was applied in each plot through Matlab. This removal process is less subjective since each plot has different condition of background removal.

a. **b.**

Fig 2.3 Comparison before and after the background removal for watershed segmentation.

(3) Watershed segmentation

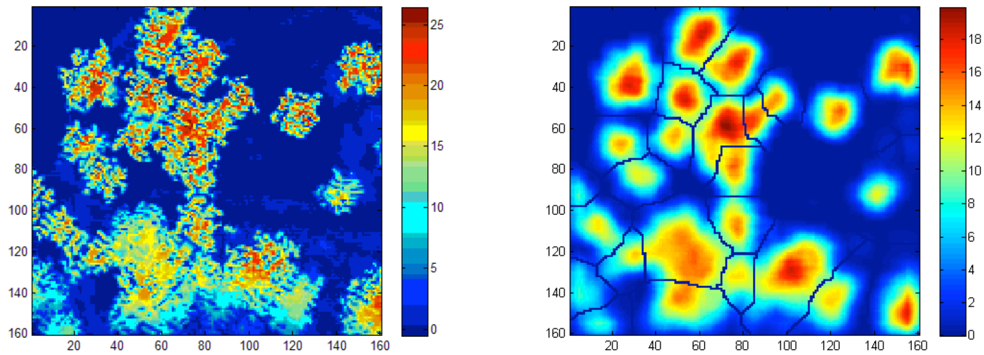
The watershed algorithm is commonly used for segmentation. The algorithm was firstly introduced in 1979 by S. Beucher and C. Lantuejoul (Beucher & Lantué., 1979). A grey scale image can be interpreted as a topographic object, where the grey level of the pixel can be treated as its altitude in the object. Water can fill from each local minimum until the flood meets to build the dam between two adjacent catchment basins. These dams are the boundaries of the objects. This process can be interpreted as watershed segmentation. In most cases, watershed segmentation is usually used for images of elevation information.

In Matlab, there are three kinds of watershed segmentation: watershed segmentation using the distance transforms, watershed segmentation using gradients, and marker controlled watershed segmentation. First, watershed segmentation using distance transform has a relatively simple explanation. The "distance" is that the distance from every pixel to the nearest non-zero value pixel. For tree crown delineation, a binary image is created, with the canopy as values of one and background set the value of 0. Distance transform is used to calculate the distance from each non-zero pixel to the nearest zero pixel. This distance helps to determine the tree top of the segment and then marker controlled watershed segmentation was applied for crown delineation boundaries. In the second place, watershed segmentation using gradients is usually used as the pre-process in the gray scale image. The assumption is that gradient magnitude object images should have relatively high value around the edges and other pixels have a relatively low value. The ideal watershed segments boundaries are made along the object edges. However, the watershed segmentation using a gradient method is usually applied after image smoothing and filtering. This process obviously reduces the over-segmentation problem.

Marker controlled watershed segmentation is the most widely used method in crown delineation. For original watershed segmentation, the most severe problem is over segmentation. Marker controlled watershed segmentation is aiming to do some preprocessing to reduce the problem. The basic assumption of this method is based on the concept of markers. A marker means the connected component in the original image. Both objects and background areas in the image should have a set of markers. The markers of objects are called internal markers while the background markers called external markers. The issue worth to be mentioned is that there are various ways to generate the markers, such as linear filtering, nonlinear filtering, or morphological process like image opening operation and close operation. All these processes are trying to the image more smoothed to contain less useless peaks and valleys. The optimum amount of markers is generated and watershed segmentation will avoid over-segmentation. For example, local maxima in CHM image with different scales can be used as the process to generate markers (Jing et al., 2012; Chen et al., 2006; Kaartinen et al, 2012; Vastaranta et al., 2011). After watershed segmentation using different markers, all the images were integrated as one image. A morphological process is also used before watershed segmentation to generate markers (Jing et al., 2012).

In this study marker controlled watershed segmentation was used for creating boundaries. The algorithm starts from an inverted CHM image. All the grey value in the image stands for the height information. The local maximum is inverted and treated as the local minimum in order to fill up the basins. For watershed segmentation, from each local minimum, flood will fill the basin and barriers will derive when the flood from different local minimum was reached together. The area within each barrier can be treated as the basin and the embankments were created. For this study, the basins equal to the tree crowns. The embankments are tree crown boundaries.

The segmentation results after filtering are shown in Fig 2.4. To avoid the over-segmentation problem, the first two steps worked as the noise removal process. Maker-controlled watershed segmentation was applied with tree tops as the makers to generate boundary of crowns.



a. **b.**

Fig 2.4 Original image and the watershed result after Gaussian filtering. X and Y axis represent the pixels in the image. Blue means low pixel value and red means high pixel value

(4) Watershed boundary improvement

For tree crown delineation, there are two issues should be included, the accuracy of the tree location, and the boundary. The former one is mostly related to the filtering condition and the latter one is more complicated. Tree crown delineation is focusing on searching for the boundary of individual trees. Even reference data for tree crown delineation needs proper definition for the boundary.

According to Fig 2.4, the original tree boundaries were created by flooding the basins when two basins were met. In this case, these boundaries may not be located exactly on the boundary of the trees. Since there are some gaps between the canopies, the original boundaries may be located within the gaps of the trees. Some studies directly detected the gaps using airborne LiDAR data (Armston et al., 2013). The boundaries derived using the original watershed algorithm from CHM images often do not accurately describe the crown boundaries. There were not many studies focusing on the boundary criteria for trees. It is common that the real boundary stands for the similarity for each pixel in one segment. The aim of the improvement for this thesis is to search for the real boundary of trees. There are some features of the pixels of the crown boundary. For example, the difference between the boundary pixel and its neighbors can be treated as the threshold for searching the boundary. Since only the boundary pixels have large difference with its neighbors other than the pixel on the ground and on the tree. The value of each pixel in the image stands for the height information, so a proper threshold should be found to distinguish the crown boundary.

Based on typical tree structure, crown boundary can be defined as the edge of all live branches. For tree crown delineation, three dimensional structure of a tree can be converted to two dimensions. From a vertical perspective, the shape of the crowns can be treated as circular or oval like objects. Trees are 3 dimensions objects. Crown shape is the silhouette of a tree, which is extended from branch tip to branch tip, including all the branches from top to bottom. However, few abnormally long branches will be excluded. The shape of the crown is varied from species to species and from time to time. However, the tree shapes from horizontal direction like pyramidal or elliptic are not crucial for crown delineation. The important issues turn to find the proper position to delineate the edge of the lowest branches for each tree based on height information. In this study, we introduced a variable called "Crown Ratio" to determine the crown boundary. Crown ratio is the percentage of total tree height that has live branches on it showed in Fig 2.5. Crown ratio can be calculated as the ratio of crown length to the tree height (Kuprevicius, Auty, Achim, & Caspersen, 2012). This concept is often used for forest studies to get the structure information from individual trees. Volume estimation, surface area estimations and other types of two dimensional crown issues are related to crown ratio. Since the value of each pixel means the height information of the area, crown ratio can be used to detect the boundary pixels of the crown. A characteristic of the boundary pixels is that the ratio between the boundary value and the local maximum value of the crown matches the crown ratio of this crown. In this case, a loop was designed to search the boundary pixel. When the ratio is found, the iteration process stops.

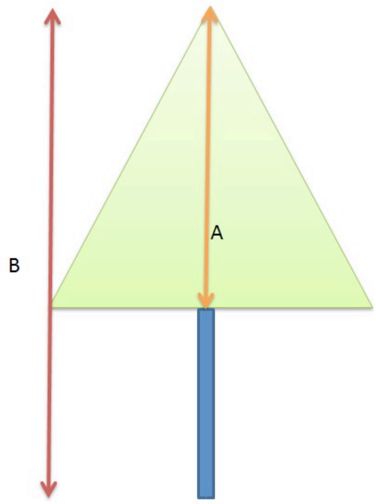
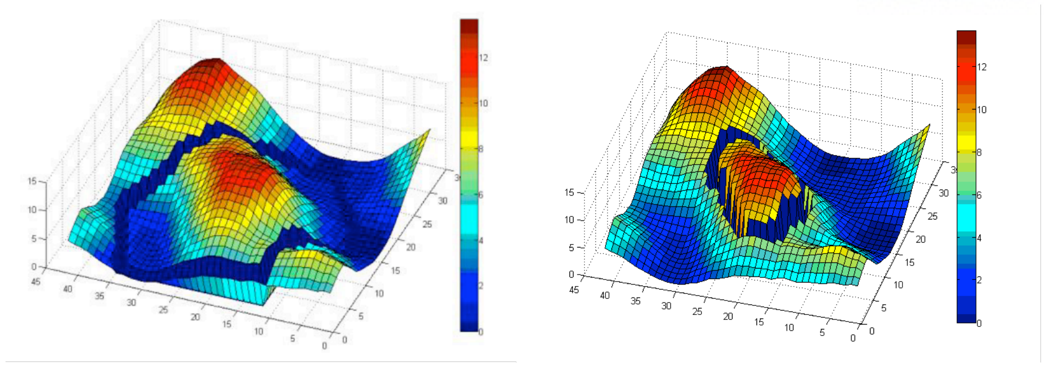


Fig 2.5 The definition of Crown ratio: A/B. A stands for the crown height and B is the tree height

Crown ratios are varied by species and age. In this case, the crown ratio used in this paper was derived from field data. The average crown ratio in these ten plots was 0.3 after several trials. The algorithm was designed below: Each segment has only one top which regard as the local maximum. The difference between the original pixel and the top was larger than 30 percent of the maximum height of the segment. For each segment after watershed segmentation, each pixel on the original boundary was trying to move towards the center of the segments. The difference is becoming smaller and smaller since the value of the pixel is kept increasing towards the center. The stop criteria for moving were the difference between the moving pixel and the top criteria was the pixel whose value is less than 30 percent of the maximum height of the segment. This criterion was simple and scientific for crown delineation since it considered the vertical structure of the tree. With this threshold, the boundary was getting closer to the center loop by loop to find the boundary. The boundaries before and after the process are shown in Fig 2.6. Image (b) fit much better compared with original watershed segmentation. The gaps between the trees especially in less dense areas can be distinguished and detected. Accuracy assessment was applied with reference data.



a. **b.**

Fig 2.6 Boundaries before and after the improved process. Dark blue line represents the boundary of crowns.

2.3.2 Algorithm comparison for segmentation: Region Growing

Region growing is part of the most commonly used algorithms for image segmentation. Region Growing is also performed as the pixel based image segmentation since it starts from given seed pixels. This method examines each neighbor of the seed pixels and detects the similarity of the seed and the neighbors to see if these neighbors should be included as the region. The sort of the measurement of similarity can be the pixel gray scale value, the texture or the color. However, it will cost some time for iterations. In this case, there are two main issues to discuss in the region growing method. One is seed selection and the other one is the similarity criteria.

First, seed selection determines the original location of the object which waiting segmentation and detection. The seed can be selected manually or automatically. Some of the studies used manually selected seeds for region growing. Manually determined seed location can be given previously or using code to get based on manually selection. However, for tree crown delineation, automatically detected seeds are commonly used. For tree crown delineation from LiDAR data, seeds for region growing are tree tops. Most studies searched local maxima using a proper scale and treat these local maximum pixels at the tree tops with the assumption that the tree tops are around the center of the trees. However, due to the complexity of forests like shrubs and high density, and the limitation of images, these local maxima may not totally match field derived reference tree tops. In addition, a scale for seed selection is also very important. A large scale will get less local maximum points so some of the small tree tops will be ignored. On the contrary, a small scale will get more local maximum points so such shrubs and small branches in the area will be misinterpreted as tree tops. In order to avoid some useless tree tops, image filtering is crucial as the preprocessing step for image segmentation.

Secondly, the quality of the region growing segmentation was also determined by the stop criteria. After searching local maxima, an automatic region growing method was applied. This method started from the local maximum. If the pixels with in this area met the threshold, this pixel can be included as the tree crown area. All the neighbors of one seed pixel should be examined. If the neighbor's value matches the criteria, this pixel will be regarded as part of the object. The algorithm will stop at the pixel whose value cannot match the given criteria or threshold. In this case, the stop criteria will largely determine the segmentation result. For individual tree crown delineation, the stop criteria are aiming at searching for the boundary of tree crowns. Some features should be defined in order to set the threshold to distinguish the boundary pixels from other pixels. An improved region growing algorithm for tree crown segmentation was presented (Solberg, Naesset, & Bollandsas, 2006). There are two restrictions for the segmentation. First, all the optimum pixels which can be included into the segment should include the steepest upslope neighbor. Secondly, in order to derive tree-like shape polygons, the optimum polygons where the line segment between the any pixel within the polygon and the local maximum will be totally contained within the polygon. These two stop criteria can help to determine the boundary of crowns and worked as part of a region growing algorithm.

In this algorithm, there were two things to consider. 1. The size of the certain area. 2. The stop method for boundary searching.

(1) Scale selection for region growing

Fig 2.7 shows the example of seeds selected from the filtered images. In the literature, the scale selection for region growing was determined through several trials or by the relationship between the tree height and crown size. For this study, the scale of window size for region growing was determined by this equation: H is the value of the local maximum (Popescu & Wynne 2004):

$$Crown\ Width = 3.75105 - 0.17919\text{H} + 0.0124H^2$$

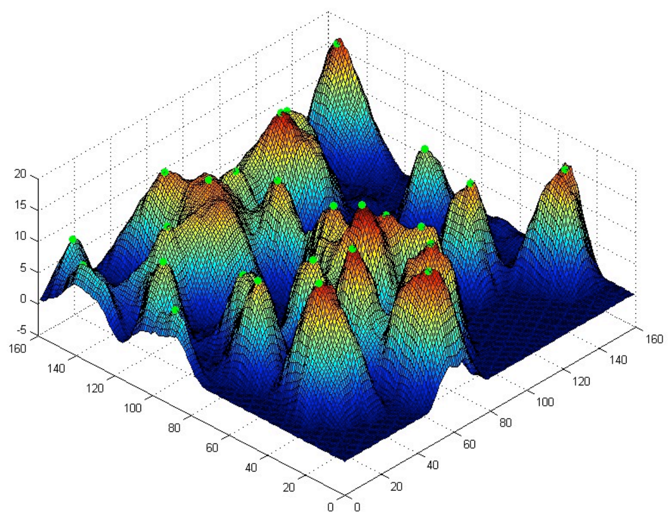


Fig 2.7 Seed selection for seed growing method.

In this case, variable window sizes were applied in this research. Variable window size is very crucial for tree crown delineation since the variety of the value of the local maximum. Using single size of region growing may ignore some small crowns or merge some crowns (Hirschmugl et al., 2007).

(2) Automatically search for tree boundary

Within each determined area, an automatically determined threshold for boundary searching was introduced. The assumption for region growing algorithm focuses on the height information of each pixel. The requirement to include a pixel into the crown was defined as: the difference between one pixel and the seed value was smaller than that of the seed and the mean height of this certain area. Areas regarded as the crowns were selected and boundaries were created. Accuracy assessment for individual tree crown delineation was applied after the region growing segmentation.

2.4 Results

Two algorithms described above were applied for 10 plots. The results of improved watershed segmentation and customized region growing are showed in Table 1 and Table 2 respectively. Based on rule-based iterations using crown ratio, new boundaries were generated. As described in Section 3, the window sizes for Gaussian filtering were varied from plot to plot. There were three levels of window size, large size, middle size and small size. All the window size information was shown in Table1 and 2. For sparse plot, trees were located more separately, so large window size was applied for filtering. The opposite happens to dense plots.

To effectively evaluate the segmentation result, reference crowns are derived by independent researchers. Manually segmented crowns were based on a 0.25m spatial resolution images after Gaussian filtering and the field detected tree locations with the help of aerial photos. These reference crowns were further verified using 3D visualization in ArcScene 10. The crown location of in-situ data was also taken into consideration. There are totally 198 trees in the ten plots. The delineated crowns were interpreted as target crowns. Target crowns were counted within the plot circle with the diameter of 22.8m.

Fig 2.8 shows the results from improved watershed segmentation visually. Fig 2.9 shows the results of the comparison group of the region growing. Outlines of target crowns are superimposed on the CHM of each plot. The red lines are the target crowns and black and white lines show the reference crown derived manually. From a visual perspective, some reference crowns can be fully delineated especially plot 3. The boundaries from reference data and the target crowns almost overlap. However, other target crowns were larger than the reference crown, and there was few target crowns was smaller than the references. Reference crowns will be merged or split due to large and small target crowns. In order to evaluate the accuracy of the segmentation quantitatively, the criteria were defined below.

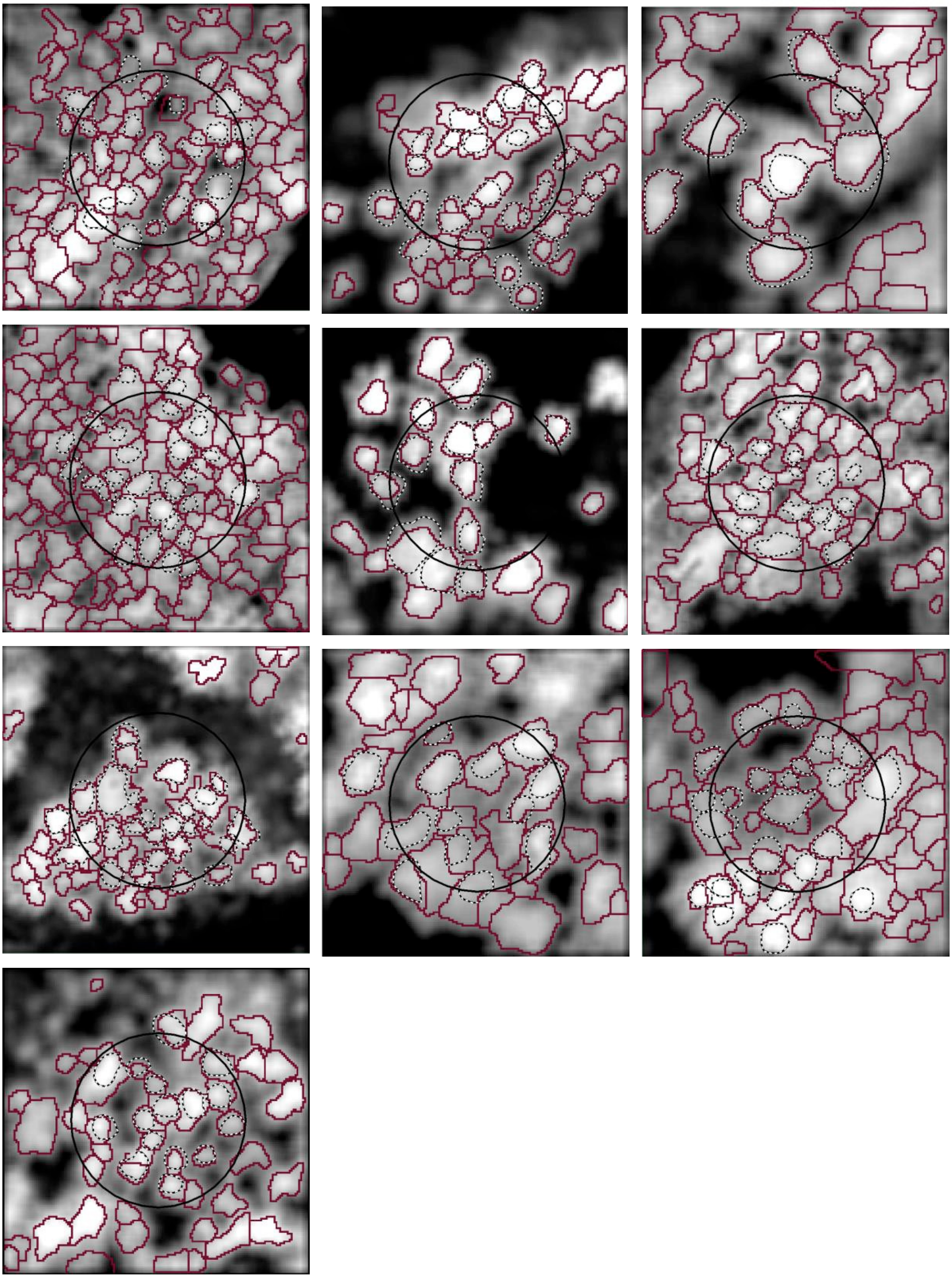


Fig 2.8 Improved watershed segmentation (Red: Delineated crowns; Black and white: Reference).

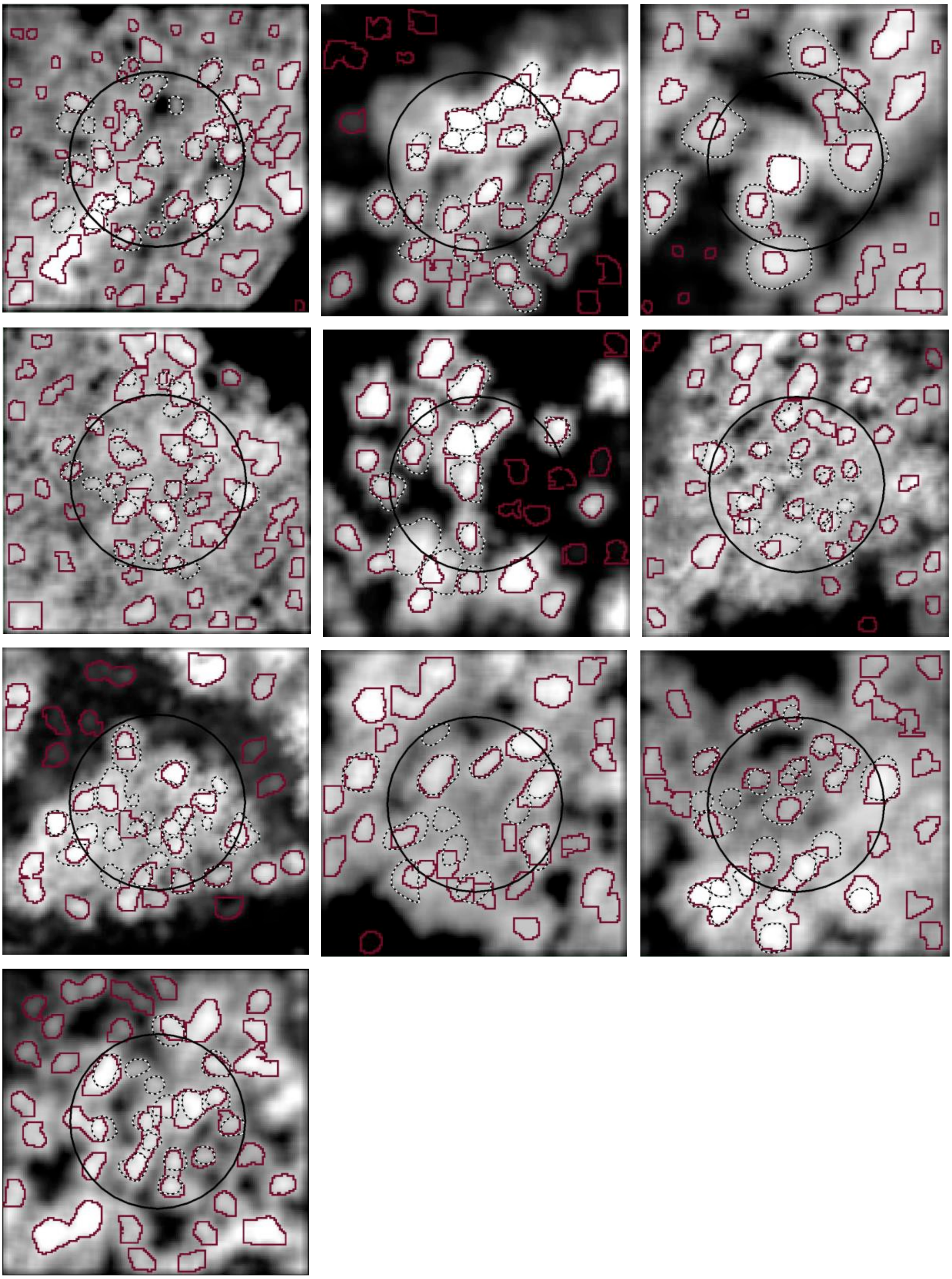


Fig 2.9 Improved region growing segmentation (Red: Delineated crowns; Black and white: Reference).

(1) Matched - For a reference crown in the plot, if more than 50% of the reference crown overlapped with only one target crown and the spatial center of the reference crown also within this target crown, the reference crown was regarded as matched crown.
(2) Merged - If more than one reference crown are included in one target crown and their spatial center are also covered by this target crown. These reference crowns are all viewed as merged crowns.
(3) Split - If more than 50% of one reference crowns are occupied by more than one target crown, this reference crown is considered as split crown.
(4) Missed – If less than half of a reference crown was overlapped with other target crowns and the spatial center doesn't belong to any target crowns, this reference crown is treated as a missed one.

Table 2.1 Accuracy assessment for improved watershed algorithm for 10 plots.

Sparse plot	**Plot**	**Reference**	**Window size**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	3	8	10	8	0	0	0	100%
	5	12	10	10	2	0	0	83.30%
	8	13	9	11	2	0	0	84.60%
Less Sparse plot	**Plot**	**Reference**	**Window size**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	6	16	7	14	2	0	0	87.50%
	9	24	7	20	3	0	1	83.30%
	10	16	7	15	0	1	0	93.75%
Dense plot	**Plot**	**Reference**	**Window size**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	1	26	5	23	0	0	3	88.50%
	2	27	5	26	0	1	0	96.30%
	4	29	5	23	6	0	0	79.30%
	7	27	3	20	4	1	2	74%

Table 2.2 Accuracy assessment for region growing algorithm for 10 plots.

Sparse plot	**Plot**	**Reference**	**Target**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	3	8	9	8	0	0	0	100%
	5	12	13	8	4	0	0	67.00%
	8	13	13	10	0	0	4	71.00%
Less Sparse plot	**Plot**	**Reference**	**Target**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	6	16	16	9	4	0	3	56.00%
	9	24	21	13	10	0	3	50.00%
	10	16	17	10	4	0	1	67.00%
Dense plot	**Plot**	**Reference**	**Target**	**Matched**	**Merged**	**Split**	**Missed**	**Accuracy**
	1	26	25	11	6	1	8	71.00%
	2	27	14	9	13	0	7	31.00%
	4	29	23	18	4	0	8	60.00%
	7	27	16	7	12	0	4	29%

According to the criteria above, Table 2.1 shows the summary of the four evaluating accuracy for watershed segmentation. In this study only matched reference crown can be treated as effectively delineated crowns. In this case, the accuracy is calculated as the percentile between matched crowns and references. Almost all the crowns can be detected and the average accuracy is more than 85%. However, as the plot getting denser, the accuracy is slightly decreased. The numbers of merged, split and missed crowns are increasing.

The table 2.1 suggests that with less density plots, the delineated method works well for extracting tree crowns compared with density plots. Merged crown happened when local maximum are excessively merged before watershed segmentation because these trees are too close together. Split crown is due to the misinterpretation of the branches of the tree. Missed crowns happened when no local maximum is found in that region because of the large window size is applied to smooth the image.

2.5 Discussion

The framework proposed in this study includes these two novel steps: (a) detect the tree top using local maximum after Gaussian filtering and delete the useless tree tops. (b) Improve the original boundary of watershed segmentation. This improved watershed algorithm is also efficient since only one parameter is needed during tree top detection and no parameters are needed during boundary improvement process. In addition, the result is promising both for the visual observation and accuracy assessment table compared with region growing algorithm.

In terms of background removal and tree top detection, the mainly novel approach is based on morphological operations and automatic Matlab algorithm. As done by Wang et al.,(2004), a Laplacian of Gaussian edge detection method at the smallest effective scale was used to mask out the background. This method was used for aerial photo. Background and useless small shrubs and branches were removed in this study by Otsu's method in Matlab. For scale selection, morphological operations are closely related to multi scale analysis and integration during the delineation process (Jing et al., 2012). 69 percent, 65 percent, 73 percent of the tree crowns can be correctly detected in coniferous forest, deciduous and mixed forests respectively. Multi-scales are selected and the integration for removing fake crowns was applied after segmentation because of the mixture species forest. However, in this study, the removal of fake tree tops was first applied before the segmentation. In this case, fake tree top were excluded using opening operation and merging process instead of multi-scales. Initially, the filtering window size was chosen according to the amount of trees. As described above in the methodology, opening operation and tree top merging process can greatly remove the small and useless tree crowns. However, size of opening operation and the threshold for merging process has to be set manually. In the future, more advanced rules with automatically determined parameters will be expected.

This study had demonstrated the novel shrink process for boundary refinement which was efficient and effective. Although it took around 500 seconds per plot for crown delineation, the boundaries are promising and scientific. The stop criteria for shrink are based on the crown ratio between the target pixel and the original pixel. Other researches (Solberg et al., 2006) also tried to control the shape of the crown segments through region growing algorithm and set changing rate of the pixels from each direction as two restricts for the boundary, or using height coefficient of variation to control the boundary of the crowns (Colgan et al,. 2012). Based on the local maximum of geodesic distance (Wang et al., 2004), makers were generated to find crown boundaries. This process needs to calculate all the geodesic distance for each maker. For tree crown delineation of LiDAR data, Solberg et al (2006) developed an algorithm controlling the shape of the crowns. The algorithm was improved based on region growing and the shape was controlled through star shaped polygons. The assumption here was that the segments should have the tree like shapes. A line based restriction was set to judge the pixel is included in the segment or not. The line between the boundary pixel and the local maximum should be entirely contained within the new polygon. This algorithm was reasonable to generate tree-shape segments; however, the algorithm only starts from the flat perspective of the image and set the restrictions. These criteria are of high complexity and less convenient compared with crown ratio. Crown ratio takes the structure of the tree in to consideration, which converts the 3D structure information into the flat image, and finds the scientific explanation of tree structure for crown boundaries. The introduction of the concept of crown ratio successfully combined image segmentation and forest features together effectively. The exact definition of crown area should be determined by using crown ratio. The shrink process is loop by loop, instead of controlling the times of loop, the circularity of the segments, crown ratio is used for the stop criteria. Segments will reflect the crown boundary in reality.

Compared with the second algorithm, improved watershed algorithm was better. First, there are too many merged segments in the region growing algorithm. In this case, the re-split process may be needed and will make the algorithm more complicated. Furthermore, if we want to avoid the merged segments, we have to decrease the size of region growing area. This will make the algorithm less objective since our expecting results affect the segmentation process. The improved algorithm started from the original boundary of watershed, which already divided the segments, no further re-split process was needed. To improve the region growing method, the size for growing was the major point to concentrate. Popescu & Wynne (2011) introduced some relationship between the crown size and the tree height, which can help to determine the window size. However, because of the complexity of the forest and variance size for different species, it is not that convenient to find a proper relationship between the size and the tree height. The relationship between the tree height and crown size should be different from species to species. So the improved watershed algorithm avoided these problems and delineated the crowns from the boundary to the center. Watershed segmentation is proper for tree crown delineation, however, the improvement of boundaries is also important is described in the methodology section.

This novel algorithm also provides methods for further studies related to image processing and object recognition. This study both includes digital image processing problem and forest related problems. Beside tree crown related topics, some other kinds of object detection like building detection, also widely study these years. Compare with building detection, individual tree detection is more difficult since the complexity of the forests. In this case, more detailed tree based parameters and variables should be taken into consideration and the process is more complicated.

There are some limitations in this study. First, the window size for Gaussian filtering needed further study. An automatic size determined algorithm is expected. However, since the complexity and the various species in the forest, determination of the window size for filtering is not easy. In addition, especially for the mountainous area, the trees lying on the slope will be somewhat inclined to one side and overlapped with other crowns. In this case, two tree tops will stay too close, which also make it more difficult to detect the tree tops and more crowns will be merged. Secondly, the threshold for watershed boundary shrink process was based on crown ratio. However, crown ratio is also individual tree based value and different from species to species and from time to time. Accuracy of boundary detection will be more precise if crown ratio is set species by species and different from individual to individual. Future studies may focus on detecting the species first and applied crown ratio threshold to generate more accurate results for individual tree crown delineation.

In conclusion, the algorithm provided in this study was efficient for forest in mountainous area especially for coniferous wood with various tree species. In terms of accuracy assessment, the algorithm performance was quite promising when compared with manually delineated crowns. More researches are required to complete testing this algorithm in deciduous forest with a deciduous forest in order to see if the tree detection and boundary detection are accurate enough.

2.6 Conclusion

The novel approach successfully removes gaps among trees and the approach yielded reasonable results even for low density forest. Novel algorithms of fake crown removal and boundary improvement work well for these 10 plots. Boundaries of improvement were firstly defined properly. The results are pretty promising and more sites can be tested including the deciduous forest area. However, future study should focus on the mixed forest and test crown ratio based on species to species. An automatic crown ratio detection algorithm is also needed. Ancillary data such as high resolution aerial orthophotos can be used as reference data to facilitate accuracy assessment. Additional tests of this approach in various forests with and multiple stories are essential to make the algorithm more robust and flexible.

3. BIOMASS ESTIMATION USING LiDAR-DERIVED METRICS AT THE INDIVIDUAL TREE LEVEL IN A RUGGED MOUNTAINOUS AREA

Abstract

Forest biomass quantification processes from LiDAR data have been applied and studied at a rapid speed these decades. This study aims to estimate the biomass use three kind of machine learning algorithm: random forest, support vector machine and Cubist based on individual tree crown delineation. The most important variables are found in the three models based on the individual tree level procedure. Results indicated that Cubist created the most accurate biomass with the R^2 of 0.8 using 5 variables: tree height, crown area, slope, crown volume and stem height. All the models suggested that slope is the important factor for estimate AGB and crown height, stem height also influence the estimation process. The effect of crown delineation accuracy on biomass estimation was also discussed. The accuracy of individual tree crown delineation will directly influenced the performance of AGB estimation using machine learning approaches.

3.1 Introduction

Biomass is biological material which derived from living or recently living organisms. In forestry area, biomass is usually defined as the above ground portion of trees. Biomass has drawn pretty much attention recently since biomass is highly related to carbon storage and the further understanding of the carbon cycle (Bortolot & Wynne, 2005). Above ground biomass can be derived from field data, however, more and more researchers are using remote sensing data for biomass estimation. Compared with field data, remote sensing data can reach every section of the ground. The accessibility of remote sensing data also makes it accessible to estimate biomass.

Forest biomass can be interpreted as the amount of living or foliage of root, stem, leaf, seed and flower. The unit for forest biomass is usually defined as the amount of dry weight for all the materials per unit area or unit time. However, biomass is closely related to the living condition of the forest, in this case, numerous factors contribute to the amount of biomass since they can influence the living conditions of forest and trees. These factors are: climate, including the annual precipitation, temperature, sunshine and so on; ground condition, include the slope of the ground (Sun, Ranson, & Kharuk, 2002; Darke et al., 2003; V. de Castilho et al., 2006), and the soil condition; the ecosystem of the area, including the competition between the species, and the interaction among the plants and the animals. In this case, there are bunch of researches focusing on the relationship of biomass estimation based on these factors.

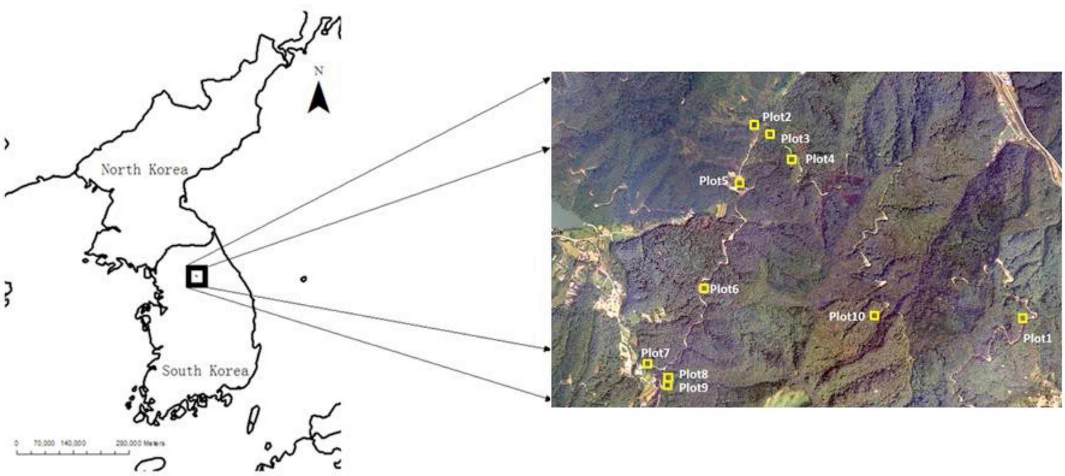
Forest ecosystem has almost 76% to 98% of the carbon element of the land system. Forest ecosystem plays the important role in the global carbon cycle, and carbon storage and exchange between the forest and the atmosphere through photosynthesis, respiration and combustion. In this case, carbon storage is drawn more and more attention since it closely related to the carbon dioxide in the atmosphere, which directly influences our daily life. Carbon storage is the crucial variable for carbon cycle all around the world. Particularly, aboveground biomass (AGB) is estimated at a landscape scale presents an attractive and effective tool to understand how forest influence the carbon cycle and the forest productivity (Lefsky, Turner, Guzy, & Cohen, 2004). Therefore, biomass estimation becomes the essential and necessary step for carbon storage estimation. Several methods for biomass estimation has been built and applied in the past researches (Bortolot & Wynne, 2005; Lucas et al., 2006; Boudreau et al., 2008; Kim et al., 2009; Zhao, Popescu, & Nelson, 2009; L.Powell et al., 2010; Becknell, Kucek, & S. Powers, 2012; Zhao, Guo, & Kelly, 2012; Ahmed, Siqueira, & Hensley, 2012).

Various variables are supposed to estimate biomass such as diameter at breast height (DBH), stem bark, or foliage. DBH performs as the primary variable to estimate biomass (Popescu, 2007). The study is aiming at assessing the accuracy of estimating DBH for individual tree using LiDAR derived data and searching for the relationship between DBH and the biomass. This research works as the leading example for estimating biomass based on individual trees. Equation is constructed with only one input: DBH. Another equation was also used which take individual tree height into consideration for DBH estimation. They got the conclusion that DBH works as the most reliable variable for biomass estimation. Another method to estimate biomass is using percentile tree metrics with both LiDAR data and Full-Waveform signals (Allouis, Duurrieu, Véga, & Couteron, 2012). Reference data for biomass are calculated based on equations and the processed data were based on smoothed CHM. Similar to the study which introduced before, Allouis et al. (2012) also derived the biomass from individual trees. Local Maximums were detected first for tree crown segmentation. Besides tree total height was derived from each segment, tree bounding volume (TBVCHM), which is similar to canopy geometric volume, was also extracted for canopy metrics. TBVCHM is the model with the shape of an elliptic cylinder of height as the tree height and the area which is the area of the crown. 25th, 50th, 75th and 90th height percentiles were retrieved from the data. The conclusion showed that TBVCHM is one of the most useful metric for volume and biomass estimation. Furthermore, biomass estimation also takes slope effect into consideration. Since the biomass estimation is largely influenced by the height information from LiDAR data, tree height is of highly importance for biomass estimation, which is also highly affected by the slope of the area. As for the topographical features, some studies keep on understanding the influence of the slope and terrain condition for biomass. Not only the LiDAR derived data will be influenced by slope factor, but also the biomass will be higher on the steep slope area since active photosynthetic area of trees is more than that of flat area. Difficulty accessing biomass and other parameters on steep slopes make it is more and more necessary to derive the proper method for calculating biomass through LiDAR data (Barbosa, Melendez-Pastor, Navarro-Pedreño, & Bitencourt 2014). A survey on the mountainous region is laborious and time consuming, and even dangerous. For example, some researchers already made some progress on biomass in mountainous region (Sun, Ranson, & Kharuk, 2002; He et al., 2012). The most difficult issue here is trying to minimize image distortion especially for mountainous areas. Biomass was estimated over precipitous slopes and they explore how topographical features affect biomass using remote sensed data (Barbosa et al., 2014). A straightforward framework and the novel method for forest biomass in steep slope areas were developed by Barbosa et al. (2014). In this study, two kinds of data were utilized, Landsat and topographical features from DEM. Illumination factor (IF) was calculated as the DEM and solar angles at the same time of the Landsat data. IF was described as one of the variable for modeling biomass and calculated in three different ways. Slopes and aspects were combined differently into three illumination factors: IF_{cos}, $IF_{hillshading}$, and $IF_{vector\ sum}$. The results showed that biomass can be best predicted when combined with a satellite image of Landsat TM5 with $IF_{hillshading}$ especially for mountainous area. Besides all the success of previous studies, there is something remaining to be improved.

First, there are few studies related to the crown ratio and stem height of the biomass estimation, especially from LiDAR data using machine learning approaches. As mentioned above, crown ratio plays an important role in estimating the quality of an individual tree, which represents the portion between the leaves and the stem. Since the biomass reference was derived based on the stem and the leaves, we assume that the variable selection should also base on the different part of the tree. In this case, crown ratio and stem height are portion based variables, which can serve to find out their relationship with the individual tree biomass. Secondly, there are some differences of the LiDAR image where the tree is situated on the steep slope and the one is not. The slope and terrain condition not only affect the height information derived from LiDAR, but also affects the productivity of the individual tree. In order to test the affect from slope and terrain conditions, different combination variables should be selected and compared to see the importance. Thirdly, there are already several researches related to estimate biomass using machine learning approaches with all the variables. There is still no standard selection for variables. Optimum variables should be found for biomass estimation. In addition, biomass estimation should be more accurate if combined with individual tree crown detection. In this case, our study focused on the slope affection and biomass estimation was applied after individual tree crown delineation.

3.2 Study area

The study area, Maehwa Mountain is situated in Gangwondo, South Korea. The location is showed in Fig 3.1. The terrain condition for this area is mountainous. The average angle of slope is around 30. Ten plots were selected with the radius of 20 meters and field measurements were implemented to derive height, DBH, species and the distance from the center of the plot for each tree. The species in the study plot are *Pinus densiflora*, and *Larix kaempferi*. These are most abundant species located in South Korea.



a. **b.**

Fig 3.1 (a) The study area and (b) Sample plots in the study area. Each plot is with the diameter of 22.8m.

3.2.1 LiDAR data

This study used discrete multiple-return LiDAR from an airborne CESSNA 208 CARAVAN sensor with the altitude of 2300m. Raw data were digital elevation model (DEM) and digital surface model (DSM). Both were unfiltered and discrete images of the whole mountainous area. Image cropping was applied as the first step to extract the ten plots and image smoothing was applied in Matlab R2012a.

3.2.2 Reference data

For the ten plots, the conditions of the trees are various. For example, trees in plot 3 are much taller than the others. However, plot 3 has the least amount of the tree. On the other hand, plot 1,2,4,7 have nearly 30 trees inside with high density. The terrain condition is also varied from plot to plot.

Reference data for individual tree crown delineation and biomass estimation were both rely on the ground data collection of tree species, DBH and tree height. All the field data was collected on August, 2013. For biomass estimation, we select 154 trees in this area.

Various studies derived biomass reference data from based on the field measurements and equations (Gonzalez et al., 2010; Popescu, 2007; Gleason & Im, 2012; Tian et al., 2012). Most of the equations, from USDA Forest Service, were estitablished for all deciduous and coniferous species. Our biomass reference data relied on the equations provided by the Korea Forest Research Institute and the Korea Forest Service. These equations are different from species to species and different coefficients are offered for different portion of the tree. The basic equation is below:

$$B=aDBH^{b}H^{c}$$

In this equation, *a,b,c* are three different coefficients which is various from species to species. These are also different according to different portion of the tree. *B* stands for biomass and *H* means to measure tree height.

The total above ground biomass for individual tree was defined as the sum of the biomass from stem, leaf and branches. Each portion has its own coefficient. For each formulation, two parameters were used: tree height, and DBH from field data. The superiority of this method for reference is that the biomass amount for trees is various from species to species and from portion to portion. In this case, the reference data were portion based and species based, which is much more in detail and accurate.

Our individual tree level biomass estimation is based on the result of individual tree crown delineation. Reference crowns were derived manually from two experienced researchers with LiDAR data using canopy height model (CHM) data. In addition, reference crowns were manually delineated after image filtering and all the process was developed in ArcMap 10. During the manually delineation process, there are a few principles. The crown dimension should be scientific and realistic. The number of reference crown should try the best to match the number of fields measured trees. For the crowns which are too difficult to be distinguished by CHM in ArcMap, ArcScene 3D image using CHM can help more to determine the boundary. Since the study area was rugged and some places are of high slope, some field record tree location showed the image is hard to be interpreted by human eyes. In this case, the reference crowns will follow the CHM height information first to detect tree shape objects. Some trees are overlapped and hard to separate. Reference crowns are trying to divide the crowns properly according to the possibility of the size of the crown.

3.3 Methodology

3.3.1 Overview

Biomass is calculated based on individual tree level in this study using three different methods: Random forest regression, support vector regression (SVR) and Cubist regression. The effectiveness of these three different machine learning method was tested. In each method, several kinds of unique combination of variables were implemented. Various combination aims at finding out the most relevant factors which contribute to biomass estimation. Results show in the comparisons between the modeled biomass and reference. All the variables were extracted based on individual tree crown delineation using improved watershed segmentation. Only one to one matched segments were chosen for biomass estimation based on the delineated crowns and the tree location measured from in situ data. One to one match process was applied using ArcMap 10.

3.3.2 Tree crown delineation

Individual tree crown were delineated using improved watershed algorithm and shows nearly 80 percent of the reference trees can be satisfactorily detected. This improved algorithm is the synthesis of tree top location detection and boundary delineation. Firstly, Gaussian filtering was applied and background and small crowns with shrubs were removed based on image opening operation and Otsu's method utilizing canopy height model (CHM) images. Secondly, traditional watershed approach was applied and the original boundary was created. Since the gaps between the trees especially in less density areas cannot be detected using the original watershed approaches. The next step and the improvements tried to search for the real boundary for tree crowns. The definition and criteria for tree boundaries were based on the concept of crown ratio. Crown ratio stands for the ratio between the height of the crown and the height of the stem. According to the features of LiDAR images, the crown boundary can be detected using crown ratio after determining the tree height of each segmented original trees. The boundary searching process was starting from the original pixels from the watershed. If the original pixel's value cannot match the crown ratio value when compared with the tree height, move this pixel towards the center of the segments until crown ratio value iss matched. This process can largely improve the shape and size of the segments since the shape and the area of the crown are very crucial for biomass estimation. Large area of crown means bigger biomass amount than small area of the crown. Our biomass estimation can be more accurate based on the improved crown boundaries since more accurate and scientific boundaries were made. All the boundaries were extracted and variables were selected on individual tree level for biomass estimation.

3.3.3 Above ground biomass estimation (AGB): variable selection

Individual tree crown delineation was implemented and acted as the basic step for biomass estimation. After individual tree crown delineation for each tree, several variables were extracted to predict the amount of biomass. The variable selection is of eminent importance for accurate biomass estimation. However, there is not a common standard for choosing the variables. In this study, we calculated the common variable as described in other papers and developed and calculated our own special variables to test their importance for biomass. The candidate variables are height (minimum height, 60th, 70th, 80th, 90th, height percentile per crown), crown area, mean slope of the crown, crown length and the height of stem, crown volume (volume of 60th, 70th, 80th, 90th, height percentile per crown), standard deviation with in the crown, and perimeter of the delineated crown.

In other studies, various variables were extracted (Gleason & Im, 2012; Breidenbach et al., 2010; Vauhkonen, Korpela, Maltamo, & Tokola, 2010; Yao, et al., 2009; Næsset et al., 2011), however there are few studies focusing on the best combination of variables for AGB estimation. In our study, several traditional and commonly used predictor variables were adopted as the previous studies (Li, Im, Quackenbush, & Liu, 2014; Kronseder, Ballhorn, Böhm, & Siegert, 2012), while some special predictor variables were also derived for the estimation. All the variables were extracted based on the delineated crowns for each individual tree. Crown height, minimum height, crown area, perimeter and crown volume are the conventional predictor variables which directly related to the volume of the tree. These four variables are used as the basic predictor variables (Li et al., 2014; Gleason & Im, 2012). Tree height is extracted as the maximum height within the delineated crown. Volume was calculated on the basis of each volume of the pixel and the sum of all the pixels can be treated as the volume of the tree. Crown height and stem height were derived based on the boundary of individual tree crown delineation. Since our tree crown delineation process used crown ratio as the threshold, the crown height and stem height are more accurate and they are a portion based variables. The mean value of all the boundary pixels stands for the stem height and the difference between the local maximum of the segments and the stem height can be regarded as the crown height. Since our crown delineation method improved the boundary of watershed segmentation using crown ratio as the threshold, the boundary pixels are the distinguishable pixels between crown and stem. Our AGB reference data were calculated on the basis of the different portion of the tree like stem and leaf. So the assumption is that the predictor variables should also base on a different portion of the tree in order to get better predicting result. Distict combinations of these variables were tested to find out the significance of stem height and crown height. Stand deviation was calculated within each crown to measure how the height of each pixel spread out from the mean height of the crown. Importance can be found if stand deviation is important for biomass of an individual tree. In addition, our study area is a mountainous area. Regarding to the complexity of the terrain and slope condition, the amount of AGB may be different when compared with flat forest (Sun et al., 2002). The assumption is that the trees lying on the slope area may have less competition with other trees and larger amount of light will be arrived. In this case, AGB amount will be larger on the slope area and the relationship between slope value and biomass should be detected. Average slope value was calculated within each tree crown area. Different combination of variables using slope as the variables are adopted and results were shown below in Table 3.1. Specifically, slope information was extracted using DSM images, and the other variables were extracted using CHM images. In conclusion, our variables selection was based on the different portion of the tree and the terrain condition. The combinations of these variables are intended to find out the significance of each variable and chose the best options for biomass estimation. 17 variables were selected and listed below.

Table 3.1 Summary of the predictor variables

Name	Description
Area	Area for each crown
MX H	Maximum height per crown
H90	90th percentile of height per crown
H80	80th percentile of height per crown
H70	70th percentile of height per crown
H60	60th percentile of height per crown
Height to crown base	bole height per crown
CL	Crown length for per tree
Slope	Mean slope of within the crown
CGV	Crown geometric volume
CGV90	CGV for 90th percentile of height per crown
CGV80	CGV for 80th percentile of height per crown
CGV70	CGV for 70th percentile of height per crown
CGV60	CGV for 60th percentile of height per crown
STD	Standard Deviation of height per crown
MN H	Minimum tree height
Perimeter	Perimeter per crown

3.3.4 Random forest (RF)

Random forest is the commonly used algorithms in machine learning, which is aiming at classification, regression and clustering using many classification or regression trees (Walton, 2008). The algorithm is first introduced by Breiman (2011), trying to train various regression trees, creating models and then generates the classification result. During the random forest process, about one third of the training set trees will be left out of the sample called oob. Oob (out of bag) data is very crucial since it's used to get a running unbiased estimation of the classification error and estimate the variable importance. As each tree is constructed, all the proximities are computed for each tree. Proximity increased by one when the same terminal node was met. The average will be calculated at the end of the run based on the proximity. Features for RF are obvious. Random forest does not overfit. Results will be slightly different each time since the training process is random selection (Walton, 2008). In addition, random selection requires large data bases to generate more efficient and accurate result. An internal unbiased estimation and balancing error can be obtained in class population unbalanced data sets. Variable importance can help to check and determine which variable contribute most during the process and analyze the relationship between the variables and the input.

With the random inputs variables, various researches used RF as the primary method for modeling quantities (Falkowski, Evans, Martinuzzi, Gessler, & 2009; Walton, 2008; Hudak, Crookston, Evans, Hall, & Falkowski 2008; Knudby, LeDrew, & Brenning, 2010; Li, Im, & Beier). Among them RF also widely used to estimate biomass (Gleason & Im, 2012; Li, et al., 2014). In this presented study, the RF model was implemented in the R statistical software, utilizing the random forest package. The predictor variables were prepared and different combinations were set up to test the variable importance for each variable. Error and proximity on out of bag data were also calculated.

3.3.5 Support Vector Regression (SVR)

Support Vector Regression (SVR) is belonging to the machine learning field which extended from Support Vector Machine (SVM) and aiming at regression analysis. Remote sensing implementations of SVM were reviewed in detail (Mountrakis, Im, & Ogole, 2011). Support Vector Machine has served to remote sensing related topics like classification and change detection (Gamps-Valls., 2004; Huang et al., 2008; Kavzoglu & Colkesen, 2009; Zhao, Popescu, Meng, Pang, & Agca, 2011). More accurate and less biased robust estimation results for biophysical parameters were proposed using SVR even compared with neural networks and classical bio-optical models (Gamps-Valls, Bruzzone, L. Rojo-Á lvarez, & Calpe-Melgani, 2006). SVM was implemented at land cover change based on the time series showed that SVM costs longer processing time (Chen et al., 2014). For forest studies, a highly automatic algorithm was designed for mapping forest cover change with SVM (Huang et al., 2008). This algorithm is particularly useful for forest change over large areas. Canopy structure was studies using SVM (Zhao, Popescu, Meng, Pang, & Agca, 2011). SVR also produced the highest accuracy among other models like RF and Cubist for model validation (Li et al., 2014). These studies provided reliable efficiency and effectiveness of SVR to generate regression for forest studies.

Conceptually, SVR adopted kernel functions to separate groups of input data and trying to adopt find out the ideal hyperplanes to separate the input data and then predict the response variable. The hyperplane is adapted according to the error. In this study, variables were first determined and the combinations were the same as we used in RF to test the effectiveness of SVR to predict biomass. However, SVR doesn't provide the variable importance in the process. The best combination and variables for biomass will be tested through several trials. All the variables were tested first and then remove the interested variables one by one to check if they have an important influence during the process.

3.3.6 Cubist

Cubist is a rule based regression tool to generate predictive models from data. It is the commercial product, so the detailed process and rules are unknown. Modified tree regression system for instance based criteria was used to predict models with balancing the need for more accurate prediction. However, cubist has been implemented several times in the remote sensing area. As tested in land cover estimation (Walton, 2008), cubist is the fastest algorithm among SVR and RF and it used all bands to predict surface cover. For forest related studies, it was found that Cubist regression trees generate the best calibration accuracy for estimating LAI with the R^2 larger than 0.8 (Im et al., 2012) using hyper spectral images. More recently, Cubist also used to estimation forest biomass carbon stock (Li et al., 2014) and an almost similar result with that of SVR. More studies should focus on Cubist to test its efficiency and effectiveness. Our study made use of the same variable as mentioned above to compare the three models.

3.4 Results

The accurate table for tree crown delineation showed below in Table 3.2. The overall accuracy is pretty high around 80% and even more. Based on these delineated crowns, we extracted the variables showed in the Table 1 before. In order to assess the accuracy of the segmentation quantitatively, the criteria are defined below. Target crown was referred to describe the delineated crowns.

(1) Matched - For a reference crown in the plot, if more than 50% of the reference crown overlapped with only one target crown and the spatial center of the reference crown also within this target crown, the reference crown was regarded as matched crown.

(2) Merged - If more than one reference crown are included in one target crown and their spatial center are also covered by this target crown. These reference crowns were all viewed as merged crowns.

(3) Split - If more than 50% of one reference crowns was occupied by more than one target crown, this reference crown is considered as split crown.

(4) Missed – If less than half of a reference crown is overlapped with other target crowns and the spatial center didn't belong to any target crowns, this reference crown is treated as a missed one.

There are 198 trees were delineated in the ten plots. However, for our biomass estimation, only the one-to –one matched crowns will be estimated. Target crowns were matched one by one as accurate as possible with the reference crowns. After the process, there are 151 trees to be modeled for biomass estimation.

Three different kinds of machine learning approaches were used in this study: Random Forest, support vector regression and Cubist. In order to discover the best variables for biomass estimation, three models were used separately with all the 17 variables at first and check the importance of each variable. The regression figures are also showed as follows and three values are calculated to test the accuracy of each model: R^2, which represents how match the input data are. The higher the R^2, the modeled data is more suitable with the reference data; RMSE, which means the root of mean square error between the modeled value and the reference. The higher the RMSE, the weaker relationship between the observed data and the modeled data; CV RMSE, which represent the cross validation RMSE, is the necessary for data validation. This value is a technique for estimating the performance of the model. Cross validation means some out of bag data will be used to validate the accuracy for the other selected data, which used as the predictor variables for predicting. RF has the unbiased out of bag process and will generate the CV RMSE automatically. Cubist can also get the value although the process is unknown. The higher the CV RMSE, the poorer performance of this combination of variables to predict the biomass.

Fig 3.2 shows the variable importance of RF in three sets of variables. Firstly, Fig 3.2(a) shows the importance of all the 17 variables. The importance values for the 17 variables are various and in order to discover the best combination, several variables with low importance are removed for the second trail. As the result showed in the figure, crown area, STD and crown height contributed comparably the least than the other variables, so we delete these variables and run RF again. The R^2 increased when using the other 14 variables (showed in Fig 3.3(b)). For the regression validation, both RMSE and CV RMSE decreased obviously. The best R^2 was derived when only 12 variables were chosen as showed in Fig 3.3 (c) since the other 2 variables, minimum height of the crown and the perimeter showed less importance than the others. The R^2 increased from 0.55 to 0.6 compared with the one using 17 variables. When attempting to decrease the amount of variables less than 12, the accuracy of RF stopped to improve. So the best variables derived from RF are: MX H, H90, H80, H70, H60, Stem, Slope, CGV, CGV90, CGV80, CGV70, and CGV60. Among all the best variables, slope contributes the most to estimate biomass as show in Fig 3.2 (c).

Table 3.2 Accuracy assessment for improved watershed algorithm for 10 plots.

Plot	Reference	Window Size	Matched	Merged	Split	Missed	Accuracy
1	26	5	23	0	0	3	88.50%
2	27	5	26	0	1	0	96.30%
3	8	10	8	0	0	0	100%
4	29	5	23	6	0	0	79.30%
5	12	10	10	2	0	0	83.30%
6	16	7	14	2	0	0	87.50%
7	27	3	20	4	1	2	74.00%
8	13	9	11	2	0	0	84.60%
9	24	7	20	3	0	1	83.30%
10	16	7	15	0	1	0	94%

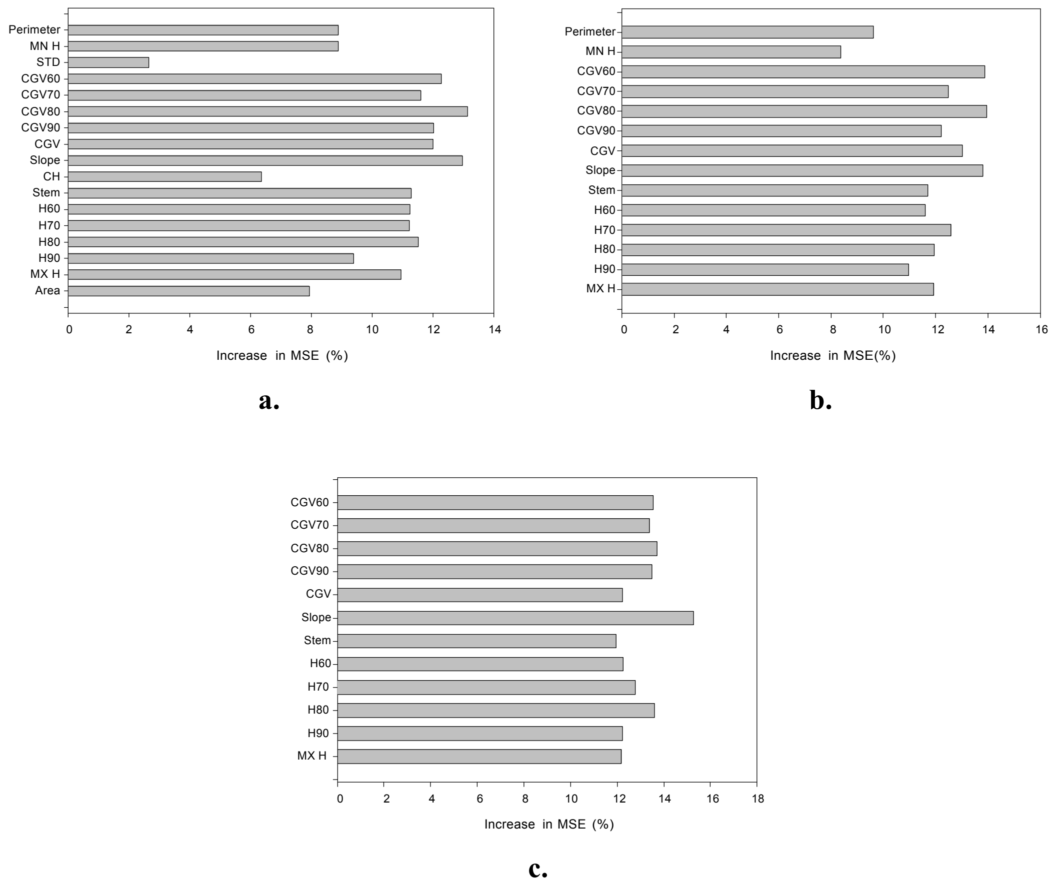
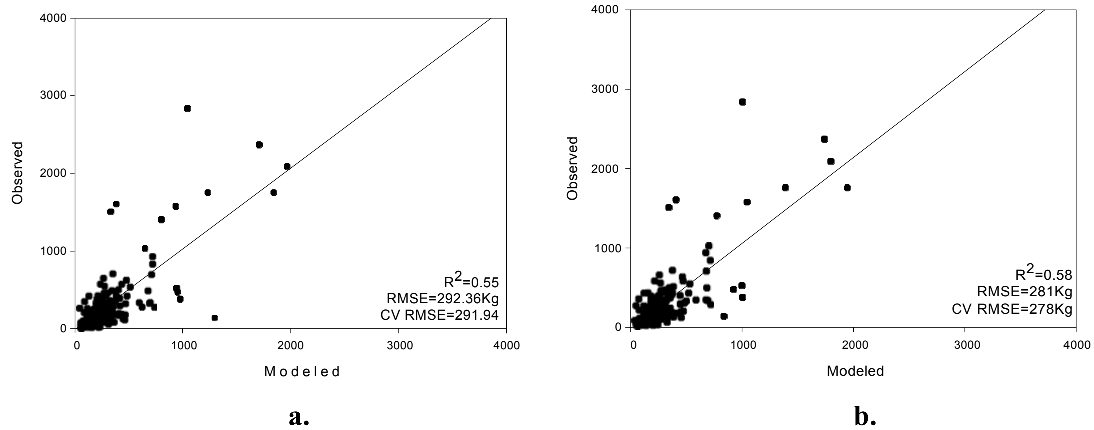
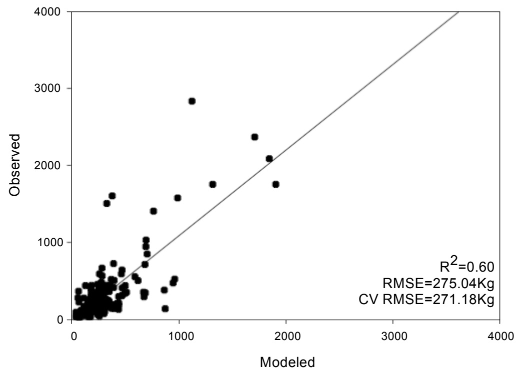


Fig 3.2 Varible importance of three sets of variables for RF.





c.

Fig 3.3 Predicted values compared with observed values using R^2, RMSE, CV RMSE with RF.

However, the results are a bit different for SVR. Since SVR cannot provide the variable importance, the variables combination for testing SVR is different from that of the RF process. First, all the 17 variables were tested in SVR. The results show in Fig 3.4 (a). The next step is trying to reduce the amount of variables. The variables we interested will be removed one by one to see their influence on the biomass. If the accuracy decreased a lot after removing the interested variable, it means the variable is important for the prediction. If the opposite happens, that means the interested variable is not involved much to the process. The whole process for the variables selection and filtering steps is presented in Table 3.3 in detail. As our assumption, the variables we interested most are crown height, stem height, slope and the perimeter.

All the variables were implemented in the first place. The R^2 is around 0.69 as showed in Fig 3.4 (a). Next, the R^2 improved a lot when perimeter was deleted as showed in Fig 3.4 (b). That means perimeter contributes less to estimate the biomass in SVR. In the third place, based on the success of perimeter removal, stem is deleted and the result is showed in Fig 3.4 (c). Similar to perimeter, the removal of the stem didn’t influence the result that much. Fig 3.4 (d) showed that the crown area is also of low importance to estimate biomass with the R^2 as high as 0.72. Until this step, 14 variables were used and the useless variables are perimeter, stem and crown area.

However, Fig 3.4 (e) showed the result is worse and R^2 is declined to 0.67 when the slope was removed. That means slope is necessary for estimating biomass. So in order to test the interested variable, crown height, slope was moved back and crown height was deleted for testing. From Fig 3.4 (f) we can see that the R^2 is also significantly influenced when crown height is missing. In this case, the best combination for the SVR to get better accuracy can be determined at 14 variables. These 14 variables are of great importance and crucial for estimating biomass using SVR. 3 unimportant variables are perimeter, stem and crown area. As the assumption, slope and crown height are very crucial for the estimating process. Slope, again, tested as the critical variable as the results in RF showed above.

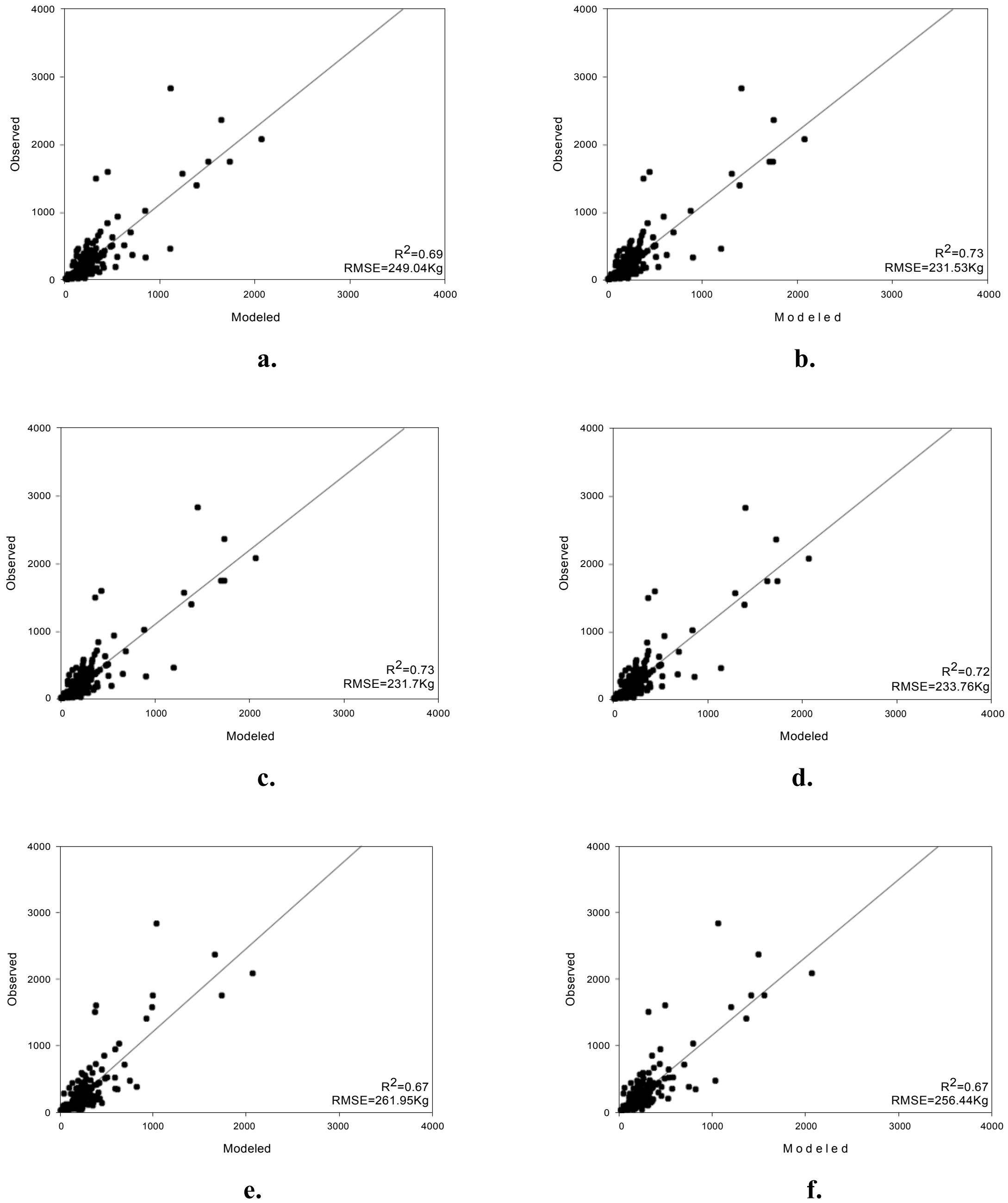
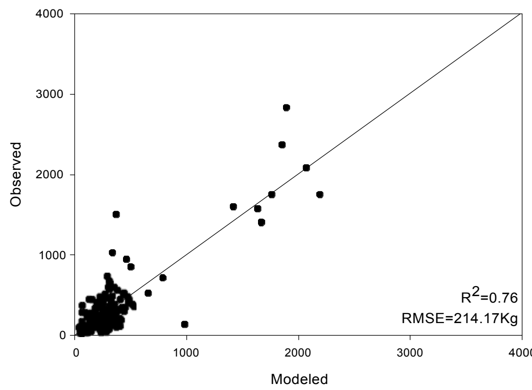


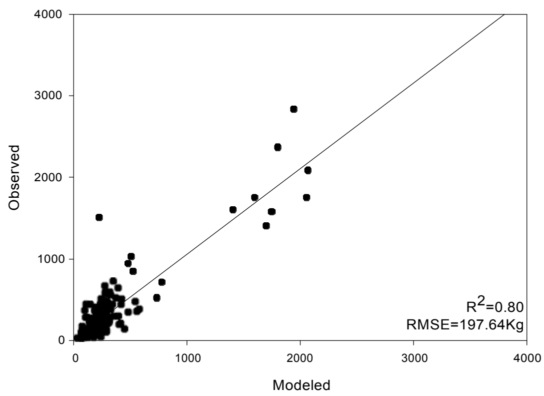
Fig 3.4 Predicted values compared with observed values using R^2, RMSE, with SVR.

Table 3.3 Variables combination using SVR and the performance

Variables selection process						
	a	**b**	**c**	**d**	**e**	**f**
Area	√	√	√			
MX H	√	√	√	√	√	√
H90	√	√	√	√	√	√
H80	√	√	√	√	√	√
H70	√	√	√	√	√	√
H60	√	√	√	√	√	√
Height to crown base	√	√				
CL	√	√	√	√	√	
Slope	√	√	√	√		√
CGV	√	√	√	√	√	√
CGV90	√	√	√	√	√	√
CGV80	√	√	√	√	√	√
CGV70	√	√	√	√	√	√
CGV60	√	√	√	√	√	√
STD	√	√	√	√	√	√
MN H	√	√	√	√	√	√
Perimeter	√					
R^2	0.69	0.73	0.73	0.72	0.67	0.67



a.



b.

Fig 3.5 Predicted values compared with observed values using R^2, RMSE, with Cubist.

Table 3.4 Variable importance in Cubist

	CGV	MX H	Slope	Stem	Area	R^2
17 variables	15%	76%	15%	65%	15%	0.76
5 variables	31%	26%	31%	69%	79%	0.80

For Cubist, since it's the black box whose detailed rules are not available, all the variables were tested in the first place. Cubist can automatically choose the variables from all the candidates can provide the variable importance. The accuracy results were most promising and the R^2 is greater 0.7 and the RMSE is as low as 214.17Kg as showed in Fig 3.5 (a). Five variables were selected and the values of variable importance are showed in Table 3.4 in the first row. Among the five selected variables, tree height and stem height contribute most to the predicting process. The next step is excluding all the other variables can just apply for Cubist to predict biomass with CVG, MX H, slope, Stem and area. The result improved obviously and the R^2 is 0.8 with the RMSE of 197.64 Kg. Stem height still contributes the most to the process. In this case, Cubist use five variables to predict biomass, crown area and stem contributes most to the process, while CGV, MX H, and slope are also necessary for the prediction.

3.5 Discussion

3.5.1 Individual tree crown delineation and the influence on biomass estimation

The innovative method used in this paper for tree crown delineation generated relatively high accuracy than other related studies (Wang, et al., 2004; Chen, et al., 2006; Jing, et al., 2012). The algorithm is based on the traditional watershed approach and improved the result especially of the tree boundaries. There are two issues considered for accuracy of individual tree crown delineation. First, the location for each delineated crown and the reference tree location should be matched. Before watershed segmentation, local maximums were detected as initial tree tops. Watershed boundary was generated from these local maximums. In this case, the amount of local maximum should be somewhat similar to that of the reference trees. If more local maximum were detected, the image will be over segmented and lots of useless trees will reduce the accuracy. On the contrary, if the amount of the local maximum is too small, some trees cannot be detected. So in the algorithm, image filtering and background removal is of high importance in order to generate the proper amount of local maximum and proper location for local maximum. Second, since the boundary of watershed segmentation is provided automatically within the algorithm, which is less related to the image itself. The image used was CHM which derived from LiDAR data. The boundary construction should follow the features of the forest and related to the tree structure. The reference data for tree crown delineation were generated by hand delineation based on the filtered CHM. In the result, crown ratio was used as the proper threshold to find out the accurate boundary of each tree. This process highly improved the accuracy since crown ratio perfectly connected the height information and boundary together. In the end, the accuracy is improved since the proper local maximums were generated and the visual result compared with manually delineated crowns was also improved. A few researches considered the boundary problem for watershed (Solberg, et al., 2006), what is more, it is the first time to introduce the forest related variable, crown ratio, for crown delineation. Individual tree crown delineation is not only related to digital processing field, but also part of forest studies. The process should base on the feature of individual trees.

Biomass estimation was applied after individual tree delineation. The biomass was predicted on individual tree level, and the reference is also calculated on the individual tree level. Before the estimation, one to one match process is significantly important since it will determine which reference tree are detected by which delineated crown. In some cases, one reference tree crown maybe segmented into two pieces after the segmentation. One reference crown will correspond to two delineated crowns. The next step was to extract the variables and values within each crown. As listed above in the method section, 17 variables were extracted, such as maxim height, minimum height, crown area, perimeter. Besides slope information is derived from DSM within each crown boundary, other variables were from CHM. In this case, the shape, size and location of each delineated crowns were very important since different shape, size and location will generate different value listed above.

The location of each tree crown was determined by the local maximum in crown delineation process; meanwhile, the location of each crown will have large influence on the maximum height within this crown. In addition, for tree crown delineation, the crown shape and size were determined by crown ratio, which are the stop criteria for boundary detection. For biomass estimation, the shape and size will influence the interested variables like perimeter, crown area. In this case, the quality of the delineated crowns will directly influence the variables needed for biomass estimation and further impact the performance of biomass estimation.

3.5.2 The performance of ABG estimation using three machine learning models

In order to predict biomass in three distinct models, various variables were extracted. Some are commonly used variables like height, volume and percentile of them. Since our biomass reference for individual tree is the sum of biomass for different portion: stem and leaves, the variables we interested in which other researches seldom thought about are crown height, and stem height. In addition, the study area is mountainous and rugged, mean slope of each crown was also used as the variables. The trees grow on the mountainous area should have different amount of biomass. The higher the slope, more light will be given so it was assumed that higher biomass will be generated. The initial assumption was that stem height, crown height and slope should have a high relationship with the biomass so highly variable importance should be given more to these variables than the other. RF, SVR and Cubist tested these variables separately to validate the assumptions.

First, RF was tested to detect the most significant variables. As it showed in Fig 3.2 (c), the combination of variables for MX H, H90, H80, H70, H60, Stem, Slope, CGV, CGV90, CGV80, CGV70, and CGV60 got the highest accuracy to predict biomass. Beside the variables related to tree height, volume and their percentiles, stem and slope were both included for the best combination. What is more, from Fig 3.2 (c), slope contributed most to the estimation process. The increase in MSE for slope is as high as 15.27%. Although stem offered the least in Fig 3.2 (c), the increase in MSE is already nearly 12%. In this case, for RF, our assumption almost validated that stem and slope should play an important role for biomass estimation. However, crown height was not included as variables in the best combination for RF. Our delineated crown was followed the crown ratio as 0.3 as the threshold, so the crown height can be interpreted as the 30 percentile of the tree height. In this case, since the percentiles from 90 to 60 have already predicted biomass properly, 30 percentile of the crown, and the crown height, is not of high importance in this case.

Secondly, because there is no variables importance provided by SVR, the method for best variables selection is different from that of RF. After several trials of variables showed in Table 2, crown area, stem height, and perimeter don’t contribute much to the whole process. However, for SVR, slope and crown height are very important to derive higher accuracy. There is no variable importance for each variable so no conclusion can be derived to see which one contribute the most in quantify condition. However, yes and no conclusion can be derived after the process. Same as the conclusion from RF, slope is still the important variables we need for biomass estimation. For SVR, crown height contributes more than the stem height. Our assumption is still validated since slope is still important, and crown height is portion based variables, which should be tested as the important variable for biomass estimation.

Lastly, Cubist was applied for detecting the most related variables for biomass estimation. Due to the unknown process inside Cubist, only part of the variables will be selected for estimating. Five variables were selected and importance for each of them is list in Table 4. For the first case, among the 17 variables, stem height is the most important variable for the estimation. For the second case, within the five variables, both stem height and crown area contribute a lot to the process. It is obviously to see that the stem height and crown area and the maximum tree height are the most essential variables for Cubist. Since the second case generated a little higher R^2, stem height was demonstrated again as the important variable. However, Cubist doesn't rely on a slope too much but it still offers more than 30% in the second case. In conclusion, slope and stem height are of high importance for estimation biomass in Cubist.

From the discussion above we can see, RF relies much on slope and stem height; Slope and crown height are necessary for SVR and Cubist use both slope and stem height for the process. In this case, slope was validated three times as the important input variable for estimating biomass. Stem height and crown height were also important variables. In general, most paper utilized the common variables like tree height, volume and percentiles. For example, 75 percentile of the tree height is of high importance to generate AGB (Ioki, Imanishi, Sasaki, Morimoto, & Kitada 2010; Sun et al., 2011). Mean height is explained as the main variable for the estimating process (Dalponte, Martinez, Rodeghiero, & Gianelle, 2011; Clark, Roberts, Ewel, & Clark, 2011). Higher percentile of heights will better help for the predicting process was also demonstrated (Li, et al., 2014). However, in our study slope plays the most important character for biomass estimation. All the studies have used the percentile of height as the input and claimed that higher or mean height are important. In our case, stem height and crown height can be also called as percentile of the tree height; however, we can give these two variables a proper explanation related to the feature of trees for the performance of biomass estimation. Crown height and stem height are all derived based on individual tree crown delineation. In that process, the boundary of each tree was set using crown ratio. When the boundary was set, crown height and stem height can be derived. This means trees are separated into two prominent and different portions. As equation we used for reference calculation, crown and stem are different portions with different relationship between their height and their biomass. Based on this scientific and logical separation for the tree height into crown height and stem height, the two variables will be more related to the biomass estimation than other percentile based height variables.

In addition, Li didn't detect a strong relationship between slope and biomass estimation on a mountainous area. However, in our case, slope was validated three times that it acts as the potent factor contributing to estimate biomass. From one hand, slope can affect the biomass since more light is provided and less competitive on the slope area. On the other hand, there are some shortcomings for slope variables. Other researchers already made some progress on biomass in mountainous region (Sun et al., 2002). They found the terrain effect on DEM will influence the biomass estimation. So they developed the algorithm to reduce the topographic effect. So the slope can be interpreted as the two fold variable which can affect the accuracy of height extraction and meanwhile has the high relationship between biomass and itself.

The result shows that we almost reached our assumption that slope, stem height and crown height are important for estimating biomass. The success depends on the superior accuracy of individual tree crown delineation and the variable selection. However, several issues need further investigation. First, there are 198 trees delineated but only 151 were used for biomass estimation after the one to one match process. That means no proper delineated crowns for these left reference trees. For instance, because of the complexity of the forest and the discrete of the LiDAR data, some reference tree locates where it seems the valley of that area. The reference data from field work can not 100% matched with the LiDAR derived images. Secondly, regardless of the crown height and stem height are first introduced in the biomass studies and contribute a lot to biomass estimation as the result shows, there are still some shortcomings. Since the crown ratio of each tree is varied from species to species and from time to time, it is not that proper to set the unique crown ratio value to the whole image for tree crown delineation. In this case, the hinted assumption is that all the trees are of the same feature with the same crown ratio. The boundaries of all the trees are not 100% identifiable since the variety of the trees age and species. So some irregular shape of the delineated crowns can be found and some crowns are much larger than the others. The irregular crowns will influence the value extracted for biomass estimation and lead some errors. In this case, proper method to determine the crown ratio automatic and multi-crown ratio is needed for different kinds of tree and different age of trees. Thirdly, except RF, the selection for SVR is based on several trials. Since there are 17 variables and more than 1000 of combinations can be found within the entire variables. Traversal is not applied for SVR so maybe there are some better combinations but we only focus on the variables we are interested in as slope, stem height and crown height. For Cubist, the detailed method inside the model is not clear, but the result is very promising.

3.6 Conclusion

In this study, three machine learning models were compared to estimate biomass and find out the important variable for the process based on individual tree crown delineation. Cubist provides the most accurate and fastest model for predicting biomass with the R^2 of 0.8. With high accuracy of individual tree crown delineation, the initial assumptions are validated and biomass was estimated properly. According to the result, slope, stem height and crown height do attribute more and of high importance for the predict process. This conclusion provides the new variables for future studies related to biomass estimation. Stem height, crown height are the important variables for biomass estimation since they are tree structure related variables and they directly represent the biomass condition from a different portion of the tree. This will make the variable more convincing and scientific. In addition, the conclusion also is also reported that AGB is largely influenced by the slope and terrain conditions. Productivity and AGB amount of trees on the mountainous area should be different from other trees on the flat ground, and the prediction process should take slope into consideration.

4. CONCLUSION

The previous two sections have introduced an effective process for biomass estimation solo from LiDAR data, and have presented an effective way to delineate tree crowns. Section two outlined the methods for individual tree crown delineation and clear defined the criteria for individual tree boundary detection. Crown ratio worked as the threshold for boundary searching which makes the result more reliable and scientific with the average accuracy of over 85%. This process provides solid foundation for the further study in section 3 about biomass estimation.

In section 3, three different kinds of machine learning approaches were applied not only to derive accurate biomass result, but also trying to find out the important variables for AGB estimation. 17 variables were extracted based on the delineated crowns in section 2. Besides the traditional variables like height and volume, slope, stem height, crown height contributed a lot to the estimation process. Since the study area is mountainous, the assumption that there is a strong relationship between the terrain condition and the AGB was validated. In addition, stem height and crown height are individual tree portion based variables, which are also closely related to the AGB. Cubist provides the most accurate result for predicting biomass with the R^2 is around 0.8. Further studies can test the novel algorithm effectiveness for individual tree crown delineation on deciduous forests. More carbon stock related studies can also be applied after the ABG estimation process to test the variables importance.

5. REFERENCE

Ahmed, R., Siqueira, P., & Hensley, S. (2013). A study of forest biomass estimates from lidar in the northern temperate forests of New England. *Remote Sensing of Environment*, *130*, 121-135.

Allouis, T., Durrieu, S., Véga, C., & Couteron, P. (2013). Stem volume and above-ground biomass estimation of individual pine trees from LiDAR data: Contribution of full-waveform signals. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*(2), 924-934.

Ardila, J. P., Bijker, W., Tolpekin, V. A., & Stein, A. (2012). Multitemporal change detection of urban trees using localized region-based active contours in VHR images. *Remote Sensing of Environment*, *124*, 413-426.

Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., ... & Goodwin, N. (2013). Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sensing of Environment*, *134*, 24-38.

Barbosa, J. M., Melendez-Pastor, I., Navarro-Pedreño, J., & Bitencourt, M. D. (2014). Remotely sensed biomass over steep slopes: An evaluation among successional stands of the Atlantic Forest, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *88*, 91-100.

Becknell, J. M., Kissing Kucek, L., & Powers, J. S. (2012). Aboveground biomass in mature and secondary seasonally dry tropical forests: A literature review and global synthesis. *Forest Ecology and Management*, *276*, 88-95.

Beucher.S., Lantuejoul. C.(1979). Use of watersheds in contour detection. *In International Workshop on Image Processing*: Real-time Edge and Motion Detection/Estimation, Rennes, France.

Bortolot, Z. J., & Wynne, R. H. (2005). Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *59*(6), 342-360.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, *112*(10), 3876-3890.

Brandtberg, T., Warner, T. A., Landenberger, R. E., & McGraw, J. B. (2003). Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America.*Remote sensing of Environment*, *85*(3), 290-303.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, *114*(4), 911-924.

Bunting, P., & Lucas, R. (2006). The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote Sensing of Environment*, *101*(2), 230-248.

Bunting, P., Lucas, R. M., Jones, K., & Bean, A. R. (2010). Characterisation and mapping of forest communities by clustering individual tree crowns. *Remote Sensing of Environment*, *114*(11), 2536-2547.

Camps-Valls, G., Bruzzone, L., Rojo-Álvarez, J. L., & Melgani, F. (2006). Robust support vector regression for biophysical variable estimation from remotely sensed images. *Geoscience and Remote Sensing Letters, IEEE*, *3*(3), 339-343.

Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Martín-Guerrero, J. D., Soria-Olivas, E., Alonso-Chordá, L., & Moreno, J. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery.*Geoscience and Remote Sensing, IEEE Transactions on*, *42*(7), 1530-1542.

Castillo-Núñez, M., Sánchez-Azofeifa, G. A., Croitoru, A., Rivard, B., Calvo-Alvarado, J., & Dubayah, R. O. (2011). Delineation of secondary succession mechanisms for tropical dry forests using LiDAR. *Remote Sensing of Environment*, *115*(9), 2217-2231.

Chambers, J. Q., Asner, G. P., Morton, D. C., Anderson, L. O., Saatchi, S. S., Espírito-Santo, F. D., ... & Souza Jr, C. (2007). Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. *Trends in Ecology & Evolution*, *22*(8), 414-423.

Chen, J., Zheng, G., Fang, C., Zhang, N., Chen, H., & Wu, Z. (2014). Time-series processing of large scale remote sensing data with extreme learning machine. *Neurocomputing*, *128*, 199-206.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering & Remote Sensing*, *72*(8), 923-932.

Clark, M. L., Roberts, D. A., Ewel, J. J., & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment*, *115*(11), 2931-2942.

Colgan, M. S., Baldeck, C. A., Féret, J. B., & Asner, G. P. (2012). Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data.*Remote Sensing*, *4*(11), 3462-3480.

Dalponte, M., Martinez, C., Rodeghiero, M., & Gianelle, D. (2011). The role of ground reference data collection in the prediction of stem volume with LiDAR data in mountain areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), 787-797.

De Castilho, C. V., Magnusson, W. E., de Araújo, R. N. O., Luizao, R. C., Luizao, F. J., Lima, A. P., & Higuchi, N. (2006). Variation in aboveground tree live biomass in a central Amazonian Forest: Effects of soil and topography.*Forest ecology and management*, *234*(1), 85-96.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., & Hofton, M. (2003).

Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global ecology and biogeography*, *12*(2), 147-159.

Ene, L., Næsset, E., & Gobakken, T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. *International Journal of Remote Sensing*, *33*(16), 5171-5193.

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

Hirschmugl, M., Ofner, M., Raggam, J., & Schardt, M. (2007). Single tree detection in very high resolution remote sensing data. *Remote Sensing of Environment*, *110*(4), 533-544.

Horvath, P., Jermyn, I. H., Kato, Z., & Zerubia, J. (2009). A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction.*Pattern Recognition*, *42*(5), 699-709.

Hu, B., Li, J., Jing, L., & Judah, A. (2014). Improving the efficiency and accuracy of individual tree crown delineation from high-density LiDAR data.*International Journal of Applied Earth Observation and Geoinformation*, *26*, 145-155.

Huang, C., Song, K., Kim, S., Townshend, J. R., Davis, P., Masek, J. G., & Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*(3), 970-985.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, *112*(5), 2232-2245.

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

Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J., & Serrato, M. (2012). Vegetation cover analysis of hazardous waste sites in Utah and Arizona using hyperspectral remote sensing. *Remote Sensing*, *4*(2), 327-353.

Ioki, K., Imanishi, J., Sasaki, T., Morimoto, Y., & Kitada, K. (2010). Estimating stand volume in broad-leaved forest using discrete-return LiDAR: plot-based approach. *Landscape and Ecological Engineering*, *6*(1), 29-36.

Jing, L., Hu, B., Li, J., & Noland, T. (2012). Automated delineation of individual tree crowns from LiDAR data by multi-scale analysis and segmentation.*Photogrammetric Engineering & Remote Sensing*, *78*(12), 1275-1284.

Jing, L., Hu, B., Noland, T., & Li, J. (2012). An individual tree crown delineation method based on multi-scale segmentation of imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 88-98.

Jung, S. E., Kwak, D. A., Park, T., Lee, W. K., & Yoo, S. (2011). Estimating crown variables of individual trees using airborne and terrestrial laser scanners.*Remote Sensing*, *3*(11), 2346-2363.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., ... & Wu, J. C. (2012). An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sensing*, *4*(4), 950-974.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*(5), 352-359.

Ke, Y., & Quackenbush, L. J. (2011). A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing.*International Journal of Remote Sensing*, *32*(17), 4725-4747.

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.

Koukoulas, S., & Blackburn, G. A. (2005). Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing*,*26*(3), 431-455.

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

Kuprevicius, A., Auty, D., Achim, A., & Caspersen, J. P. (2013). Quantifying the influence of live crown ratio on the mechanical properties of clear wood.*Forestry*, *86*(3), 361-369.

Kwak, D. A., Lee, W. K., Lee, J. H., Biging, G. S., & Gong, P. (2007). Detection of individual trees and estimation of tree height using LiDAR data.*Journal of Forest Research*, *12*(6), 425-434.

Larsen, M., Eriksson, M., Descombes, X., Perrin, G., Brandtberg, T., & Gougeon, F. A. (2011). Comparison of six individual tree crown detection algorithms evaluated under varying forest conditions. *International Journal of Remote Sensing*, *32*(20), 5827-5852.

Leckie, D. G., Gougeon, F. A., Tinis, S., Nelson, T., Burnett, C. N., & Paradine, D. (2005). Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sensing of Environment*, *94*(3), 311-326.

Lefsky, M. A., Turner, D. P., Guzy, M., & Cohen, W. B. (2005). Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modeled forest productivity. *Remote Sensing of Environment*, *95*(4), 549-558.

Li, J., Hu, B., & Noland, T. L. (2013). Classification of tree species based on structural features derived from high density LiDAR data. *Agricultural and Forest Meteorology*, *171*, 104-114.

Liu, J., Shen, J., Zhao, R., & Xu, S. (2013). Extraction of individual tree crowns from airborne LiDAR data in human settlements. *Mathematical and Computer Modelling*, *58*(3), 524-535.

Lucas, R. M., Cronin, N., Lee, A., Moghaddam, M., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, Queensland, Australia. *Remote Sensing of Environment*,*100*(3), 407-425.

Moffiet, T., Mengersen, K., Witte, C., King, R., & Denham, R. (2005). Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry and Remote Sensing*, *59*(5), 289-309.

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.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Ståhl, G., & Weydahl, D. (2011). Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area.*Remote Sensing of Environment*, *115*(12), 3599-3614.

Popescu, S. C., & Wynne, R. H. (2004). Seeing the Trees in the Forest.*Photogrammetric Engineering & Remote Sensing*, *70*(5), 589-604.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, *31*(9), 646-655.

Pouliot, D. A., King, D. J., Bell, F. W., & Pitt, D. G. (2002). Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, *82*(2), 322-334.

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., & Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, *114*(5), 1053-1068.

Reitberger, J., Schnörr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(6), 561-574.

Solberg, S., Naesset, E., & Bollandsas, O. M. (2006). Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest.*Photogrammetric Engineering & Remote Sensing*, *72*(12), 1369-1378.

Straub, C., & Koch, B. (2011). Estimating single tree stem volume of Pinus sylvestris using airborne laser scanner and multispectral line scanner data.*Remote Sensing*, *3*(5), 929-944.

Sun, G., Ranson, K. J., & Kharuk, V. I. (2002). Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains, Siberia. *Remote Sensing of Environment*, *79*(2), 279-287.

Sun, G., Ranson, K. J., Guo, Z., Zhang, Z., Montesano, P., & Kimes, D. (2011). Forest biomass mapping from lidar and radar synergies. *Remote Sensing of Environment*, *115*(11), 2906-2916.

Tian, X., Su, Z., Chen, E., Li, Z., van der Tol, C., Guo, J., & He, Q. (2012). Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. *International journal of applied earth observation and geoinformation*, *14*(1), 160-168.

Vastaranta, M., Kankare, V., Holopainen, M., Yu, X., Hyyppä, J., & Hyyppä, H. (2012). Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 73-79.

Vauhkonen, J., Korpela, I., Maltamo, M., & Tokola, T. (2010). Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, *114*(6), 1263-1276.

Véga, C., & Durrieu, S. (2011). Multi-level filtering segmentation to measure individual tree parameters based on Lidar data: Application to a mountainous forest with heterogeneous stands. *International Journal of Applied Earth Observation and Geoinformation*, *13*(4), 646-656.

Walton, J. T. (2008). Subpixel Urban Land Cover Estimation. *Photogrammetric Engineering & Remote Sensing*, *74*(10), 1213-1222.

Wang, L., Gong, P., & Biging, G. S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering & Remote Sensing*, *70*(3), 351-357.

Wulder, M., Niemann, K. O., & Goodenough, D. G. (2000). Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of environment*, *73*(1), 103-114.

Yang, X., Strahler, A. H., Schaaf, C. B., Jupp, D. L., Yao, T., Zhao, F., ... & Ni-Meister, W. (2013). Three-dimensional forest reconstruction and structural parameter retrievals using a terrestrial full-waveform lidar instrument (Echidna®).*Remote sensing of environment*, *135*, 36-51.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., ... & Strahler, A. (2011). Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. *Remote sensing of Environment*, *115*(11), 2965-2974.

Zhang, W., Quackenbush, L. J., Im, J., & Zhang, L. (2012). Indicators for separating undesirable and well-delineated tree crowns in high spatial resolution images. *International Journal of Remote Sensing*, *33*(17), 5451-5472.

Zhao, F., Guo, Q., & Kelly, M. (2012). Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. *Agricultural and Forest Meteorology*, *165*, 64-72.

Zhao, K., Popescu, S., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning.*Remote Sensing of Environment*, *115*(8), 1978-1996.

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, *113*(1), 182-196.

6. ACKNOWLEDGEMENTS

When constructing this thesis, I have been a student who study geography related topics for six years. The thesis is the culmination from what I have learned from the past time, and reflecting the support and effort I have received during the past six years. I really appreciated all the people that have instructed and supported me.

My most sincere thanks go to my advisor, Dr. Jungho Im. His wide knowledge and sincerity attracted me a lot and lead me to construct researches in South Korea, to fulfill my research goal and aims. My research ability and study ability improves a lot within these two years. He is always patient and available to answer my questions, share his ideas and guide me to generate novel thoughts. I also feel amazing about is high standards for teaching which assist me a lot during the research process.

In the second place, I would like to thank my committee. Two professors: Dr. Myong In Lee, and Dr. Yoon Dong Keun. I appreciated their help and effort on my graduation thesis. In addition, I have benefited a lot from their courses. I have expanding my knowledge range in Dr. Lee's courses like Climate-Environment Modeling and enhanced my GIS and remote sensing skills in Dr. Yoon's Disaster Risk Assessment course.

I also would like to thank all the professors in our department, especially when I took some courses under their instruction. All the professors are kind and patient to offer me ideas and knowledge. The most important, they always tried to help me to combine what I have learned in the class with my undergoing research. I would also like to thank all the classmates and friends I met during the class. They also shared new ideas with me and whenever I have some difficulties, they were pretty willing to help.

At last, I also would like to thank all my friends both in South Korea and China, and my family members. Their help and encouragement help a lot for me to keep pursuing my academic goal. Thanks to the chance to study here is UNIST, where I learned a lot both for study, and for my entire life.