Derivation of Forest type information from large-footprint spaceborne Lidar waveforms:

- a case study in cool temperate forests of Northeastern China -

Zhang Junjie

February, 2008

Derivation of Forest type information from large-footprint spaceborne Lidar waveforms:

- a case study in cool temperate forests of Northeastern China -

by

Zhang Junjie

Thesis submitted to the International Institute for Geo-information Science and Earth Observation (The Netherlands) and the School of Resources and Environmental Sciences (SRES) of Wuhan University (China) in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geo-information for Natural Resource and Environmental Management.

Thesis Assessment Board

Dr. Michael. Weir, Chairman (ITC)
Dr. Liu Xuehua, External examiner (Tsinghua University)
Prof. Dr. Ir. Alfred de Gier, Supervisor (ITC)
Prof. Dr. Du Qingyun, Supervisor (Wuhan University)

Supervisors: **Prof. Dr. Ir. Alfred de Gier** (ITC), **Prof. Dr. Du Qingyun** (Wuhan University)

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS
SCHOOL OF RESOURCES AND ENVIRONMENTAL SCIENCES (SRES)
WUHAN UNIVERSITY (CHINA)

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Dedicated to

My Family

and lanlan

Abstract

Forest aboveground biomass is of great importance for a better understanding of the global carbon cycle. Lidar (Light detection and ranging) remote sensing has proven to be a promising way of estimating forest canopy height, which has a strong relationship with forest aboveground biomass. However, forest type information is essential for a more accurate aboveground biomass estimation, due to the fact that forest biomass may vary considerably between different forest types, even when of similar height, density and age.

Lidar emits a shot duration laser pulse, and records the time and amplitude of the backscattered laser energy reflected back from the Earth's surface. The composite return form a large-footprint lidar, referred to as a lidar waveform, provides a finely resolved measure of the vertical distribution of the illuminated surface area within the footprint. However, no study have successfully built a relationship between lidar waveform profiles and forest types.

In this study, the relationship between the lidar waveform characteristics and different forest types was further investigated, to achieve the goal of deriving forest type information from large-footprint spaceborne lidar. Fieldwork was conducted in the cool temperate forest of Northeastern China, in September 2007, which consists of a mixture of conifers and deciduous tree species. A full waveform dataset of the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat) with cloud-free profiles over the study area was acquired.

The study developed a new method to extract the required forest type information from large-footprint lidar waveforms based on full waveform analysis. For this purpose, the raw waveform was decomposed into Gaussian components, and canopy return and ground return of the waveform were separated. Two types of metrics hypothesised to have relationship with forest types were derived. The first type of metric, R25, R50 and R75, is quantile-based metrics based on the vertical distribution of canopy return energy; and the second type, AGS, SGS and MSGS, are statistical characteristics based on the decomposed Gaussian components of canopy return part. Different combinations of the metrics derived were then used as input parameters in Support Vector Machine classification to find their relationship with different forest types. And the highest accuracy of 90.57% and Kappa statistic of 0.7868 were achieved applying this method to separating broadleaved and needleleaved forests.

The results showed that the first type of metrics, R25, R50 and R75, reflecting canopy vertical structure distribution failed to distinguish broadleaved and needleleaved forests. In contrast, the second type of metrics, especially metrics combination of AGS and MSGS, showed great promise in distinguishing the two types of forest. An overall accuracy of 90.57% and Kappa statistic of 0.7868 was achieved using AGS and MSGS as input parameters in SVM classification. However, the metrics AGS and MSGS seem to fail to distinguish mixed broad-/needle-leaved forest from broadleaved and needleleaved forest.

The result indicated that the decomposed Gaussian components in the lidar waveform can well mirror the characteristics of the reflecting Gaussian surfaces, in the forest sense, the layers of branches and foliage, which differ between broadleaved forest and needleleaved forest. Utilizing this feature of large-footprint lidar waveforms, the broadleaved and needleleaved forest can be successfully separated.

The study established a new way of deriving forest type information from spaceborne large-footprint waveform lidar, thus further extend the application of large-footprint lidar.

Key words: Lidar, ICESat, large-footprint, full waveform analysis, Gaussian component, metrics, forest type, broadleaved forest, needleleaved forest, classification

Acknowledgements

Upon the accomplishment of my MSc thesis, I have so many persons and organizations to give my sincere thanks and deep appreciation to, for their great support and encouragement during the whole course of my study.

My deep gratitude goes to my supervisors. To Prof. Dr. Alfred de Gier, thank you for your invaluable guidance in my study. Your humours talking, critical comments, and encouragement always inspire me to come over the difficulties encountered and to go even further in my research. To Prof. Dr. Du Qingyun, thank you for your great support from the right beginning of my postgraduate study. Your generosity, precise attitude towards science, and support, helped me to complete my work in the end. All these makes me feel lucky to have you as my supervisors.

I would like to acknowledge the International Institution for Geo-Information Science and Earth Observation (ITC) and Wuhan University for giving my this precious chance to study in the Netherlands. This experience broadened my view and greatly extended my knowledge.

To Dr. Jan de Leeuw and Dr. Michael Weir, I thank you for your kindness, concern and continuous help during our study in ITC from the first day we went there. I also want to express my appreciation to Prof. Dr. Andraw Skidmore, Prof. Dr. George Vosselman, Dr. Norman Kerle, Dr. Y.A. Hussin and some other I may not list your names, who gave me instructions in ITC and enriched my knowledge and skills in the field of Remote Sensing and GIS. My genuine thanks also goes to all the Chinese and International friends for the wonderful time together. I believe the friendship will last forever.

My special thanks goes to Ph. D. candidate Wu Guofeng (now Dr. Wu Guofeng), who is also my former undergraduate advisor, for your consist help, valuable advice and unlimited support during my research; to Ph. D. candidate Xing Yanqiu, for introducing me into your research project and giving me this great opportunity to undertake the research in this area. I am also grateful to Yanqiu, David, Ashu, Ram, Xu Huadong, and the staffs of Wangqing Forestry Farm who assisted me in the fieldwork in the forest area, for their help, companion and sharing of joy. It was an incredible experience.

I will always remember the days I spent in ITC, especially the happy times enjoyed with the fellow Chinese friends went together to the Netherland. I will also remember the wonderful times I had in Room 501, SRES, Wuhan University in China, very kind teachers and good friends, creating a strenuous atmosphere to study and work in. Thank you all the people who become part of my unforgettable experiences.

Finally, I wish to express my everlasting gratitude to my family, my uncles and aunts, for their infinite support, both physically and mentally, and what's more important, eternal love, during the course of my life.

My heartfelt thanks are sent to my special someone, my love Peng Shengnan, for helping me find myself back and teaching me what is love.

Table of contents

List of figures

List of tables

List of abbreviations

AGBM	Forest Aboveground Biomass
AGS	Average of the decomposed Gaussian curve slopes
CH25	Height of 25% lidar canopy return energy
CH50	Height of 50% lidar canopy return energy
CH75	Height of 75% lidar canopy return energy
DBH	Average diameter at breast height
GLAS	Geoscience Laser Altimeter System
GRND	Ground Return ratio
HOME	Height of Median Energy
HTRT	HOME/LHT ratio
ICESat	The Ice, Cloud, and land Elevation Satellite
LAI	Leaf Area Index
LHT	Lidar canopy height
MSGS	Modified standard deviation of the decomposed Gaussian curve slopes
NDVI	Normalized Difference Vegetation Index
R25	CH25/LHT ratio
R50	CH50/LHT ratio
R75	CH75/LHT ratio
SAR	Synthetic Aperture Radar
SGS	The standard deviation of the decomposed Gaussian curve slopes

1. INTRODUCTION

1.1. Research Background

1.1.1. Carbon and global climate change

The average temperature of the earth's near-surface and ocean has increased dramatically as compared to the historical data during the last decades which is referred to as global warming. Greenhouse gases, including water vapor, carbon dioxide, methane, nitrous oxide, and ozone, are believed to have played an important role in this global climate change (Patenaude et al., 2005; UNFCCC, 1997). The sunlight reaches the earth as short-wavelength and most are absorbed and warmed the Earth. In turn, the Earth emits long wave radiation back to the atmosphere. The greenhouse gases absorb a fraction of the energy and then emit long wave radiation both towards space and back to the Earth. The downward emitted energy further warms the surface of the Earth which resulting to more warming of the Earth's surface. This so-called "greenhouse effect" process is increased by anthropogenic emission of greenhouse gases, a large portion of which is Carbon dioxide.

Carbon dioxide (CO_2) is considered as one of the most abundant greenhouse gases and a primary agent of global warming. It constitutes 72% of the anthropogenic greenhouse gases and cause 9-26% of the greenhouse effect (Kiehl and Treberth, 1997). Hence, several important international agreements, such as the United Nations Framework Convention on Climate Change (UNFCCC) which was signed in 1992 (UNFCCC, 2002), and the Kyoto Protocol which was adapted in 1997 and entered into force on 16 February 2005 after 55 Parties had ratified, address the issue of reducing emissions of anthropogenic greenhouse gas, especially CO_2, into the atmosphere.

1.1.2. Forest carbon stock and forest biomass assessment

Forest ecosystems play an important role in global carbon budget because they are recognized as one of the largest reserves of terrestrial carbon (Figure 1-1). About 80% of all aboveground and 40% of all belowground terrestrial organic carbon are stored in forests (IPCC, 2001). Forests can influence climate change by affecting the amount of carbon dioxide in the atmosphere through natural (e.g., photosynthesis and respiration of plants) and anthropogenic (e.g. deforestation and aforestation) process (Brown et al., 1996; Dixon et al., 1994; Patenaude et al., 2005). During productive season, forest receives the solar energy and takes the CO_2 from the atmosphere and stores it in plant biomass (Phat et al., 2004; Losi et al., 2003). When vegetation decomposes or disturbed by human or natural causes, carbon is released back to the atmosphere. For this reason, it is necessary to make the forest a carbon sink rather than source.

Carbon sequestration in forest can be studied through the assessment of forest aboveground biomass (AGBM) (Aboal et al., 2005; Kraenzel et al., 2003; Laclau, 2003; Losi et al., 2003), because it accounts for the majority of the total biomass in the forest ecosystem and carbon is approximately 46% of all dry wood weight (DeGier, 2003) which is more generalized as 50% (Drake et al., 2002; IPCC,

2003). That is why a transparent reporting of forest removal and accumulation (biomass change) is required by the Kyoto protocol.

What's more, forests not only function as carbon stock, they are also of great importance in providing renewable raw materials and energy, maintaining biodiversity, protecting land and water resources (FAO, 2007). Quantification of forest structure characteristics (e.g., forest biomass) is essential to provide us with information to ensure sustainable forestry (Parresol and Bernard, 1999). For resource use purpose, it is important to know how much fuel wood or timber, thus how much biomass is available there. For environmental management purpose, it is important to assess the productivity and sustainability of forest by quantification of biomass. Thus, it is crucial to address acceptable methods, skills and terminology are important for accurately estimating the forest aboveground biomass at broad spatial scale.

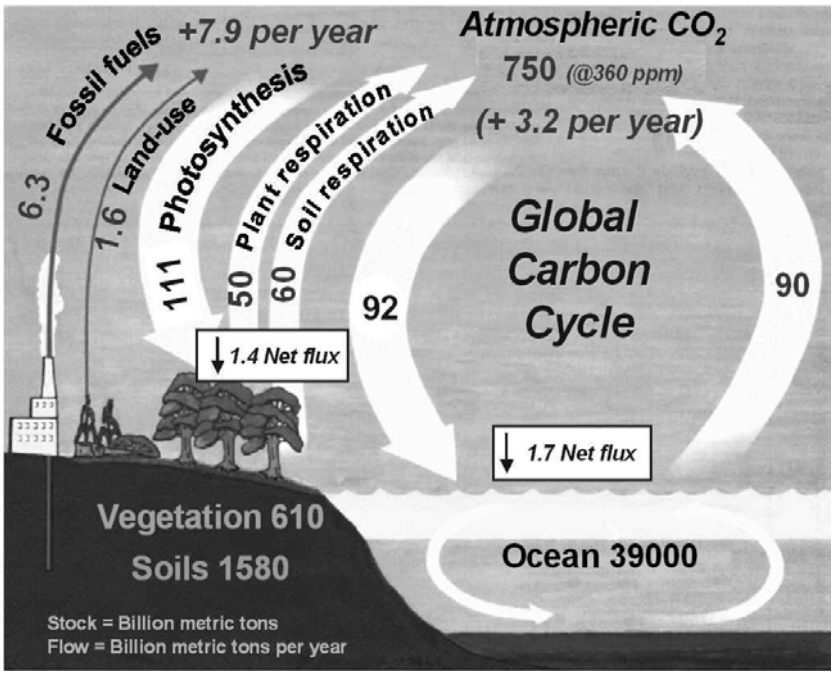


Figure 1-1 The global carbon cycle
(Source: http://www.globalchange.umich.edu/globalchange1/current/lectures/kling/carbon_cycle/carbon_cycle.jpg)

1.1.3. Remote sensing application in forest biomass estimation

Remote sensing provide us with a ideal tool in forest biomass estimation, for it is suitable for monitoring terrestrial ecosystems at various temporal and spatial scale. However, the conventional field measurement, through accurate (Lu, 2006), proved to be very costly and time consuming (DeGier, 2003), which made it unrealistic in biomass estimation over a large spatial scale. Nevertheless, also in remote sensing-based biomass estimation, field measurement is needed for predictive modeling or validation purpose.

Remote sensing techniques have been widely tested for land cover mapping and other forest applications since 1970s (Patenaude et al., 2005; Rosenqvist et al., 2003). It is useful for biomass estimation since there is an empirical correlation between forest biomass and the intensity of EM energy (or the ratio of energy at different wavelength) received by the instrument (Drake et al., 2002). Furthermore, remote sensing provides a timely, economic and systematic way (Sandra Brown, 2002;

Mickler et al., 2002) which can reduce uncertainties in a global forest biomass estimation (Dong et al., 2003; ISPRS, 1999).

Regarding remote sensing-based biomass estimation, passive optical remote sensing and Radar (Radio Detecting and Ranging) and the newly emerged Lidar (Light Detecting and Ranging) technique are assessed in the following two sections.

1.1.3.1. Optical remote sensing and Radar remote sensing

Two kinds of remote sensing techniques are widely used in forest biomass estimation before the emergence of lidar (Light Detecting and Ranging): passive optical remote sensing and radar (Radio Detecting and Ranging). Estimation of aboveground biomass using remote sensing are usually achieved by establish regression equations between field plot biomass and the remote sensing data (Austin et al., 2003; DeGier, 2003; Drake et al., 2002). However, these two types of remote sensing instruments suffer from the same problem: they are sensitive to altering in biomass in relatively young and/or homogeneous forests, but don't work well in older or heterogeneous forests where signal becomes less predictable with respect to changes in biomass (Drake et al., 2002).

The passive optical remote sensing relies on the solar illumination reflected from the outer canopy in the visible, near and middle infrared portion of the electromagnetic spectrum (~0.4 to 2.5 μm), thus only provide a two-dimensional view of the forest canopy. Spectral data will be related to biophysical attribute via vegetation indices such as NDVI, LAI, SAVI etc (Lefsky et al., 2001; Patenaude et al., 2005). However, various studies have shown that the accuracy in deriving volume and biomass from passive optical data has been inconsistent and varied much from case to case (Patenaude et al., 2005), the cause of which was found to be the saturation of vegetation index, and cloud cover. In dense forest conditions, biophysical parameters, such as NDVI, will no longer increase linearly with of biomass. Meanwhile, external factors such as cloud cover also impose a strong influence on the spectral responses of forest canopies (Treitz and Howarth, 1999).

Radar is a kind of active remote sensing system and is operated in the microwave portion of the electromagnetic spectrum (approximately 1cm to 100cm). It fires a pulse of microwave radiation to the ground and then detect and record the wave properties of the returned signal (the backscatter amplitude). Its advantage over optical remote sensing is its all-weather capability and not being impeded by cloud cover. SAR (Synthetic Aperture Radar) backscatter has been used in forest biomass estimation because longer radar wavelengths can penetrate the vegetation and interact with vegetation structure (Balzter et al., 2003; Rauste, 2005). The longer the wavelength, the stronger the ability of penetration into the canopy, and the more sensitive to volume and biomass (Rauste, 2005). The ability of SAR in forest biomass estimation has been investigated in various forest types, including pine forests (Castel et al., 2002; Rauste, 2005), coniferous forests (Dobson et al., 1992) and mixed deciduous and coniferous forests (Ranson and Sun, 1994) and better results were reported than passive optical remote sensing. However, the SAR signal also tends to saturate in dense forest conditions and has been shown to be insensitive to forest aboveground biomass above 150 Mg/ha (Austin et al., 2003; Ranson and Sun, 1994).

1.1.3.2. Lidar remote sensing

Laser altimetry, or Lidar (Light Detecting and Ranging) is an active remote sensing technique using a laser beam as the sensing carrier (Wehr and Lohr, 1999). Lidar systems measure the time elapsed from a pulse of the laser energy (usually in the near-infrared, for vegetation studies) generated from the sensor and reflected back from the target, and can finally provide the elevation information of targets with respect to the speed of light (Hudak et al., 2002).

According to the width of the laser beam, Lidar instruments can be divided into two categories: large-footprint or small-footprint systems. Large-footprint systems have a laser beam that is greater than 5m in diameter on the ground, whereas small-footprint have a more narrowly focused beam that is typically less than 50cm in diameter (Bortolot and Wynne, 2005). Furthermore, large-footprint lidar instruments fully digitize the returning signal which is referred to as lidar waveform, thus providing enhanced vertical architecture information of forests (Patenaude et al., 2005), while small-footprint systems typically record one to five returns (e.g. the first and last significant returned signal) which are referred to as discrete return lidar (Evans et al., 2006; Hudak et al., 2002) as shown in Figure 1-2. In respect to forest biomass estimation, large-footprint Lidar hold more advantages than small-footprint Lidar (Drake et al., 2002; Dubayah and Drake, 2000) and many studies conducted to estimate forest biomass mainly used large-footprint Lidar (Drake et al., 2002; Lefsky et al., 2001; Means et al., 1999), for Lidar waveforms offer the potential to directly derive vegetation height, and other canopy structure characteristics, which ensure the possibility of a high accuracy of forest biomass estimation.

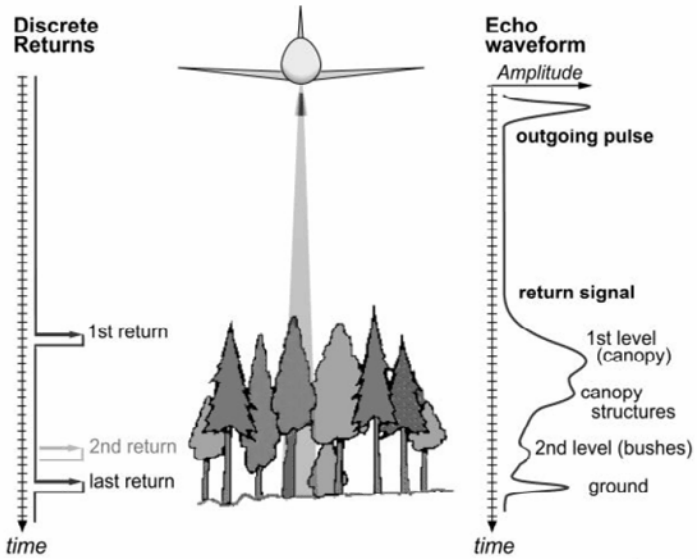


Figure 1-2 Discrete return lidar and waveform lidar in canopy height measurement
(Source: http://www.riegl.com/airborne_scannerss/airborne_scanners_literature_/airborne_literature_menu_.htm)

The Lidar system also holds promise to overcome the saturation problem in forest biomass estimation as found in both passive optical and radar remote sensing and has a great potential to increase the accuracy of forest biomass estimation (Drake et al., 2002). The ability of estimating aboveground biomass has been tested in dense forest conditions using lidar where passive optical and active radar sensor typically fail to do so (Drake et al., 2002; Patenaude et al., 2005). For example, depending on the respective study area, total aboveground biomass measured from field data could be predicted from airborne-derived lidar metrics with R2 value up to 0.95 for biomass levels of 1300 Mg/ha for coniferous forests in the western cascades of Oregon, USA (Means et al., 1999). Lefsky et al. (2001) used lidar-derived metrics to predict field measured aboveground biomass at sites in the temperate

deciduous, temperate coniferous, and boreal coniferous biomes in United States, for maximum aboveground biomass up to 716.3 Mg/ha, 1329.0 Mg/ha and 58.5 Mg/ha with R^2 value up to .65, .87 and .76, respectively.

Estimation of aboveground forest biomass using large-footprint lidar can be achieved through the following phases, as is shown in the conceptual model of forest AGBM estimation in Figure 1-3:

- The first phase is to establish tree allometric equations based on the field measurement for a restricted number of felled trees, which are also representative for the study area. For each tree, aboveground forest biomass and relative variables, such as tree DBH (average diameter at breast height) and height, are measured. Regression analysis is used to establish the relationships between tree biomass and tree variables, so called allometric equations (DeGier, 2000).

- The second phase is to locate the field plots that correspond to lidar foot-prints and calculate field plot forest biomass based on the allometric equation established in the first phase, by summing up individual tree biomass.

- The third phase is to derive metrics, such as canopy height, HOME, from lidar waveform, then regression equations are built between the field plot biomass and lidar-derived metrics. It is suggested horizontal canopy structure (***forest types***) information should also be taken into consideration when developing regression equations in this phase.

- The forth phase is to apply the regression equations (for ***different forest types***) to all lidar footprint plots (with corresponding forest types) to estimate the aboveground forest biomass.

As can be seen from the phases, the forest type information is suggested to be taken into consideration at the third and forth phases in forest AGBM estimation process. Though Lidar systems have the potential to retrieve vertical canopy structure directly, however, for spatially consistent biomass assessment, horizontal canopy structure (forest types) information is also essential (Lefsky et al., 2005; Patenaude et al., 2004; Patenaude et al., 2005) for the studies that are not conducted in homogeneous forest areas or forests of consistent horizontal spatial pattern, due to the facts that forest biomass may vary considerably between different forest types, even when of similar height, density and age.

The relationships between lidar-derived metrics and canopy structures, such as canopy height, canopy cover, canopy height profile, quadratic mean stem diameter (QMSD), biomass and basal area have been investigated (Drake et al., 2002; Drake et al., 2002; Dubayah and Drake, 2000; Lefsky et al., 1999; Lefsky et al., 1999; Lefsky et al., 2001; Means et al., 2000; Means et al., 1999). Ronson et al. (2004) indicated that different forest types are expected to have different waveform profiles.

However, there have not been any studies attempting to develop an effective way to derive accurately forest horizontal characteristics using large foot-print lidar. The only research that tried to establish a relationship between the lidar waveform and different forest types was by Ranson et. al. (2004), who investigated the possibility of using metrics front slope angle and centroid, both derived from ICESat-GLAS lidar waveform on terrain with less than 5 degrees slope, as indicators of forest types. These two metrics, however, failed to explain the relationship. Another study (Duong et al., 2006b), doing landcover classification using waveform lidar on flat areas, only tried to classify vegetation as a whole class, while no sub-classification of different forest types was attempted.

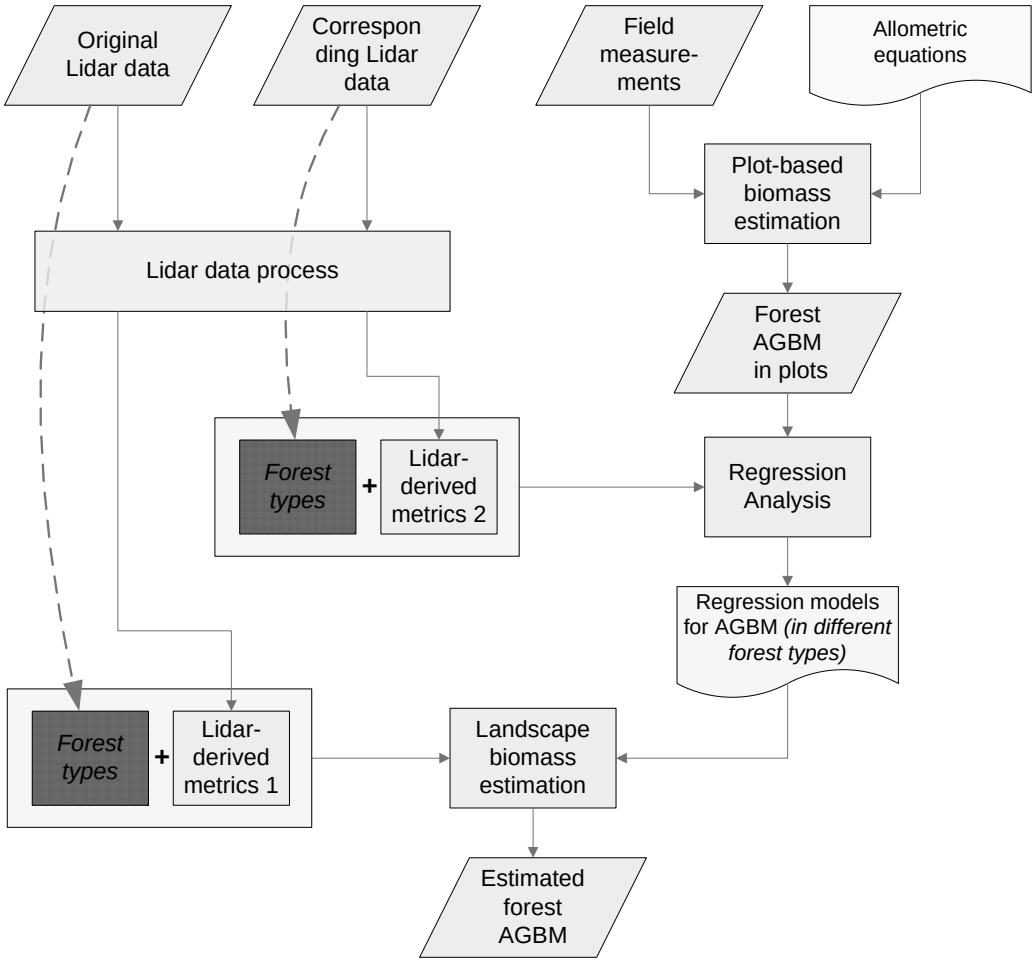


Figure 1-3 Conceptual model of estimating AGBM forest biomass using large-footprint lidar

The lidar instrument used in this study is spaceborne ICESat-GLAS lidar system. Its footprint diameter on the Earth's surface is nominally 70m, spacing between footprints is about 175m and spacing between tracks is about 2km in Northeastern China (NASA, 2007). One of its designed science objective of ICESat mission is the measurement of canopy height. This research aims to assess the ability of using this large-footprint waveform lidar to detect forest types and derive useful metrics from lidar waveform that can be used to retrieve forest horizontal characteristics.

1.2. Research Problem

The advantage of large-footprint lidar in forest application is that its waveform provide an apparent vertical canopy structure of the forest within the footprint. The conceptual basis of how large-footprint lidar waveforms reflect the vertical canopy structure (or referred as canopy height profile in some literatures) is illustrated below in Figure 1-4. The return waveform gives the vertical distribution of surfaces intercepted by the incident beam. At any particular height, the amplitude of the waveform records the strength of the return energy. Thus, for canopy surfaces with similar reflectance and under similar atmospheric conditions, a larger amplitude indicates more canopy material (leaves and branches). This vertical canopy structure information provided by the waveform profiles of large-footprint lidar give the possibility to distinguish different forest types, for different forest types are expected to have their typical vertical canopy structures, thus producing different typical lidar waveforms.

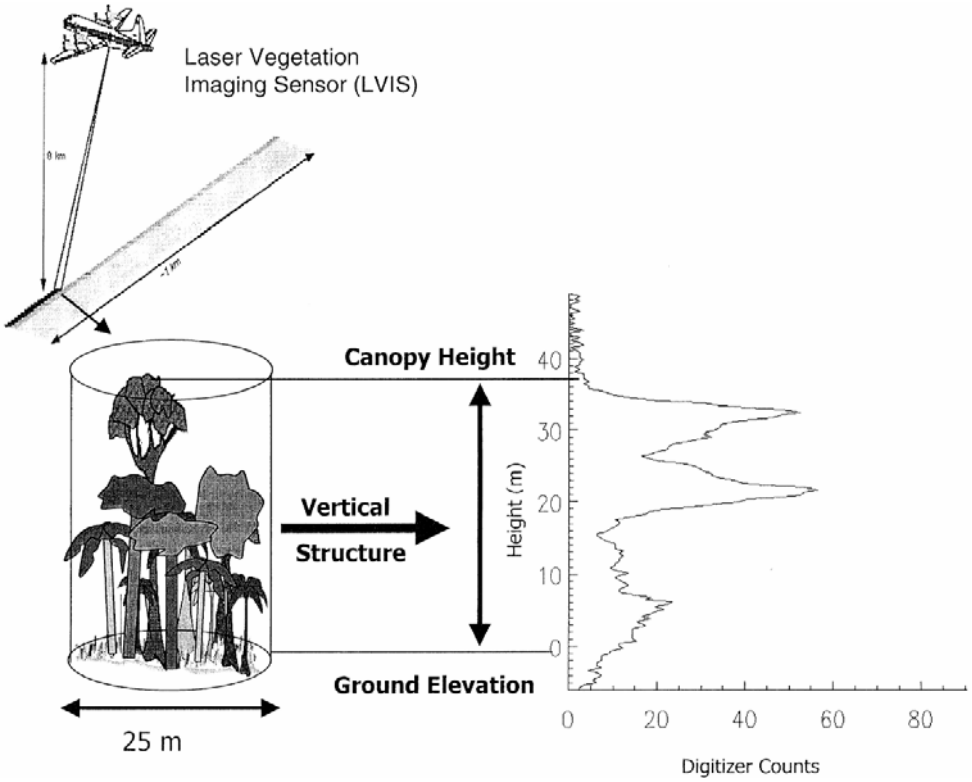


Figure 1-4 Conceptual basis of how large-footprint lidar waveforms reflect the canopy vertical structure (height profile) (Drake et al., 2002)

Five existing lidar waveform-derived metrics: Lidar canopy height (LHT), the height of median energy (HOME), the height/median ratio (HTRT), ground return ratio (GRND) and front slope angle are commonly used to describe vertical canopy structure characteristics (Brenner et al., 2003; Drake et al., 2002; Ranson et al., 2004). Figure 1-5 shows a typical Lidar waveform profile and waveform-derived metrics:

- *Lidar canopy height (LHT)* is calculated by identifying the signal start and the center of the last Gaussian pulse and then calculating the distance between the two.

- *Height of Median Energy (HOME)* is the height at which half of the return energy is above and half below. It is not only affected by the tree height, the vertical canopy structure distribution, but also sensitive to canopy openness (tree density).

- *HOME/LHT ratio (HTRT)* is the HOME divided by canopy height. Compared with HOME, HTRT normalizes tree height. However, it is still affected by vertical canopy structure and canopy openness these two factors, thus not an ideal metrics for distinguishing forest types.

- *Ground return ratio (GRND)* is the ratio of ground return energy and the total energy contained in the waveform. GRND only provide an approximation of the degree of canopy closure with certain canopy assumption and it is unsuitable to be used to derive forest type information.

- *The front angle* is the angle from the vertical to the vector from the signal start to the peak of canopy return energy. This metrics has a reflection on both forest density and the vertical distribution of the top canopy structure, thus it is unable to derive forest type information from this single metrics.

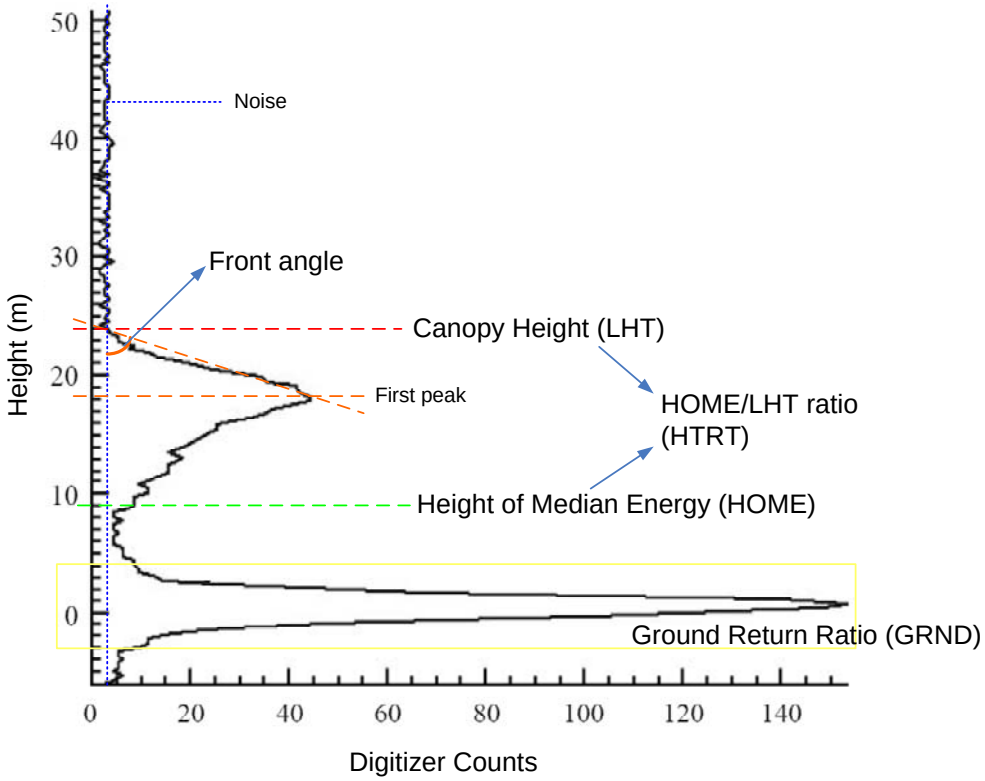


Figure 1-5 Typical Lidar waveform profile (Drake et al., 2002; Ranson et al., 2004)

However, factors such as, forest height, forest density, subcanopy, and topographic slope, as well as forest types are all known to affect the waveform profiles which make the waveform profiles differ a lot even of the same forest type (Brenner et al., 2003; Dubayah and Drake, 2000; Michael Andrew Lefsky, 1997). Figure 1-6 illustrates the theoretical waveform for uniform canopy characteristics and how these factor influence the waveform characteristics, thus the lidar-derived metrics.

- ***Forest height*** can be immediately reflected by *Lidar canopy height (LHT)* on the lidar waveform. As forest height increases or decreases, the centroid of canopy return energy will shift upwards or downwards, so then the *Height of Median Energy (HOME)*. *Height/LHT ratio (HTRT)*, determined by *Height of Median Energy (HOME)* and *Lidar canopy height (LHT)*, will change accordingly.

- ***Forest density*** has an influence on the *Height of Median Energy (HOME)* and *The front angle*, for an increase in forest density can lead to more reflection by canopy and correspondingly less by the ground, thus an increase of *HOME* and *HTRT*. And *the front angle* will also increase for there will be more top canopy return energy.

- ***Subcanopy*** can result in more than one peak on the waveform profile and make it more complex. It can also influence the *Height of Median Energy (HOME)* and *Height/LHT ratio (HTRT)* by increasing the canopy return.

- ***Slope*** can cause not only the ground return to spread, but also the canopy return (Brenner et al., 2003). Canopy return and ground return can even become severely mixed on steep slopes (Figure 1-7). It has a big influence on the whole five metrics. However, it is impossible to ascertain the

extent of the influence most of the times, for there is no effective way to separate canopy return and ground return on steep slopes (Brenner et al., 2003).

- And for different ***tree species***, their different spatial distribution of canopy structure (e.g. leaves and branches) can lead to different waveform profiles.

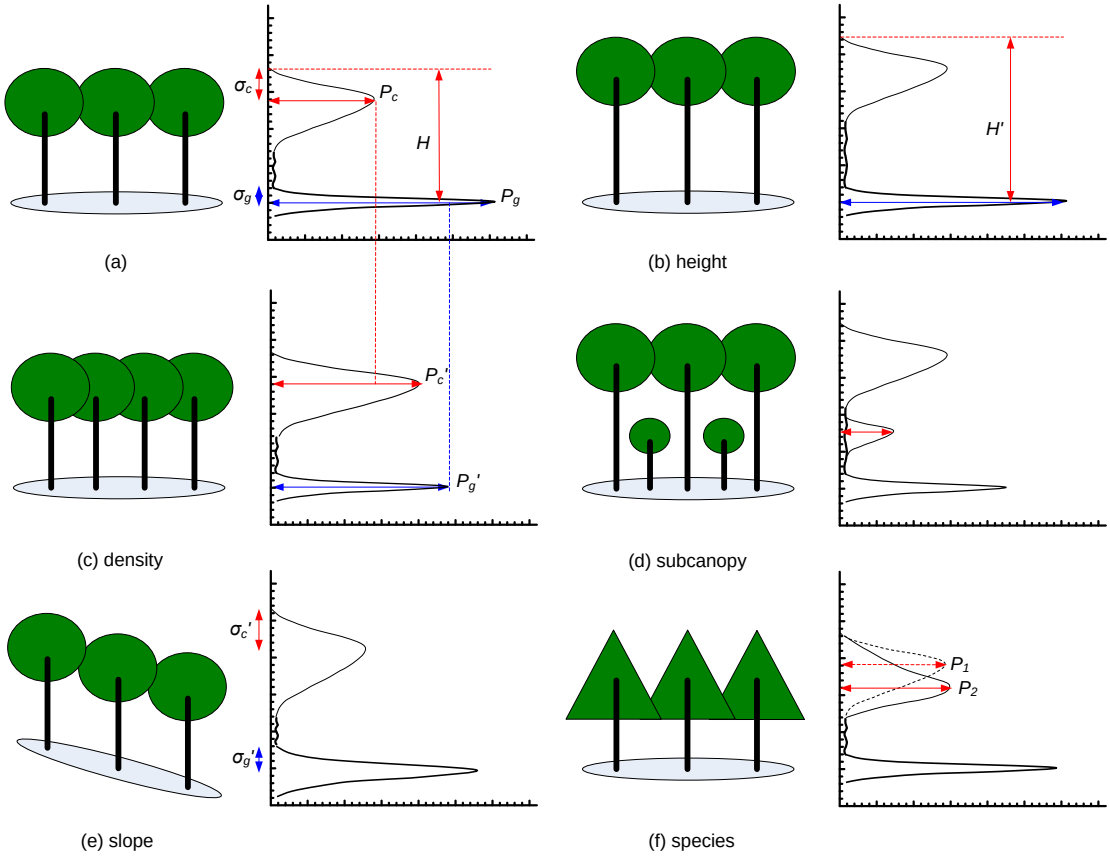


Figure 1-6 Theoretical waveforms for uniform canopy characteristics and factors influence the waveform characteristics.

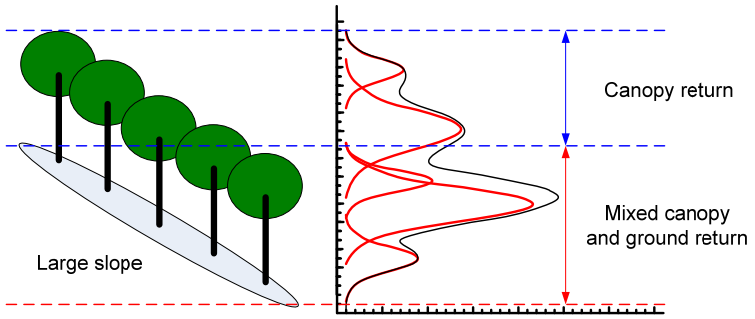


Figure 1-7 Slope can result in mixed canopy and ground return in the waveform

Through this analysis, we conclude that the existing metrics are all incapable in respect to detecting different forest types, for they can vary a lot of the same forest type with the influence of factors mentioned above. In order to derive forest type information from large foot-print lidar waveforms, metrics that can reflect the different vertical canopy structures between different forest types without biases of the factors, such as forest height, density and slope, have to be found.

To achieve this goal, we will focus more on the canopy return part, as canopy return energy reflect the canopy structure distribution, thus forest type information, for different forest types are expected to have the different vertical distribution of canopy structures (Lamotte et al., 1998), thus different canopy return energy profiles.

Therefore, several quantile-based metrics , R25, R50 and R75, as explained below, will be tried to derive from the lidar waveforms as shown in Figure 1-8. Blair et al. (1999) and Kimes et al. (2006) used H25, H50 and H75 as a relatively direct measure of the vertical profile of canopy components. H25 is the 25% quartile height calculated by subtracting the elevation at which 25% of the return energy occurs from the ground elevation (Kimes et al., 2006). In this study, CH25, CH50, CH75 are proposed and are heights of 25%, 50% and 75% canopy return energy, respectively. This group of quartile heights is an improved version of H25, H50 and H75, which exclude the influence of ground return and serve as a more direct measure of the vertical profile of canopy components. Then the ratios R25, R50 and R75 will be calculated by dividing CH25, CH50 and CH75 by canopy height, to normalize the influence of different canopy heights and the waveform spreading effect caused by topographical slope. These metrics will then be applied to statistical methods to find their relationships between different forest types.

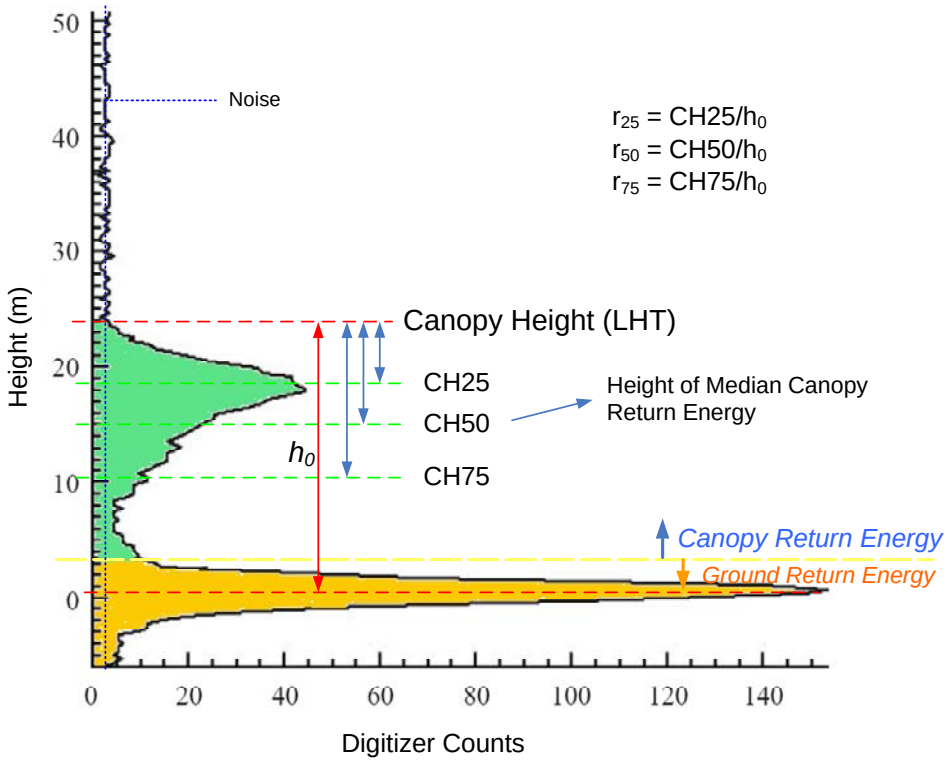


Figure 1-8 Metrics R25, R50 and R75 proposed to be derived from lidar waveform, in which ground return and canopy return can clearly separated

What's more, the return waveform can be regarded mathematically as a composition of several Gaussian curves (Figure 1-9), and reflect Gaussian surfaces (Brenner et al., 2003; Wagner et al., 2006), i.e., the layers of branches and foliage in the forest. The different vertical stratum characteristics of different tree species, especially broadleaved trees and needleleaved trees, are supposed to be reflected by the characteristics of the decomposed Gaussian curves, i.e., the Gaussian curve slope (Figure 1-10).

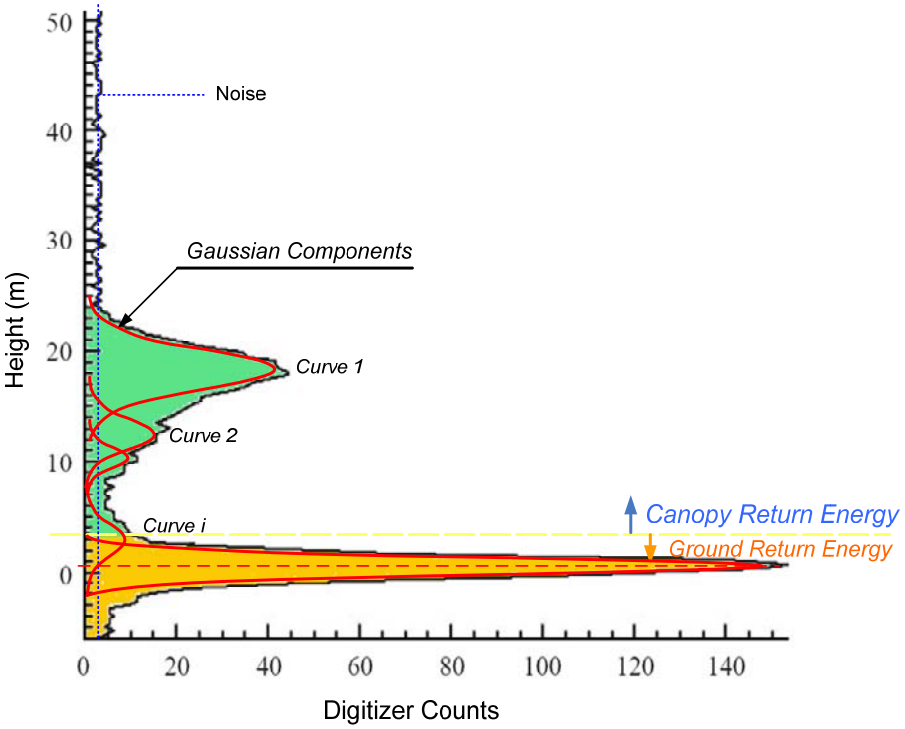


Figure 1-9 Decomposed Gaussian Curves of lidar waveforms

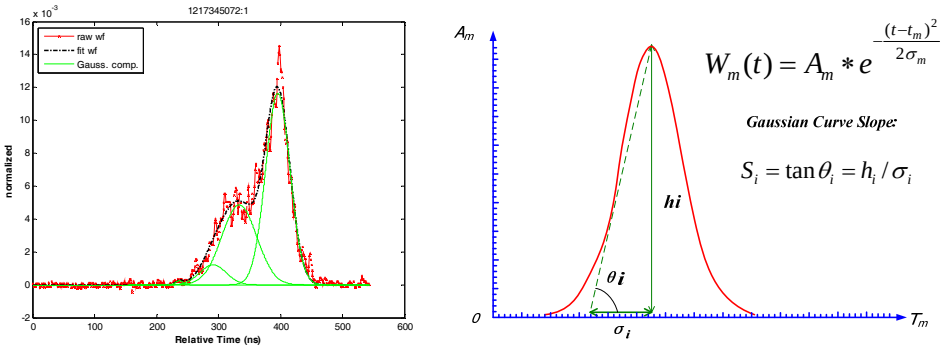


Figure 1-10 ICESat-GLAS waveform decomposition and Gaussian Curve Slope

For example, the vertical strata of needleleaved trees such as Pine (*Pinus koraiensis Sieb. et Zucc.*) is more obvious than broadleaved trees like Betula (*Betula platyphylla Suk.*). So when the lidar energy is reflected by this thin slice of layer of needleleaved trees, the wave range will be much smaller than that by the broadleaved trees, as shown in Figure 1-11. The averages of the decomposed Gaussian curve slopes (AGS) of the needleleaved forests are expected to be larger than those of the broadleaved forests. Moreover, considering the homogenous nature of vertical structure of the needleleaved forest, the standard deviations of the decomposed Gaussian curve slopes (SGS) of needleleaved forest are expected to be smaller than those of the broadleaved forest.

So in the research, Gaussian fitting will be firstly applied to the waveforms and the two kinds of statistical characteristics of decomposed Gaussian curve slopes, namely AGS and SGS as shown in Figure 1-12, will be investigated to examine whether they can serve as indicators of different forest types.

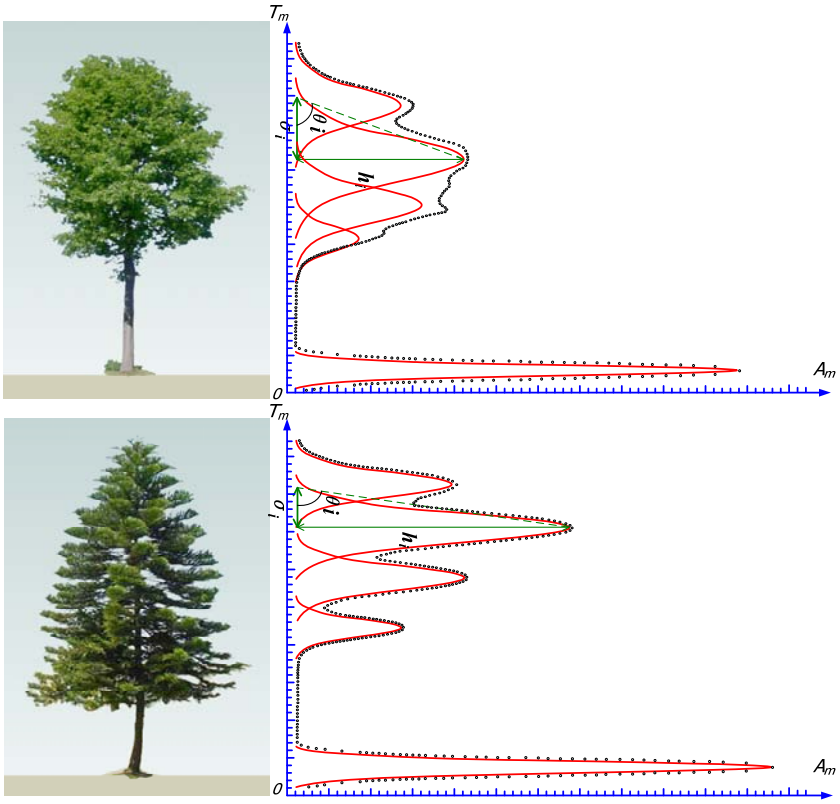


Figure 1-11 Theoretical comparison of vertical strata and waveforms of broad- and needleleaved forest

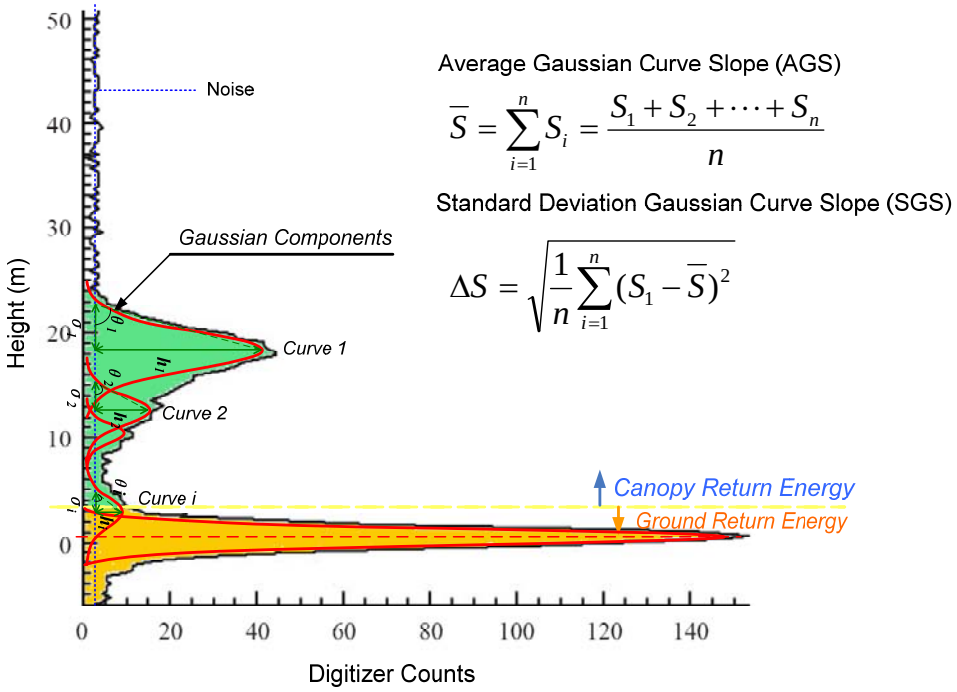


Figure 1-12 Metrics AGS and SGS proposed to be derived from lidar waveform, in which ground return and canopy return can clearly separated

However, the topographical slope imposes such a great impact on waveform profiles that slopes can cause the lidar ground return to spread, leading to inaccurate ground determination, and consequently, canopy heights; what's still worse, on really steep slopes, return from trees can be convolved or blurred with the surrounding topography (Hyde et al., 2005) (Figure 1-7).

Our study area is a montane area with a wide range of slopes, lying in Mountain Changbai forests, Northeastern China. This cool temperate forest has variety of forest types, including broadleaved forest, needle-forest forest and mixed forest (Xiao et al., 2002). Due to the complex composition of tree species in the forest, in our study, when the percentage of needleleaved trees in number in the forest plot surveyed exceeds 70%, the plot will be classified as Needleleaved forest. If the percentage of needleleaved trees is under 15%, the plot will be labelled as Broadleaved forest. And plots with the percentage of needleleaved trees in number between 15-70% will be classified as Mixed forest.

What's more, considering the effects of slope on waveforms, our study firstly aims to derive forest type information from waveforms located on terrain with low relief, in which there are obvious ground peaks, thus ground return and canopy return can be separated. Then investigate the possibility of deriving forest type information from waveforms of which footprints plot on steep slopes.

1.3. Objectives

The main objective of the research is to investigate the possibility of using large-footprint lidar to derive forest type information, thus extending the application of large-footprint lidar and contributing to the global methodology of a more accurate biomass estimation using large-footprint waveform lidar.

The specific objectives are as follows:

- To investigate whether broadleaved and needleleaved forest can be distinguished by the lidar.
- To investigate which combination of the lidar waveform metrics proposed can best distinguish broadleaved and needleleaved forest.
- To investigate the extend to which lidar can classify mixed broad-/needle-leaved forest.

1.4. Research Questions

- For what slope range, can the waveform metrics proposed be successfully derived?
- How strong is the relationship between the metrics of canopy energy distribution (R25, R50, R75), which reflect canopy vertical structure distribution, and broadleaved and needleleaved forest?
- How strong is the relationship between the metrics of statistical characteristics of decomposed Gaussian curve slopes (AGS, SGS), which reflect canopy vertical stratum characteristics, and broadleaved and needleleaved forest?
- Which combination of the derived metrics can best predict the presence of broadleaved and needleleaved forests? What is the accuracy of the classification?
- How strong is the relationship between the metrics and mixed broad-/needle-leaved forest?

1.5. Hypotheses

- Different forest types have typical different vertical canopy structure, and those differences can be reflected by ICESat-GLAS waveform profiles (energy distribution of the canopy return and pattern of decomposed Gaussian waveforms).

- The metrics proposed can be derived from ICESat-GLAS waveforms on low relief areas and can be used to derive forest types information.

- The slope exceeding a certain degree can make it impossible to distinguish canopy return from ground return within ICESat-GLAS waveforms, for return from trees will be convolved or blurred with the surrounding topography.

1.6. Research Approach

The general framework of the research methods is demonstrated in Figure 1-13. The methods used in the framework will be described specifically in the methods chapter.

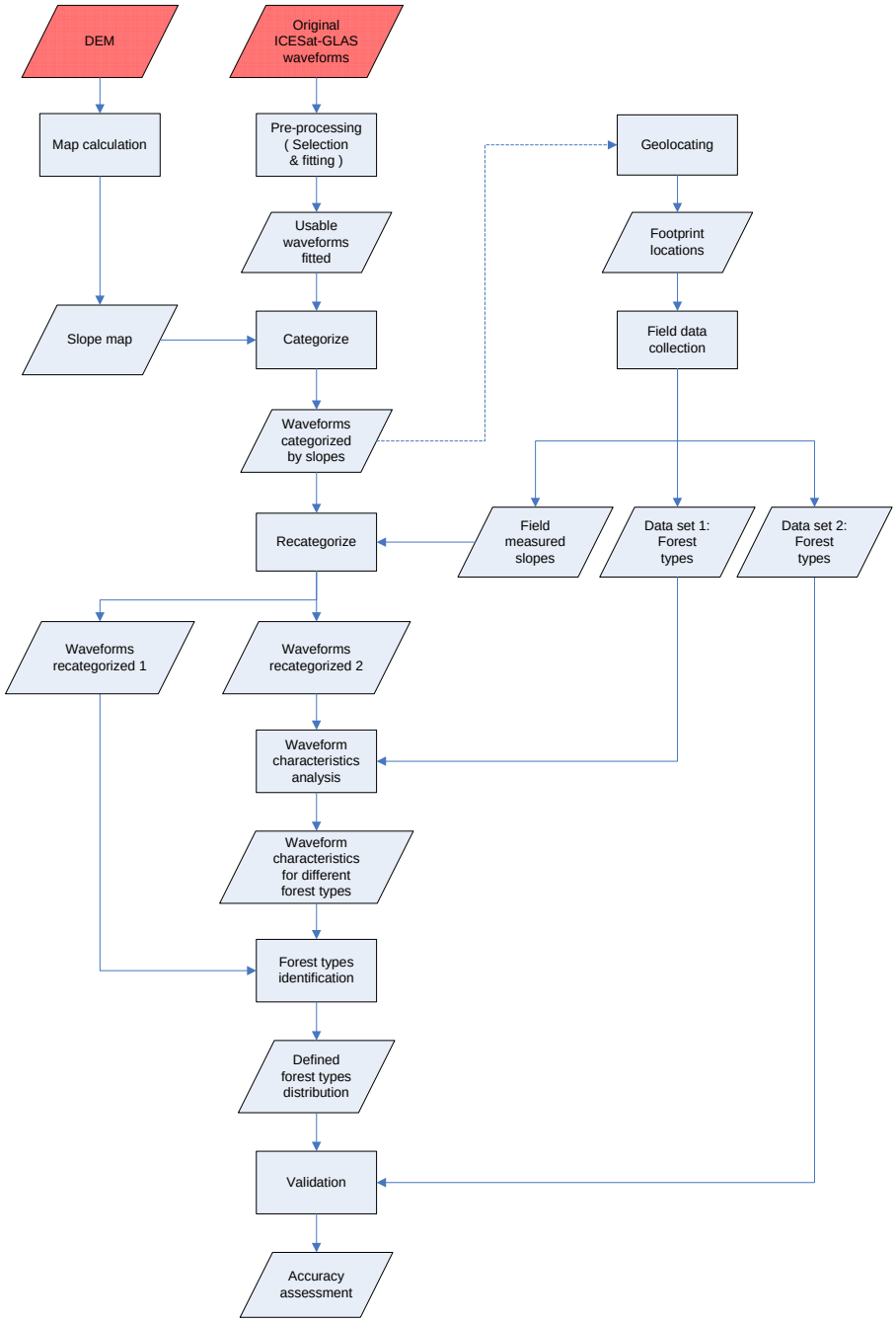


Figure 1-13 General flow chart of research approach

2. METHODS AND MATERIALS

2.1. Study Area

To answer the research questions and test the hypotheses of the research, the study area is selected in Wangqing forest area of Jilin province, China (Figure 2-1). It lies along the border between China and North Korea and belongs to the Changbai mountain system, which is one of the most valuable reserves because of its rich gene pool of many plant species with the altitudinal vegetation zone in the Mountain. Geographically, the study area is located 43°05′ N through 43°40′ N latitude, and 129°56′ E through 131°04′ E longitude, covering an area of approximately 85×60 km.

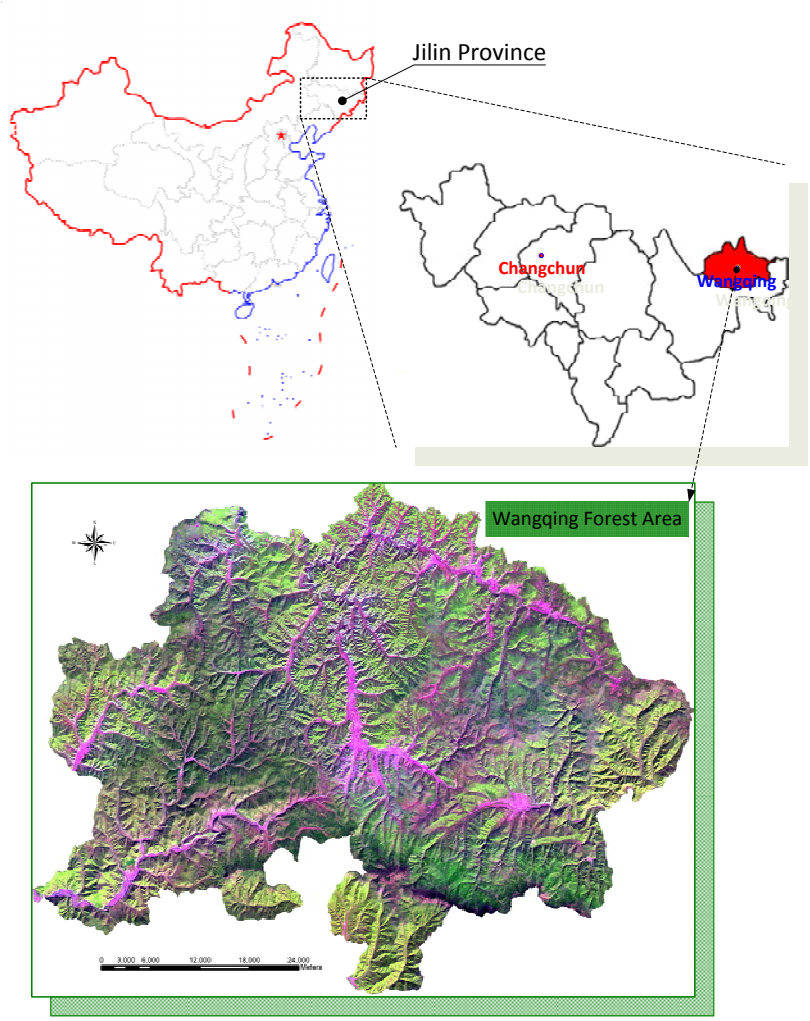


Figure 2-1 Location of Wangqing Forest Area of Jilin Province, China

Wangqing forest is a cool temperate forest and dominated by a cool temperate continental climate influenced by monsoon, and has four clearly different seasons: windy spring, hot and rainy summer,

cool autumn and cold winter. Mean annual temperature is 3.9 ^{o}C. The mean annual precipitation is 438 mm, about 80% of which takes place between May and September. The elevations of the study area vary, ranging from 360m to 1,477m above sea level, and the steepest slopes are more than 100 percent (Yanqiu, 2006).

The natural vegetation in the study area is dominant by Pinus koraiensis, Tilia amurensis, Quercus mongolica, Fraxinus mandshurica and Acer mono, and the mean forest canopy height is 26 m (Yanqiu, 2006). According to the forest structure and the growth rules and biology features, the forests in the study area can be classified to 3 types: Broadleaved forest, Needleleaved forest and Mixed forest, as defined in Section 1.4.. The dominant tree species of boardleaved forest are *Betula platyphylla, Quercus mongolia, Betula castata, Populus ussuriensis, Fraxinus mandshurica.*, while *Pinus korinensis, Picea koyamai var. koraiensis, Larix olgensis* and *Abies nephrolepis are* dominant in the needleleaved forest. The main tree species of mixed forest are *Pinus korinensis, Picea koyamai var. koraiensis, Pinus sylvestris var. mongolica*, *Larix olgensis, Abies nephrolepis, Tilia amurensis, Ulmus pumila, Betula platyphylla, Betula castata* and *Acer mono*. There are usually shrubs growing in the low-layer of the forest.

2.2. Approach in Forest Type Information Derivation

The overall framework of research method is shown in the Figure 2-2 below:

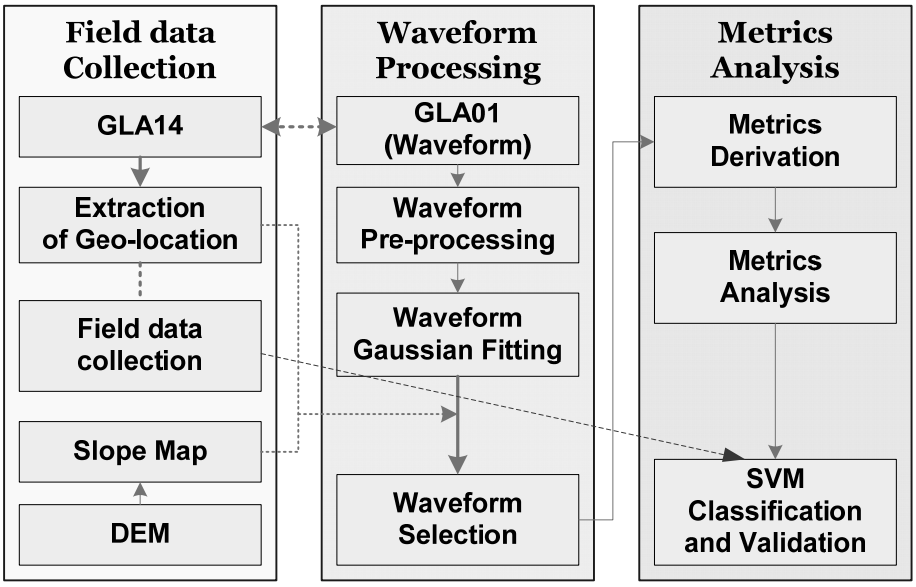


Figure 2-2 Overall framework of the study

2.3. Data Available

- Satellite Images

Two cloud-free Landsat TM satellite images were available (October 2, 2001 and September 29, 2006), both of which have a resolution of 30×30m including 7 bands of information. Geo-reference of the TM image was applied to make the imageries match with the real-world coordinates. The accuracy of the geometric correction was checked with the remarkable surface features in the topographic maps.

The reason why these two images were chosen was because the dates on which they were acquired was in the same period of year when the fieldwork will be conducted. The main usage of the two images in the study is for visual interpretation in the fieldwork stage for sample footprints selection.

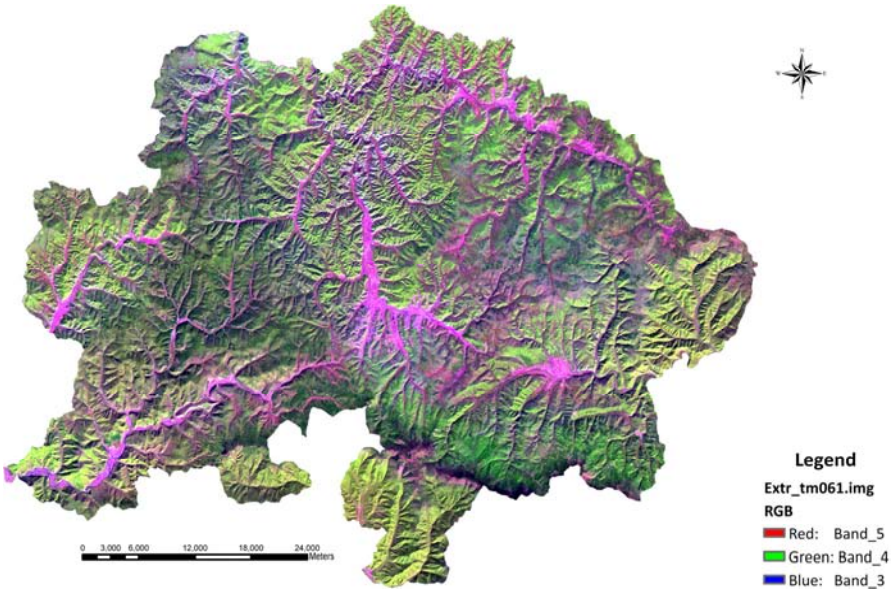


Figure 2-3 False color composite (Band 5, Band 4, Band 3) of ETM+ image (29 September 2006) of the study area

♦ Digital Elevation Model

A Digital Elevation Model (DEM) of the study area produced by Chinese Academy of Forestry was also introduced in the study. The resolution of the DEM is 30×30 meters horizontal and vertical 20m. The DEM was applied in this study to derive slope map (Figure 2-4), which will assist in ICESat-GLAS lidar waveform selection. The waveform will be firstly categorized by slopes, to determine the slope range for which the proposed metrics can be successfully derived from. Also, the slope map will be of help to determine the accessibility of footprints and design sample strategy in the fieldwork.

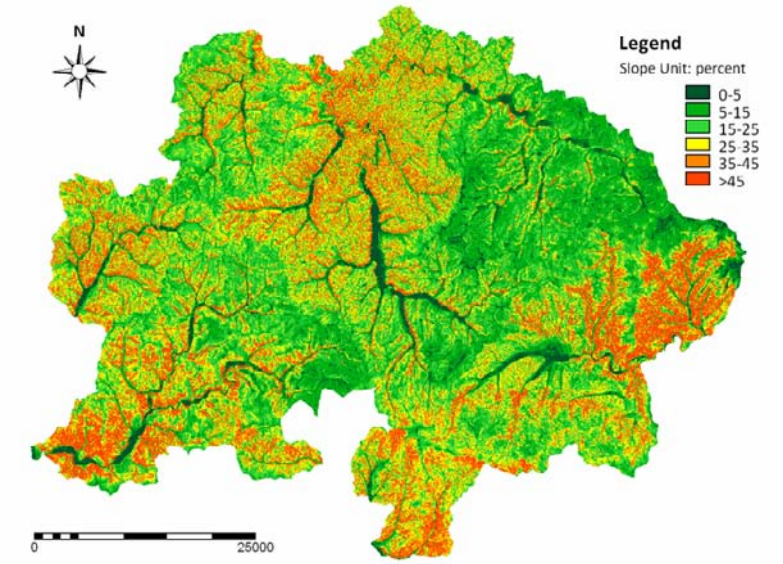


Figure 2-4 DEM derived slope map of Wangqing forest area

♦ Thematic Maps

An administration distribution map, road map and river map of the study area (Figure 2-5), were also acquired from Wangqing forest bureau, to assist in footprints selection and field data collection.

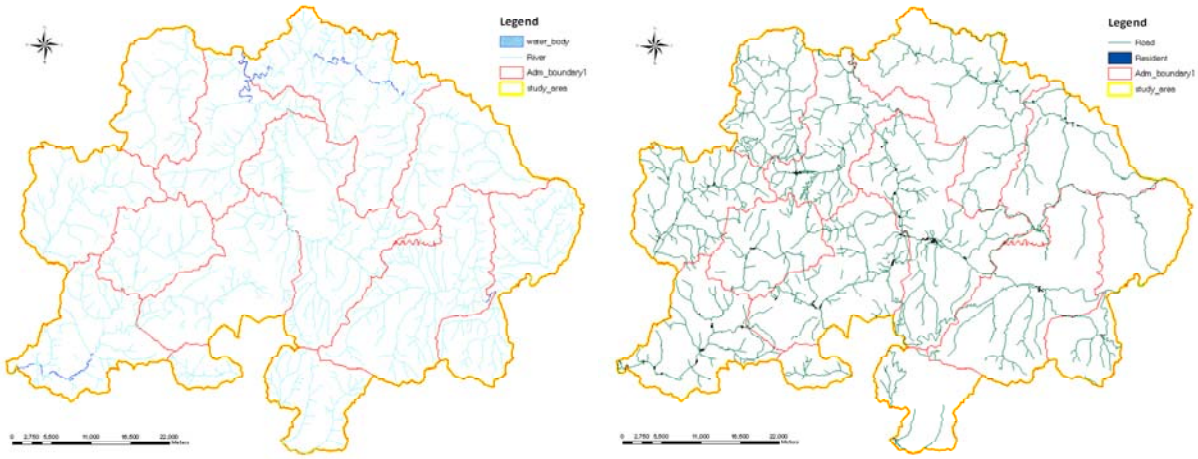


Figure 2-5 Thematic map of the study area: river map (left) and road map (right)

♦ Secondary Dataset

Around one hundred (102) forest type information plots with GPS location collected by Xing Yanqiu (2006), using the same sample strategy as illustrated in Section 2.4.2., were obtained as the secondary data. The dataset contains plot information such as forest type (broadleaved forest, needleleaved forest and mixed forest, as defined in Section 1.4.), species of each tree, related forest vertical structure data, and slope value, of forest plots locating within the sampled 2003-2005 ICESat footprints of the study area. The corresponding ICESat-GLAS lidar waveform dataset is also available.

2.4. Data Collection

With the intention of exploring and analyzing the relationship between lidar waveform profile and forest type information, the primary data shown below are collected.

2.4.1. ICESat-GLAS Full Waveform Data Acquisition

A full waveform dataset of Geoscience Laser Altimeter System (GLAS) onbroad the Ice, Cloud, and land Elevation Satellite (ICESat) with cloud-free profile during the forest growth season of Oct. 27, 2006 over the study area was acquired. There are 15 products of ICESat data, however, only the products GLA01 and GLA14 are used for the research. The GLA01 Global Altimetry Data Product includes the waveforms and timing. The GLA14 Global Land Surface Altimetry Data Product contains precise geo-location of footprint centers with height information. A GLA01 waveform is linked to a GLA14 location by index and shot number.

The footprint of the waveforms are elliptical with the size of 95×52m on average (Harding and Carabajal, 2005). The separation of footprints along track is 175m with the horizontal geo-location accuracy of 3.7m. The waveform is digitized in 544 nanoseconds over the land area, which means the

waveform can acquire a height of up to 81.6m over land (NASA, 2007; NSIDC, 2007a). A visualization of the footprints over the study area is shown in Figure 2-6.

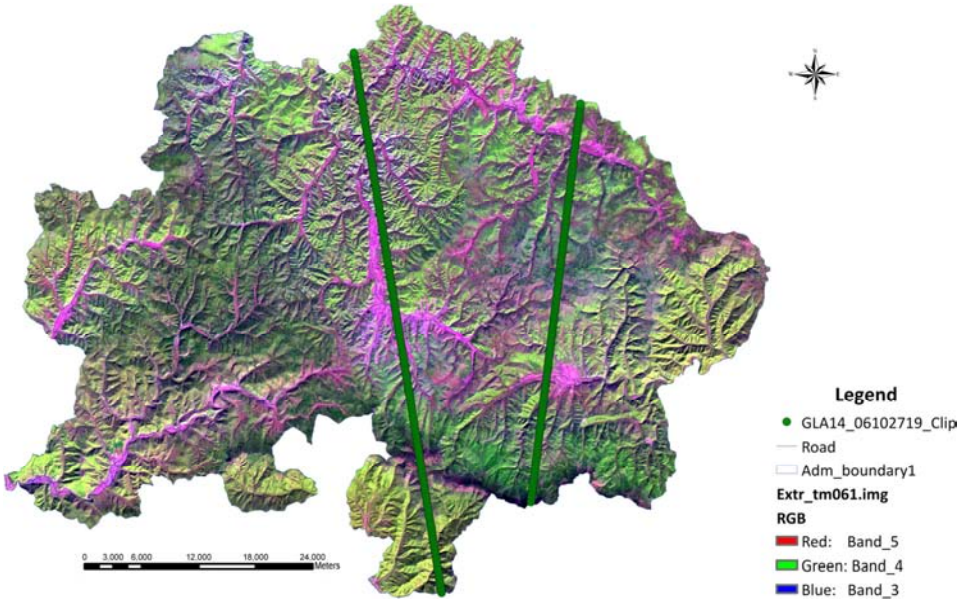


Figure 2-6 ICESat ground tracks (27 October 2006) over the study area

2.4.2. Field Data Collection

Fieldwork was implemented in the Wangqing forest area from September 18th to October 1st, 2007. Although it is always suggested to use random sampling to avoid bias in a study, due to the complex and rough terrain in the mountainous forest area, what's more important, plot data have to be collected within the footprints, which in most cases are locating in the remote forest area far away from roads, making it difficult and time consuming to access and collect plot data, a modified stratified random sampling method by segment slope was applied in the fieldwork.

A total of 103 circular plots centered on ICESat-GLAS lidar footprints distributed on different slope range were selected and measured in the footprints throughout the study area (Figure 2-7). To correspond to the lidar measured metrics within footprints and avoid the biases of GPS locational error, and in respect of the data collecting load of each field plot, the size of the circular plot was 25.24 m in diameter on flat land, which is smaller than the 70 m-diameter footprint (Figure 2-8). Considering the effect of slope, the radius of the plots were corrected according to slope correction table (in Appendix I Table 12) on slope areas.

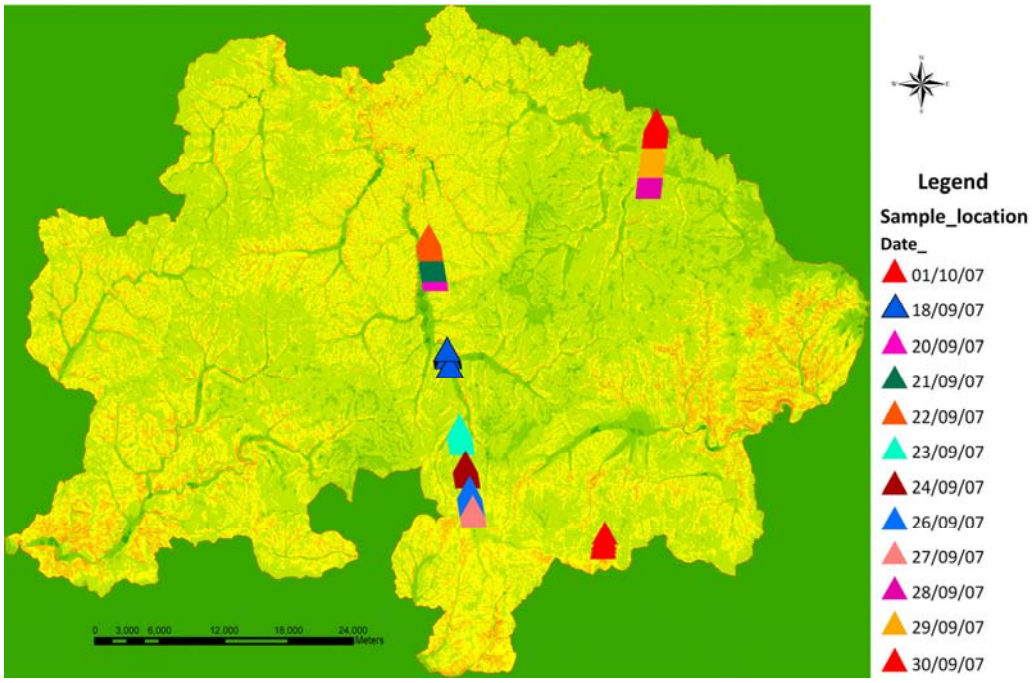


Figure 2-7 Distribution map of sample plots

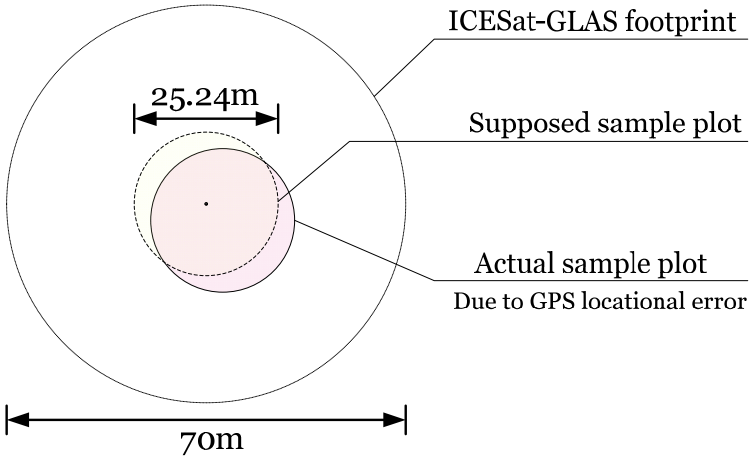


Figure 2-8 Sample plot design on flat land

The study focused on the relationship of the forest type and lidar waveform profile, so forest type of each field plot was firstly recorded. Meanwhile, the tree species and height of each tree located within the sample plot with DBH>10cm were also recorded, since they normally contribute more to characteristics of the waveform profiles. The vertical canopy structure information by visual interpretation of the sample plots was also collected, namely the height and coverage of three forest layers (high-/medium-/low- canopy layers) as it is shown in Appendix I Figure 0-1. What's more, slopes of the sample plots were measured, to compare with the DEM derived slope and categorize the ICESat-GLAS lidar footprints in the data processing stage. Photos were also taken of each sample plots. The detailed data collected in the field can be found in Appendix I Table 13.

2.5. Data Processing

2.5.1. Waveform Pre-processing

Before we could make use and do the analysis of the ICESat-GLAS lidar waveform data, processing of the raw dataset acquired during the data collecting stage, was necessary. The method employed to process the ICESat-GLAS lidar waveform data is adopt from Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document Version 4.1 (Brenner et al., 2003) and the methodology developed by Duong et al. (2006a).

2.5.1.1. Waveform Conversion

As ICESat full waveform data is distributed in binary format, the GLA01 and GLA14 is firstly converted into ASCII format by an IDL program *IDL_Reader* developed by National Snow and Ice Data Center (NSIDC, 2007b). The produced waveform data originally in counts (from 0 to 255) is then be converted into voltage units for further process.

2.5.1.2. Waveform Normalization

After conversion, the voltage waveform is then normalized, by dividing each amplitude by the total received energy. The purpose of the operation is to make the waveforms comparable, since different waveforms were captured in different epochs. Due to reasons such as different atmospheric condition or changes in the behaviour of the laser device, the amount of return energy may vary with time, even the ground don't change (Duong et al., 2006a). This operation makes it possible to compare the relative energy levels of particular constituents of different waveforms.

The normalization process is completed by dividing the received energy V_i at moment i by the total energy V_T, calculated by $V_T = \sum_{i=1}^{544} V_i$. Then the normalization is described as $V_N(i) = V_i / V_T$. Two normalized full waveform are shown in Figure 2-9.

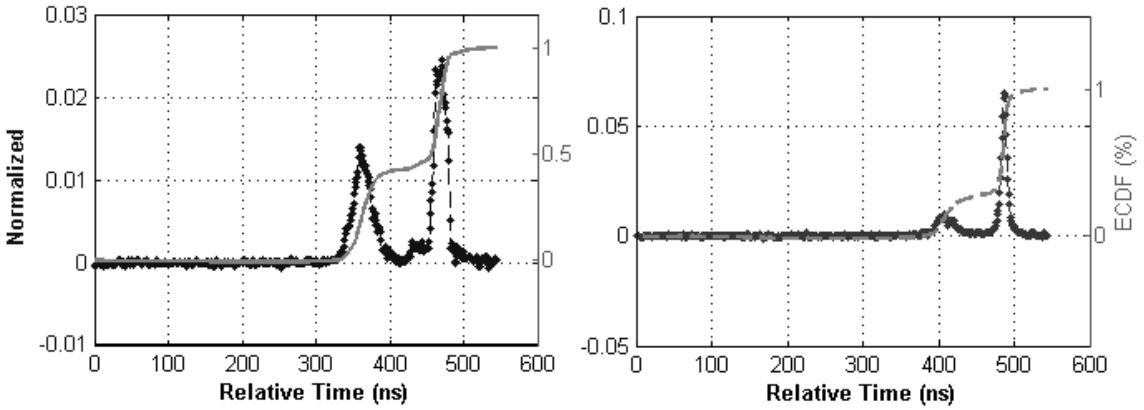


Figure 2-9 Two normalized waveforms, displayed together with their cumulative distribution curves (in gray) (Duong et al., 2006a)

2.5.1.3. Waveform Smoothing

To help remove noise and determine initial estimates for the waveform parameters, such as estimating the locations and amplitudes of the peaks in the waveform, the voltaged waveform has to be further smoothed by a Gaussian filter (Brenner et al., 2003; Duong et al., 2006a).

In this filter process, the method referenced in the two above article was used. The Gaussian shape is positioned where it maximum coincides with the filter location. The width of the Gaussian which is defined in terms of sigma, is usually described as the Full Width at Half Maximum (FWHM). The FWHM is linked to sigma by the formula: $FWHM = sigma * sqrt(8 * \log(2))$. A FWHM value of 3 is used in the actual implement of the method (Duong et al., 2006a).

2.5.2. Waveform Gaussian Fitting

A waveform can be considered as composition of Gaussian curves, and reflect Gaussian surface, i.e., the layers of branches and foliages in the forest. And the second set of waveform metrics proposed, AGS, SGS and MSGS, are statistical characteristics of the decomposed Gaussian components. Wagner et al. (2006) gave a theoretical basis for modelling the lidar waveform as a series of Gaussian pulses. The Gaussian fitting step is to model the normalized and smoothed waveform w(t) with Gaussian components using the algorithm developed by Brenner et al. (2003).

A smoothed waveform $w(t)$ is a sum of Gaussian components $W_m(t)$, as shown in the equation below:

$$w(t) = \sum_{m=1}^{N} W_m(t), \text{ with } W_m(t) = A_m \times e^{\frac{-(t-t_m)^2}{2\sigma_m^2}}$$

Where $w(t)$ is the amplitude of the waveform at time t; $W_m(t)$ is the *m*-th Gaussian component; *N* is the number of Gaussian components in the waveform; A_m is the amplitude of the *m*-th Gaussian component, t_m its position and σ_m its standard deviation.

The least squares approach is used to calculate the model parameters A_m, t_m, and σ_m in the above equation, by fitting the theoretical model to the observed waveform in such a way that the difference between the model and observation is minimized in the least-square sense. Figure 2-10 shows the results of the fitting algorithm.

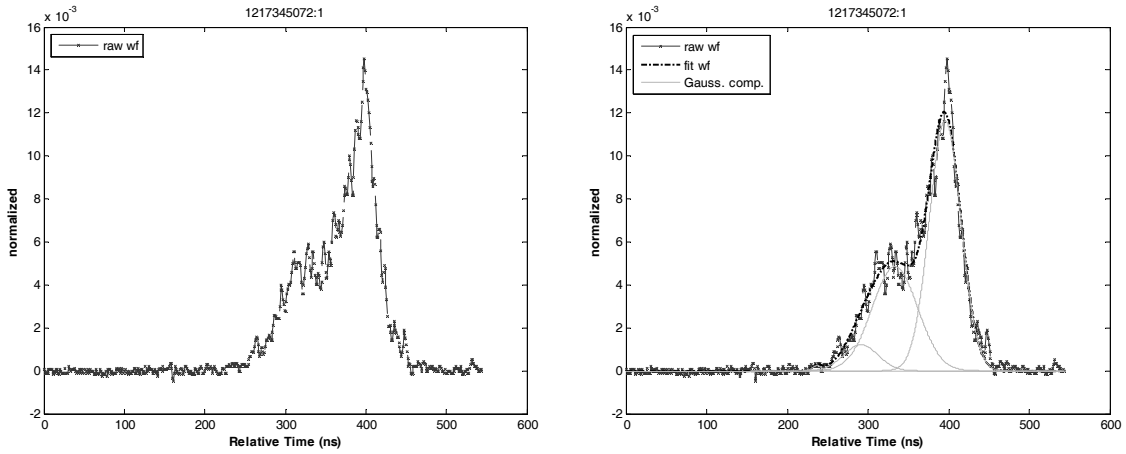


Figure 2-10 Lidar waveform on track 121745072 with shot number 1: raw waveform (left) and its fitted waveform (right), with raw waveform—red, the Gaussian component—green and fitted waveform—dashed black

As can be seen from the fitting result, the leftmost Gaussian component is referred to as the first mode, and this component corresponds to the first feature in the laser footprint that is reflecting, which over the forest areas, mostly originate from the reflection by the tree canopy. On the other hand, the right most Gaussian component corresponds to the energy reflected by the surface hit last, which in forest applications, corresponding to the last ground return below the tree.

2.5.3. Waveform Selection and Parameters Calculation

After the waveform is smoothed and fitting with Gaussian components, waveforms validate for the study are selected and some important waveform parameters are calculated, for the purpose of deriving the proposed metrics from.

2.5.3.1. Waveform Categorization and Selection

All the waveforms that have corresponding field survey data are firstly stratified by slope value extracted from DEM. Waveforms located on terrain with slope larger then 25 percent are excluded from the next analysis, because theoretically, within more than 18.75m of waveform range, return from trees will be convolved or blurred with the surrounding topography, which makes it impossible to derived the metrics proposed.

The remaining waveforms of each slope range are further checked to exclude those from which no clear ground return is discernable or in which there is only ground return (no canopy return), leaving waveforms validate for analysis.

2.5.3.2. Background Noise Threshold, Signal Start and Signal End

The locations of signal start and signal end can be determined, by locating the leftmost and rightmost position where the amplitude of smooth waveform firstly exceed the background noise threshold (Figure 2-11):

$$Signal\_start = (t \in I) \cap (t \le (\forall i \in I)),$$

$$Signal\_end = (t \in I) \cap (t \ge (\forall i \in I)), \text{ with } I = \{i \mid V_i > threshold, 1 \le i \le 544\}$$

Where V_i is the received energy of the smoothed waveform at moment i, and background noise threshold is set to mean noise value plus 4 times noise standard derivation (Lefsky et al., 2005), shown as the formula below:

$$Threshold = noise\_mean + 4 * noise\_std\_derivation$$

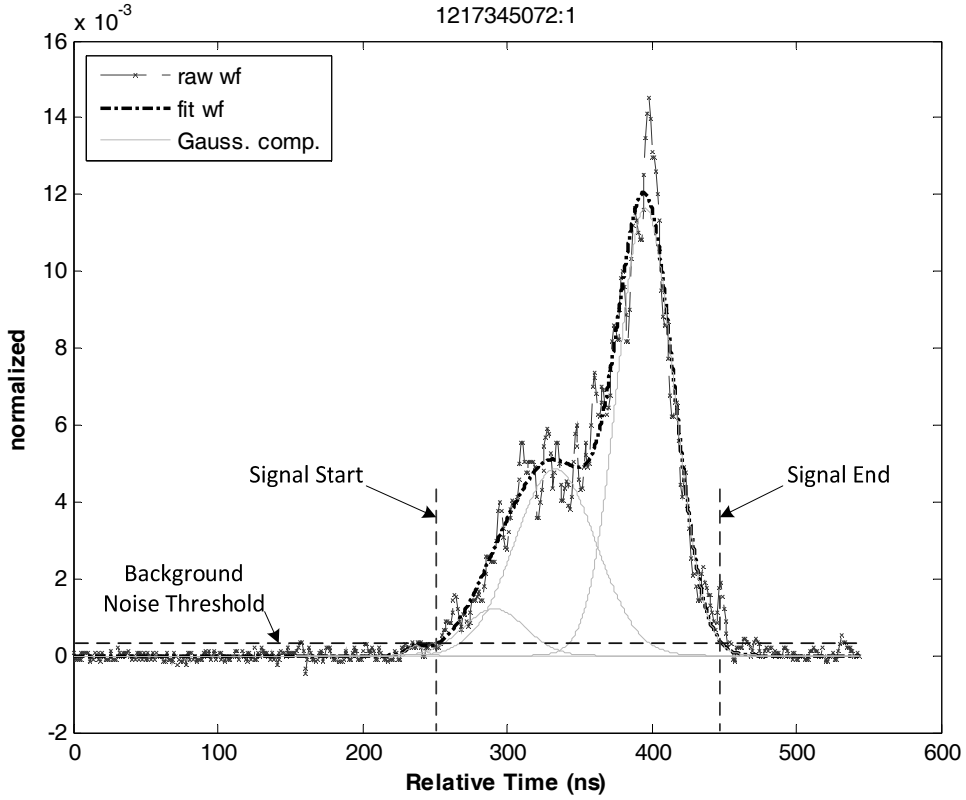


Figure 2-11 Waveform parameters as Background Noise Threshold, Signal Start and Signal End

In the study, the first 100 bins of the waveform are used to determine the noise component of the waveform signal, namely mean noise value and noise standard deviation, for the actual waveform signal often starts after the 100th bin within the 544 bins of the whole waveform, as indicated in the formula below:

$$noise\_mean = \overline{V} = \sum_{i=1}^{100} V_i / 100$$

$$noise\_std\_deviation = \Delta V = \sqrt{\sum_{i=1}^{100} (V_i - \overline{V})^2 / 100}$$

Where V_i is the received energy of the smoothed waveform at moment i.

2.5.3.3. Ground Return, Canopy Return and Canopy Height

To prepare for the waveform metrics derivation in the next step, it is important to determine the boundary between the canopy return and ground return of the waveform (Figure 2-12). In most cases, the ground return and canopy return are not clearly separated, or to say, there is always a overlap of ground return and canopy return in the waveform, especially when the topography relief is large and vegetation height is low within the lidar footprint. Below is the method used in the study to determine the boundary between the ground return and canopy return of the waveform.

Firstly, the Ground Return is identified as the Gaussian component with the highest peak within the right half of the waveform range,

$$Ground\_\mathrm{Re}turn = W_i(t), \text{ with } W_m(t) = A_m \times e^{\frac{-(t-t_m)^2}{2\sigma_m^2}},$$

$$\text{and } \forall t_i, t_m \in [(signal\_start + sig\_end)/2, signal\_end], A_i > A_m$$

Then the Boundary between the canopy return and ground return is then determined by formula:

$$Boundary = A_i - 1.5 * \sigma_i,$$

And Canopy Height:

$$Canopy\_Height = A_i - signal\_start,$$

Where $W_i(t)$, $W_m(t)$ is the *i*-th, *m*-th Gaussian component, i is the index of Ground Return of the waveform, A_i, A_m is the amplitude of the *i*-th, *m*-th Gaussian component, t_i, t_m their position and σ_i, σ_m their standard deviation.

The waveform profile and Gaussian components to the left of the boundary are canopy return part of the waveform.

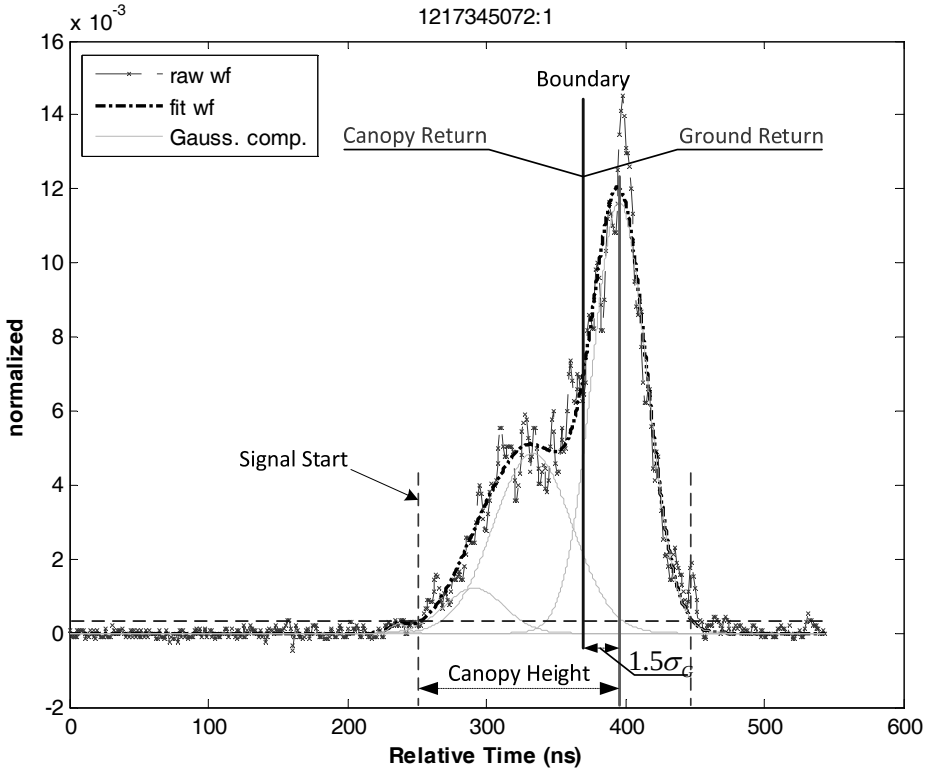


Figure 2-12 Waveform parameters as Ground Return, Canopy Return and Canopy Height

2.5.4. Waveform Metrics Derivation

As is proposed, two sets of waveform metrics are derived from the above prepared waveforms. The first set of waveform metrics, R25, R50 and R75, are quantile-based metrics of the canopy return energy; the second set of waveform metrics, AGS, SGS and MSGS, are statistical characteristics of the Gaussian component contributed by canopy reflecting.

2.5.4.1. Metrics R25, R50 and R75 Derivation

Quantile-based metrics , R25, R50 and R75, are derived from the lidar waveforms (Figure 1-8). CH25, CH50, CH75 are heights of 25%, 50% and 75% canopy return energy, respectively (Figure 2-13). Then the ratios R25, R50 and R75 will be CH25, CH50 and CH75 divided by canopy height, to normalize the influence of different canopy heights and the waveform spreading effect caused by topographical slope. These metrics will then be applied to statistical methods to find their relationships between different forest types.

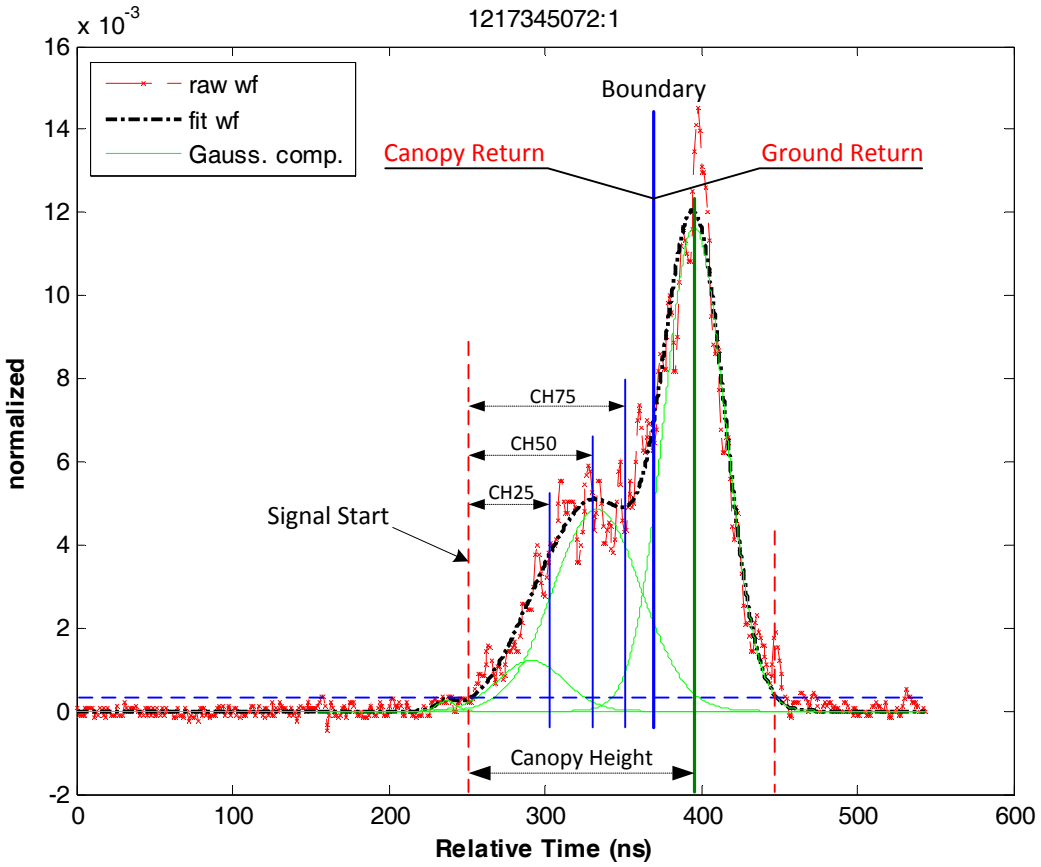


Figure 2-13 Heights of 25%, 50% and 75% canopy return energy, CH25, CH50 and CH75

2.5.4.2. Metrics AGS, SGS and MSGS Derivation

The slopes of the Gaussian components (see Figure 1-10, Figure 2-14) contributed by the canopy reflecting are expected to have relationships with the different forest types. The average and standard deviation of the decomposed Gaussian curve Slopes of the canopy return of each waveform, namely AGS and SGS (see Figure 1-11), are calculated by formula below:

$$AGS = \overline{S} = \sum_{i=1}^{n} S_i$$

$$SGS = \Delta S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - \overline{S})^2} \text{ with } S_i = A_i / \sigma_i$$

Where A_i is the amplitude of the *i*-th Gaussian component, n is the number of Gaussian components of canopy return part, and σ_i its standard deviation.

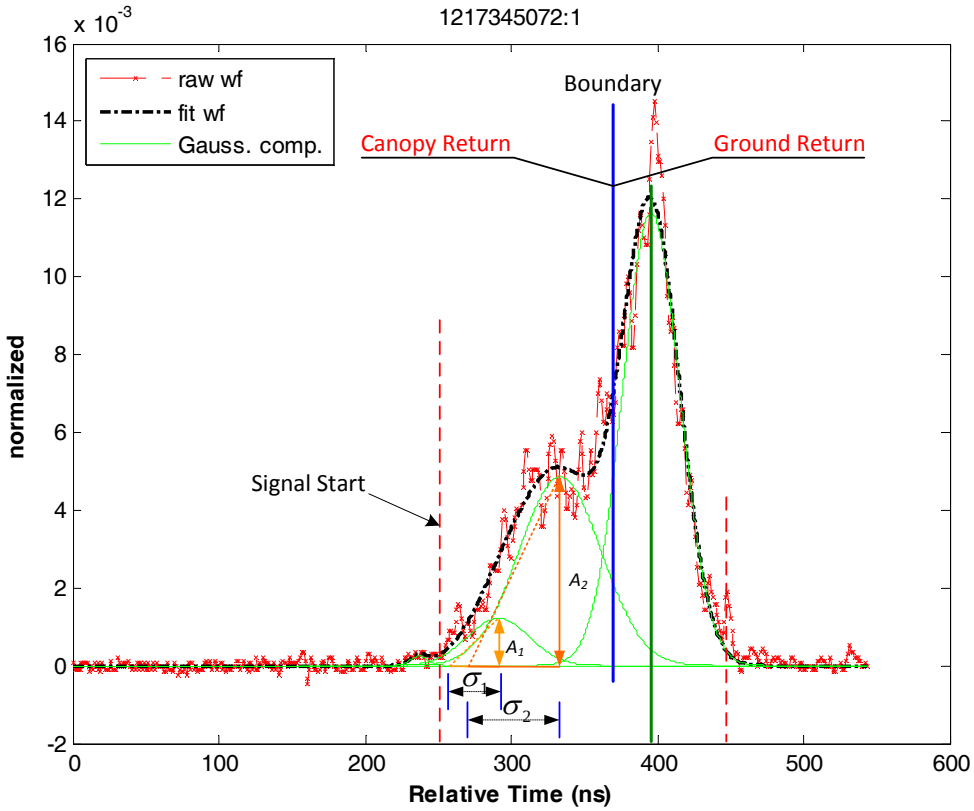


Figure 2-14 The slopes of the Gaussian components contributed by the canopy reflecting

A Modified Standard deviation of the decomposed Gaussian curve Slopes of the canopy return (MSGS) is also calculated using the formula shown below:

$$MSGS = \Delta S' = \sqrt{\sum_{i=1}^{n}\frac{E_i}{E_T}(S_i - \overline{S})^2} \text{ , with } E_T = \sum_{i=1}^{n} E_i$$

Where S_i is the Gaussian curve slope of the *i*-th Gaussian component of the canopy reflecting, E_i is the energy of the *i*-th Gaussian component of the canopy reflecting, E_T is the total energy of the Gaussian components of the canopy reflecting, and n is the number of Gaussian components of canopy return part.

This Modified Standard deviation of the decomposed Gaussian curve Slopes of the canopy return (MSGS) is expected to amplify the main vertical structure characteristics of the canopy, for the energy of each Gaussian component is used as weight in the formula. The main vertical structure of the canopy will have a stronger energy, thus have a larger weight in the calculation.

2.6. Data Analysis

2.6.1. Statistical Analysis of Metrics Derived

In order to find out the relationships between the metrics derived and different forest types, namely broadleaved forest, needleleaved forest and mixed forest, box-plots of different groups of metrics are produced to examine which metrics have the best potential to separate those different forest types.

2.6.2. Support Vector Machine (SVM) Classification and Validation

2.6.2.1. Introduction of SVM Classification

Support Vector Machine (SVM) is a powerful methodology for solving problems such as general (nonlinear) classification, regression and outlier detection with an intuitive model representation (Scholkopf and Smola, 2002; Vapnik, 1995). The foundation of Support Vector Machines (SVM) was developed by Vapnik (1995) and is gaining popularity in the machine learning community, creating an enthusiasm that Artificial Neural Networks used to create before, due to many attractive features and promising empirical performance (Gunn, 1998; Meyer, 2007).

Support Vector Machine, having no restriction on the dimensions (attributes) of input data, is superior to traditional Neural Network approaches, which suffered difficulties in generalization and resulted in overlearning (overfitting) (Gunn, 1998). What's more, SVM classifier can be trained to any size of training set, while a Neural Network need definitely more training nodes so as to achieve a reasonable classification result. Considering the relatively small sample size of the study, SVM is again a better classification method suitable for this study, than a Neural Network approach.

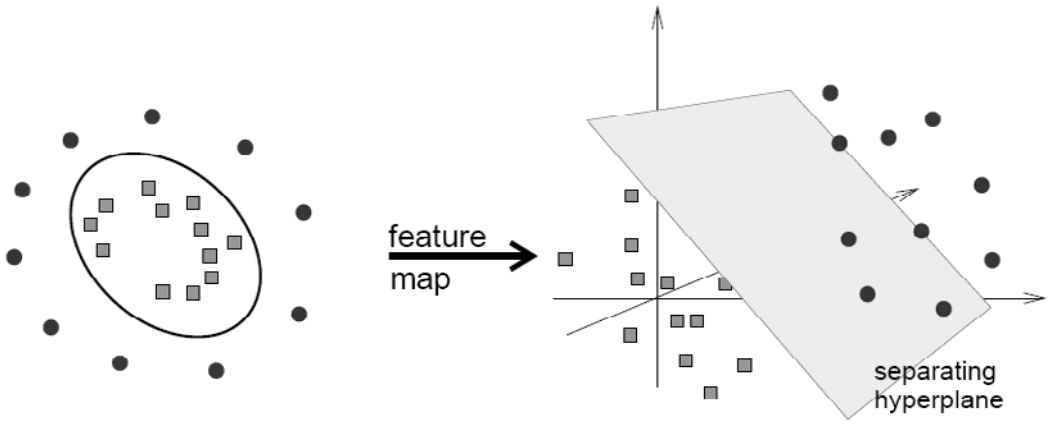


Figure 2-15 The mapping of vectors from gene (low demensional) space to feature space (hyperspace) using kernel function in the SVM classification (Markowetz, 2003)

The concept of SVM classification is to transfer the data (vector) that are often hard to separate in gene low dimensions, to a higher dimensional space by kernel functions, such as linear, polynomial and radial basis function, where the linear separability of the data is possible (Scholkopf and Smola, 2002;

Vapnik, 1995) as shown in Figure 2-15. In this way, an optimal separating hyperplane between data points of different classes in a (possibly) high dimensional space is calculated. And the actual Support Vectors are the points that form the decision boundary between the classes used to calculated the optimal separating hyperplane (Kumar et al., 2006). For more details of SVM classification methods, one can refer to more technical literature, such as *An Introduction to Support Vector Machine (and other kernel-based learning methods)* (Shawe-Taylor and Cristianini, 2000). Due to its advantage and promising performance, in our study, SVM classification method is applied to the derived metrics to classify the waveforms.

2.6.2.2. Application of SVM Classification on the Metrics Derived

In our study, the open-source data mining software Weka is used in the application of SVM classification to the metrics derived. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization (Weka, 2007).

The two sets of derived waveform metrics, reflecting canopy vertical structure distribution and canopy vertical stratum characteristics respectively, are grouped into five combinations, as shown in Table 1. and each of the combination of metrics is used as input attributes in the SVM classification in Weka platform, to test which metrics have the strongest relationship with and can best predict different forest types. The method is firstly applied to distinguish broadleaved and needleleaved forest, to find out the metrics combination with the highest classification accuracy. Then, classicisation using this metrics combination as input parameters is applied to the dataset when mixed forest is added in, to ask the research question “the extend to which the metrics can classify mix broad-/needle-leaved forest”.

In the SVM classification, the Libsvm algorithm (Chang and Lin, 2001) is used and linear Kernel function is selected. The detailed procedure of using Weka to classify the derived metrics is illustrated in Appendix II.

Table 1 Metrics combinations whose relationship with different forest types will be tested in SVM

	Metrics Combination	Metrics Attributes
1	R25, R50, R75	Canopy Return Energy Distribution
2	AGS, SGS	Statistical Characteristics of Gaussian Component
3	AGS, MSGS	Statistical Characteristics of Gaussian Component
4	AGS, SGS, MSGS	Statistical Characteristics of Gaussian Component
5	R25, R50, R75, AGS, SGS, MSGS	Canopy Return Energy Distribution and Statistical Characteristics of Gaussian Component

2.6.2.3. Classification Validation

10-fold cross-validation method is used in the SVM classification in Weka to perform classification validation. This cross-validation procedure can prevent the overfitting problem and help select the SVM model which will produce the best classification result (Chang and Lin, 2001). In this procedure, the training set will be firstly divided into 10 subsets of equal size, then sequentially one subset is tested using the classifier trained on the remaining 9 subsets. Thus, each instance of the whole training set is predicted once. The cross-validation accuracy is the highest percentage of data which are correctly classified during the procedure.

During this stage, confusion matrix, one of the common means of expressing classification accuracy, of the classifications is produced. How well a classification is categorized, as well as how bad the classification errors of commission and omission are, are shown in the matrix.

Classification result table, in which several main descriptive measures of the classification (Lillesand et al., 2004) is contained, is also obtained from the confusion matrix. The overall accuracy is the probability that a pixel randomly taken from the classification has the same class as the corresponding pixel in the reference data, and it is calculated by dividing the total number of correctly classified pixels in each class by the number of reference pixels. The producer's accuracy is the probability that a pixel taken randomly from the reference data class *X* has the same class as the corresponding pixel in the classified data. It is computed by dividing the number of correctly classified pixels in each class by the number of reference pixels of the class, and it indicates how well the reference data of the class are classified. The user's accuracy is the probability that a pixel randomly taken from the classified data class *X* has the same class as the corresponding pixel in the reference data. It is indicates the probability that a pixel classified to a class actually represents that class in the reference data. The user's accuracy results from dividing the number of correctly classified pixels in each class by the total number of pixels that were classified in that class.

3. RESULTS

3.1. Field Data Discription

Within the total 103 sample plots collected, the number of plots of broadleaved forest, needleleaved forest and mixed forest are 60, 16 and 19, respectively. And 6 plots are located on agriculture land and 2 on bare land. The histogram below (Figure 3-1) shows the distribution of sample plots of different forest types on slope ranges. As can be seen on the histogram, the largest slope of the sample plots is more than 55 percent, and the majority of sample plots are located on slopes of less than 30 percent. It can be also found out that when slope of the sample plots exceeds 25 percent, the ground return on the corresponding waveforms becomes hard to distinguish visually. Table 2 is a summary of sample plots with slope of less than 25 percent.

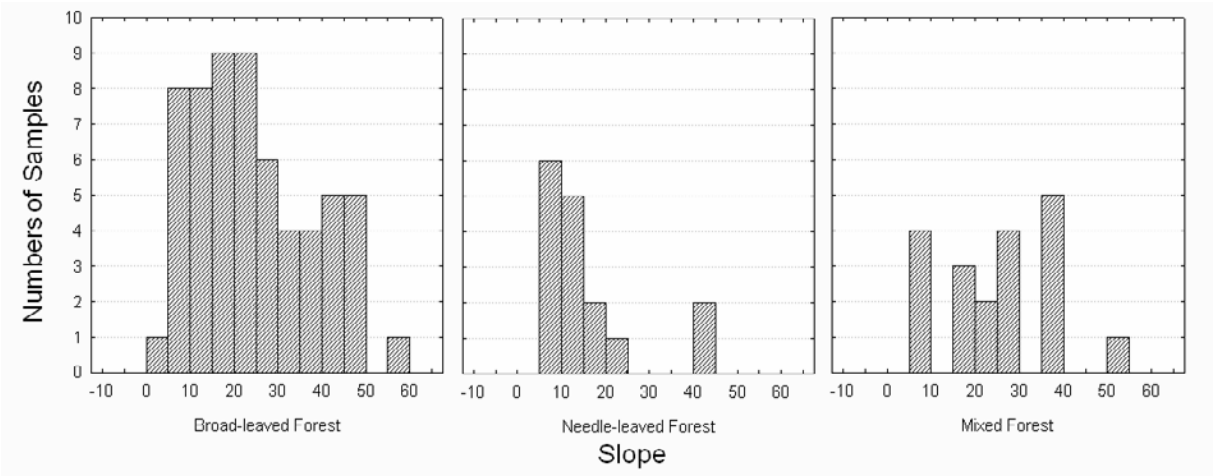


Figure 3-1 Histogram of sample plots distribution on slope ranges (slope unit: percent)

Table 2 Summary of sample plots with slope of less than 25 percent

	Broadleaved		Needleleaved		Mixed	
	0-25%	>25%	0-25%	>25%	0-25%	>25%
Number	35	25	14	2	9	10
Percentage*	58.33%		87.50%		47.37%	
Summary	(35+14+9)/(60+16+19)=61.05%					
Note: * the percentage of plots with slope range of 0-25%						

Field plots data with corresponding ICESat-GLAS waveform data acquired during forest growth season (September and October) were also selected from secondary dataset collected by Xing Yanqiu in 2006, to supplement to the primary dataset. Finally, A total number of 42, 27 and 16 plots of broadleaved forest, needleleaved forest and mixed forest were prepare for the study (see Table 3).

3.2. ICESat-GLAS Waveform Selected

Table 3 shows the result of the ICESat-GLAS waveforms selected which are used for the metrics derivation and further analysis. Consider for example the row of slope range of 5-10%, the total number of waveforms with corresponding field survey forest type of broadleaved forest is 9, and 8 of them are used for metrics derivation. One of them is excluded for further analysis because no clear ground return is discernable or there is only ground return (no canopy return) in the waveform. For the waveforms located on slope range large than 25%, they are all excluded, because theoretically they are unsuitable to derive the metrics proposed, for more than half of the waveform range, return from trees is convolved or blurred with that from the surrounding topography. The final waveforms selected validate for further analysis are 35 of broadleaved forest, 18 of needleleaved forest and 11 of mixed forest.

Table 3 ICESat-GLAS Waveforms selected categorized by slope range

	B			N			M		
Slope	**Total**	**Validate**	**%**	**Total**	**Validate**	**%**	**Total**	**Validate**	**%**
0-5%	1	1	100%	1	1	100%	0	0	0%
5-10%	9	8	89%	10	7	70%	5	4	80%
10-15%	8	6	75%	8	6	75%	0	0	0%
15-20%	11	10	91%	4	2	50%	4	3	75%
20-25%	13	10	77%	4	2	50%	7	4	57%
>25%									
Total	*42*	***35***	83.33%	*27*	***18***	66.67%	*16*	***11***	68.75%
							Overall valid: (35+18+11)/(42+27+16)=75.30%		

***B--Broadleaved; N--Needleleaved; M--Mixed;**

3.3. Statistical Analysis of Metics Derived

The two types, totally six, of the proposed waveform metrics were derived from the ICESat-GLAS waveforms selected in Section 3.2.. All the metrics can be found in Appendix III Table 15. Figure 3-2 and Figure 3-3 show the statistical analysis applied to the metric derived categorized by different forest types.

As can be seen from Figure 3-2, in the first row are box plots showing the metrics reflecting canopy return energy distribution and different forest types, and in the second row, the metrics reflecting canopy vertical stratum characteristics and different forest types. From a glimpse of all the six graphs, it can be concluded that none of the six metrics can separate the three different forest types alone. In the box plots of R25, SGS and MSGS, the boxes of the three forest types are largely overlapping. And in the box plots of R50, R75 and AGS, though the extend of the overlapping of the three forest types is smaller than that of R25, SGS and MSGS, however, these three forest types are still not clearly separable, especially when mixed forest is there.

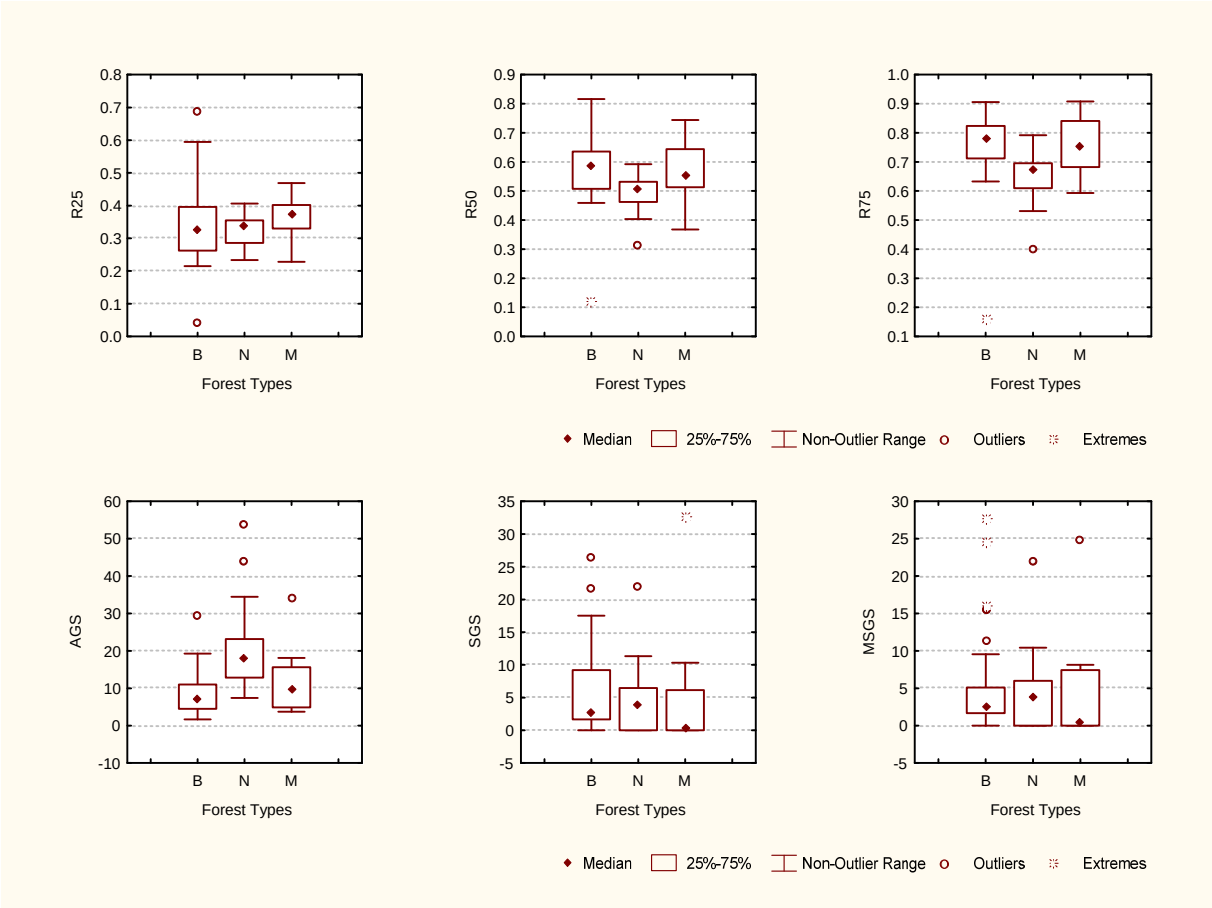


Figure 3-2 Box plots of lidar derived metrics and the three different forest types

Figure 3-3 shows the box plots of lidar metrics and two forest types – broadleaved and needleleaved forests. For the metrics of R25, SGS and MSGS, the boxes of broadleaved forest and needleleaved forest are largely overlapping, which means these metrics alone cannot be good indicators of the two types of forest. However, the metrics AGS of the broadleaved and needleleaved forest indicate a good separation, which shows the great potential in classifying the two types of forest in the next step.

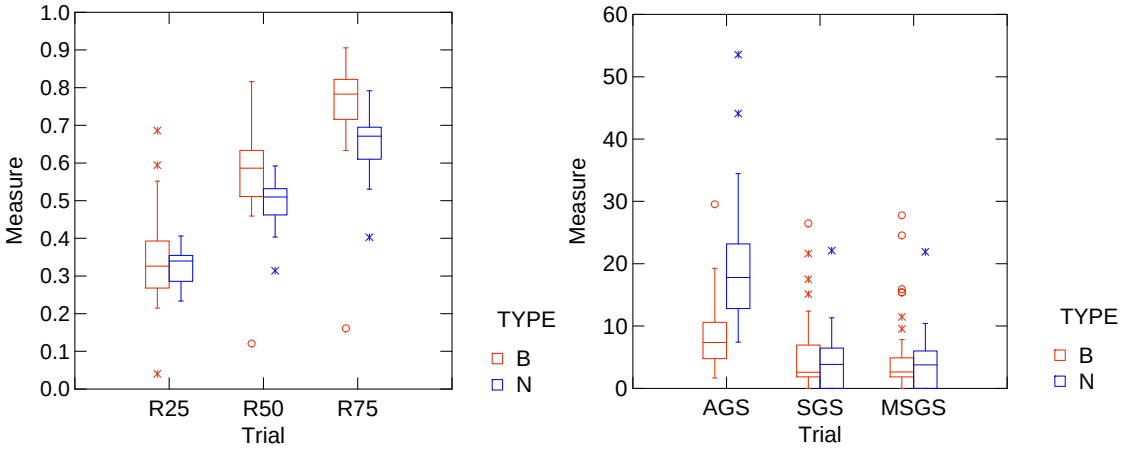


Figure 3-3 Box plots of the two types of metrics and broadleaved and needleleaved forest

3.4. Support Vector Machine Classification Results and Validation

3.4.1. Broadleaved and Needleleaved Forests Classification Result

In this section, SVM classification results of broadleaved and needleleaved forest using different types of metrics are firstly presented. Then the results of different combinations of metrics as input parameters in SVM classification are compared.

3.4.1.1. Metrics R25, R50 and R75 – Reflecting Canopy Return Energy Distribution – as Input Classification Parameters

The SVM classification result of the two forest types using metrics R25, R50 and R75 is presented in this section.

Table 4 shows the confusion matrix between the SVM classification and reference data collected in the field. The columns contain reference data with known forest types, and the rows are classified waveform data. From the table, we can see though 35 broadleaved forest area are all correctly classified as broadleaved forest, the needleleaved are poorly classified – all of them are misclassified as broadleaved forest. Or in other words, the two forest types are not separated by the classification.

Table 5 shows the classification result. The producer's accuracy of the broadleaved forest is 100% and the user's accuracy is 66.4%. As all the needleleaved forest are misclassified, the producer's and user's accuracy are both 0%. And the Kappa statistic of the classification is 0, which means these two forest types can not be separated, using the metrics reflecting canopy vertical structure distribution.

Table 4 Confusion matrix of SVM classification of broadleaved and needleleaved forest using R25, R50 and R75 as input parameters

	Reference data		
Classified data	Broadleaved	Needleleaved	**Row Total**
Broadleaved	**35**	18	53
Needleleaved	0	**0**	0
Column Total	35	18	**53**

Table 5 Classification results of broadleaved and needleleaved forest using R25, R50 and R75 as input parameters

Class name	r	Classified total	Number correct	Accuracy (%)		Kappa
				Producer	User	
Broadleaved	35	53	35	100.00%	66.04%	
Needleleaved	18	0	0	0.00%	0.00%	
Total	53	53	35			
Overall Classification Accuracy =			**66.04%**			**0**

3.4.1.2. Metrics AGS, SGS and MSGS – Reflecting Canopy Vertical Stratum Characteristics – as Input Classification Parameters

SVM classification results using three combinations of metrics reflecting canopy vertical stratum characteristics, 1) AGS and SGS; 2) AGS and MSGS; 3) AGS, SGS and MSGS respectively, are produced. Classification result produced by AGS and MSGS is specified in this part.

Figure 3-1 shows the SVM classification results of the two forest types using metrics AGS and MSGS. The cross symbols represent for correctly classified forest, and blue cross for broadleaved forest, while red for needleleaved forest. The square symbols represent for misclassified forest. And we can see from the graph that 5 points are misclassified, with 3 of them needleleaved forest and 2 of them broadleaved. However, a large portion of the points, totally 48 in fact, are correctly classified. Also, we can see the red crosses (needleleaved forest) are posited at the up-left part of the graph, while the blue crosses (broadleaved forest) are located at the down-right part of the graph. With the same AGS value, the needleleaved forests usually have a smaller MSGS.

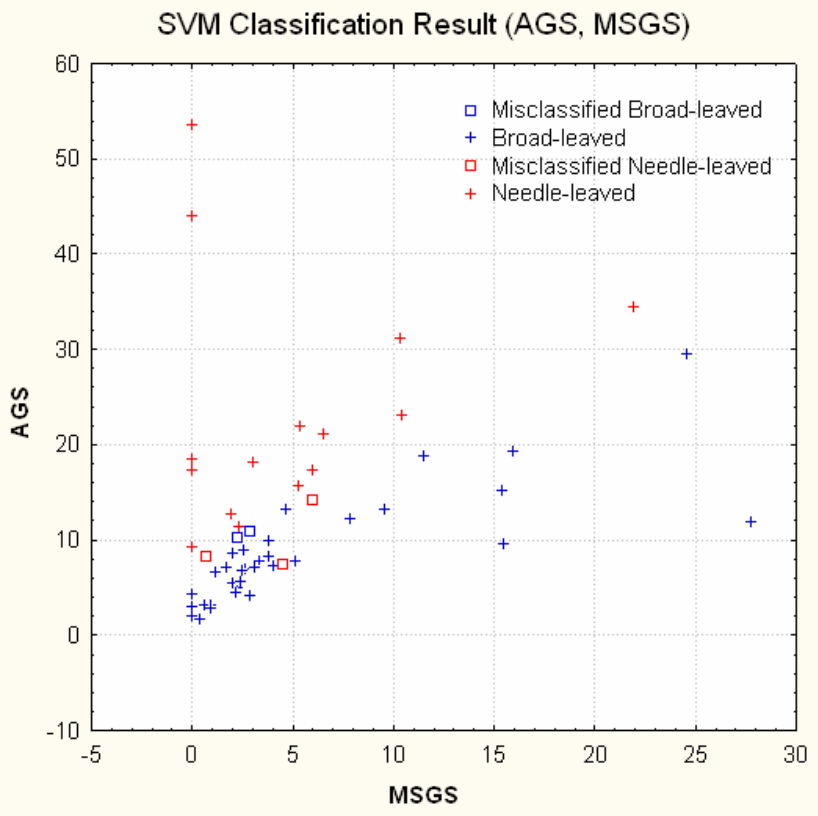


Figure 3-4 SVM classification results of broadleaved and needleleaved forest using AGS and MSGS as input parameters

Table 6 shows the confusion matrix between the SVM classification and reference data. The columns contain reference data with known forest types, and the rows are classified waveform data. From the table, we can see 33 out of 35 broadleaved forest area are correctly classified as broadleaved forest, and 15 out of 18 needleleaved forest are correctly classified.

Table 7 shows the classification result. The producer's accuracy of the broadleaved forest is 94.29% and the user's accuracy is 91.67%, while for the needleleaved forest, the producer's accuracy is 83.33% and the user's accuracy 88.24%. Moreover, an overall accuracy of 90.57% is obtained, with the Kappa statistic of 0.7868, which means the classification is 78.68 percent better than one resulting from chance. The Kappa statistic value, together with the high producer's and user's accuracy of both broadleaved and needleleaved forests, indicates this is a quite promising result for the robust classification method based on waveform analysis.

Table 6 Confusion matrix of SVM classification of broadleaved and needleleaved forest using AGS and MSGS as input parameters

	Reference data		
Classified data	Broadleaved	Needleleaved	**Row Total**
Broadleaved	**33**	3	36
Needleleaved	2	**15**	17
Column Total	35	18	**53**

Table 7 Classification results of broadleaved and needleleaved forest using AGS and MSGS as input parameters

Class name	**r**	**Classified total**	**Number correct**	**Accuracy (%)**		**Kappa**
				Producer	**User**	
Broadleaved	35	36	33	94.29%	91.67%	
Needleleaved	18	17	15	83.33%	88.24%	
Total	53	53	48			
Overall Classification Accuracy =			**90.57%**			**0.7868**

3.4.1.3. Comparison of Classification Results of Different Combinations of Metrics

Table 7 shows the comparison of SVM classification results using different combinations of metrics as input parameters. We can see some interesting points from Table 7. The classification of metrics combination (the row in red of Table 8), R25, R50 and R75, which reflect canopy vertical structure distribution, get the worst classification result, with the overall accuracy of 66.04% and Kappa statistic 0, which means the metrics of canopy return energy distribution cannot separate the two types of forest.

However, the metrics combinations reflecting canopy vertical stratum characteristics (shown in the 2nd to 4th row of Table 8) achieve better classification results of the two types of forest. Among the three combinations, AGS and MSGS (the row in blue of Table 8) achieve the highest overall accuracy of 90.57% and Kappa statistic value of 0.7868 and best predicts the broadleaved and needleleaved forests. The metric combination AGS and SGS, though having the same overall accuracy as that of AGS and MSGS, gets a lower Kappa value of 0.7808. When SGS is added into the combination AGS and MSGS, rather than obtaining a better classification result, the overall accuracy and Kappa decreases to 88.68% and 0.7406 respectively, which means SGS can even bias the classification. Moreover, when metrics R25, R50 and R75 are also added as input parameters of the classification, the result shows no improvement, which again proves that the metrics R25, R50 and R75 cannot be used as indicators of different forest types.

Table 8 Comparison of SVM classification results of different combinations of metrics

Metrics	Classified Total	Number correct	Overall Accuracy	Kappa	Broadleaved		Needleleaved	
					Producer	User	Producer	User
R25, R50, R75	**53**	**35**	**66.04%**	**0**	***100.00%***	***66.04%***	***0.00%***	***0.00%***
AGS, SGS	53	48	90.57%	0.7808	97.14%	89.47%	77.78%	93.33%
AGS, MSGS	**53**	**48**	**90.57%**	**0.7868**	***94.29%***	***91.67%***	***83.33%***	***88.24%***
AGS, SGS, MSGS	53	47	88.68%	0.7406	94.29%	89.19%	77.78%	87.50%
R25, R50, R75, AGS, SGS, MSGS	53	47	88.68%	0.7406	94.29%	89.19%	77.78%	87.50%

3.4.2. Classification Result when Mixed Forest is added

After finding out the best metrics combination, AGS and MSGS, as indicator of broadleaved and needleleaved forests, the mixed broad-/needle-leaved forest data is added into the classification. As is shown in Table 9, these two metrics appear unable to distinguish mixed forest from broadleaved and needleleaved forest. All 11 mixed forest are misclassified, with 6 of them are misclassified as broadleaved forest and 5 of them needleleaved forest. Thus, Table 10 shows the classification has an overall accuracy of only 76.56% and Kappa statistic value of 0.5642. The result shows that these two metrics cannot separate mixed broad-/needle-leaved forest from broadleaved and needleleaved forest.

Table 9 Confusion matrix of SVM classification of broadleaved, needleleaved and mixed forest using AGS and MSGS as input parameters

Classified data	Reference data			Row Total
	Broadleaved	Needleleaved	Mixed	
Broadleaved	**35**	3	6	44
Needleleaved	0	**14**	5	19
Mixed	0	1	**0**	1
Column Total	35	18	11	**64**

Table 10 Classification results of broadleaved, needleleaved and mixed forest using AGS and MSGS as input parameters

Class name	r	Classified total	Number Correct	Accuracy (%)		Kappa
				Producer	User	
Broadleaved	35	44	35	100.00%	79.55%	
Needleleaved	18	19	14	77.78%	73.68%	
Mixed	11	1	0	0.00%	0.00%	
Total	64	64	49			
Overall Classification Accuracy =			**76.56%**			**0.5642**

4. DISCUSSION

In this study, the applicability of retrieving forest type information from spaceborne large-footprint waveform lidar is further investigated. The large-footprint waveform Lidar has unique advantage of providing vertical structure information of the Earth's surface, over the conventional remote sensing technology. Researchers have attempted to utilize the vertical information provided by lidar waveforms to derive land cover and forest vertical structure information. However, derivation of forest type information from large-footprint lidar waveforms has not been attempted, which are import in applications such as mapping, forest resource management, ecological modelling and forest aboveground biomass estimation.

The study by Duong et al. (2006b) tried to classify the land cover into four classes, water, urban, high vegetation and bare land, on flat areas using ICESat-GLAS lidar waveforms, while no sub-classification of different forest types was attempted. The only research tried to establish a relationship between the lidar waveform and different forest types by Ranson et. al. (2004), investigated the possibility of using metrics front slope angle and centroid, both derived from ICESat-GLAS lidar waveform on terrain with less than 5 degrees slope, as indicators of forest types. However, these two metrics failed to explain the relationship.

The study developed a new method to extract the required forest type information from large-footprint lidar waveforms based on full waveform analysis. For this purpose, the raw waveform was decomposed into Gaussian components, and canopy return and ground return of the waveform were separated. Two types of metrics hypothesised to have relationship with forest types were derived. The first type of metric, R25, R50 and R75, is quantile-based metrics based on the vertical distribution of canopy return energy; and the second type, AGS, SGS and MSGS, are statistical characteristics based on the decomposed Gaussian components of canopy return part. Different combinations of the metrics derived were then used as input parameters in Support Vector Machine classification to find their relationship with different forest types. And the highest accuracy of 90.57% and Kappa statistic of 0.7868 were achieved applying this method to separating broadleaved and needleleaved forests.

The results showed that the first type of metrics, R25, R50 and R75, reflecting canopy vertical structure distribution failed to distinguish broadleaved and needleleaved forests. This may be due to the complexity of the forest vertical structures. In theory, the homogenous broadleaved forest and needleleaved forest have typical different canopy vertical structures. However, most of the forests are a mixture of different tree species and of different canopy height in the study area. What's more, lots of shrubs are growing in the under-layer of the broadleaved forest (see Figure 4-1), all of which makes the canopy vertical structure distribution of broadleaved forest and needleleaved forest are not so distinct; Another reason may lie in the topographic effects on the vertical distribution of canopy return energy, which will also weaken the differences of vertical distribution of canopy return energy of different forest types.

(a) broadleaved forest

(b) broadleaved forest

(c) needleleaved forest

(d) needleleaved forest

Figure 4-1 Broadleaved and needleleaved forest senses. Lots of shrubs exist in under-layer of some broadleaved forests; Canopy strata is more obvious in needleleaved forests.

In contrast, the second type of metrics, especially metrics combination of AGS and MSGS, showed great promise in distinguishing the two types of forest. An overall accuracy of 90.57% and Kappa statistic of 0.7868 was achieved using AGS and MSGS as input parameters in SVM classification. It indicated that the slope of the decomposed Gaussian components in the lidar waveform can well mirror the characteristics of the reflecting Gaussian surfaces, in the forest sense, the layers of branches and foliage. The Average canopy return Gaussian component Slope (AGS) of waveforms produced by needleleaved forest is often larger that that produced by broadleaved forest. Moreover, the Modified Standard deviation of canopy return Gaussian component Slopes (MSGS) of needleleaved forest are smaller than that of broadleaved forest which have similar AGS value (see Figure 3-1). This could be explained by that needleleaved forest used to have more obvious canopy vertical strata than broadleaved forest (Figure 4-1). The metrics combination AGS and MSGS got better classification result than the combination AGS and SGS may because metrics MSGS amplify the main vertical structure characteristics of the canopy.

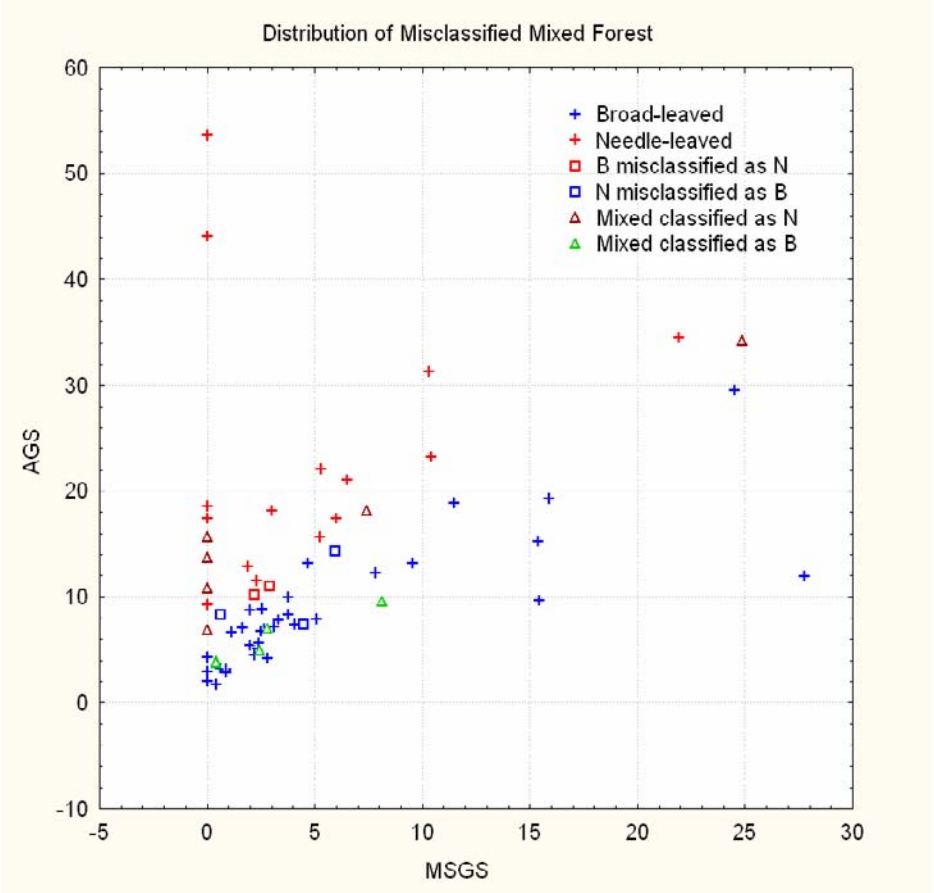


Figure 4-2 Classification result of all the waveforms using the model developed by AGS and MSGS

However, the metrics AGS and MSGS seem to fail to distinguish mixed broad-/needle-leaved forest from broadleaved and needleleaved forest. The distribution of the mixed forest in the two dimensional plane defined by AGS and MSGS is shown in Figure 4-2 above. This may be because the waveform metrics of mixed forest, as a transition from broadleaved forest to needleleaved forest causing an increase of the portion of needleleaved tree in the forest, may behave more like needleleaved forest when needleleaved trees structure are dominant the canopy of the forest plots. Table 11 shows a re-categorization of mixed forest according the portion of needleleaved trees in the sample plots. And of course whether needle-leave trees structure dominate the canopy structure of the forest plot, is not only determined by the portion of needleleaved trees in number, but also factors such as the tree height, canopy size of the needleleaved trees as well as the broadleaved trees in the plot. In this case we just consider the portion of needleleaved trees in number of the sample plot. If we reclassify mixed forest as needleleaved forest when the percentage of the number of needleleaved trees in the plots exceeds 40%, or reclassify it as broadleaved forest, only four of the "mixed forest" will be wrongly predicted by the model developed by metrics AGS and MSGS to separate broadleaved and needleleaved forests.

This also suggests that when applying the method developed in this study to other forest areas, or using large-footprint lidar waveforms to do classification of other types of features on Earth's surface, researchers have to be sure there are remarkable differences on vertical structure can be reflected by lidar waveforms between the objective classes in the study area, so as to get a good classification result.

Table 11 Information of misclassified mixed broad-/needle-leaved forest

SPOT_ID	FOREST TYPE	PREDICT TYPE	Num	Needle%	Threshold by 35%	Threshold by 40%	Note
1221245452:07	M	N	35	40	N	N	*Populus ; larix*
1221245452:11	M	B	15	26.7	B	B	*Betula ; larix*
1221245452:31	M	B	19	36.8	N	B	*Populus ; Larix*
1221245452:35	M	B	13	46.2	N	N	*larix ; Betula*
1221245452:37	M	B	21	38.1	N	B	*Larix ; Betula*
898951817:19	M	N	48	54.2	N	N	*Abies nephrolepis ; Betula costata*
898951817:20	M	N	45	53.3	N	N	*Picea koraiensis, Abies nephrolepis, Betula costata*
898951817:21	M	N	52	46.2	N	N	*Abies nephrolepis, Picea koyamai*
898951817:22	M	N	85	36.5	N	B	*Betulla castata, Abies nephrolepis*
898951847:05	M	B	13	61.5	N	N	*Betula platyphylla, Larix olgensis*
902852247:40	M	N	55	18.2	B	B	*Betula castata, Betula platyphylla*

Note: Needle%--percentage of number of needleleaved trees in the plot; B—Broadleaved forest; N—Needleleaved forest; M—Mixed forest;

Threshold by *%--When Needle% exceeds *%, then reclassified Mixed forest as Needleleaved forest, or Broadleaved forest.

The rows in gray are Mixed forests predicted by the model as Broadleaved forests; N and B are forests mis-predicted.

The method can be improved to apply to classify waveforms located on terrain larger than 25 percent. The solution of this matter is to employ a relief correction model, such as the one proposed by Lefsky et al. (2005) to get the canopy height on the up-slope, thus in turn get the canopy return range on the waveforms. Through this method, the problem of separating canopy return from ground return in the waveforms on large slopes can be solved (However, the results produced by the model vary across studies. In the case of Lefsky et al. (2005), the model explained 59%-68% variance in field-measure canopy height in three forest ecosystems. The study by Gwenzi (2008) in Wangqing forest area, showed that when slope level exceeds 22%, the mode fails to accurately predict the field-measured canopy height), and metrics reflecting canopy stratum characteristics, namely AGS, SGS and MSGS can be derived. In this way, this method can be applied to waveform on terrain with larger slopes to some extend, and eliminate the subjective in ground return determination exist in the many studies (In this study, as only waveforms located on slope less than 25 percent were selected, we assume the Gaussian component with the highest peak within the right half of the waveform range as the ground return), to investigate the relationship between the metrics and different forest types on large slope areas.

Furthermore, the number of sample plots collected in the study may be a constraint. Due to the limited time and rough terrain, only a total number of 103 plots was collected during the 12 days of fieldwork. What's more, the dataset was far from ideal: the forest is not homogeneous, and most of them are in montane area. Together with 102 secondary data, after excluding the waveforms unsuitable for analysis, only 64 data were validate in the study. More field data are expected to further examine the effects of slopes on the metrics derived and to get a more reliable result.

5. CONCLUSIONS

In this thesis, a method for deriving forest type information was developed, and broadleaved forest and needleleaved forest classes could be successfully separated. In this way, the aims of investigating the possibility of using large-footprint lidar to derive forest type information, thus extending the application of large-footprint lidar and contributing to the global methodology of a more accurate biomass estimation using large-footprint waveform lidar was achieved. This research further verified that the canopy vertical stratum characteristics could be well reflected by the canopy return part of the decomposed Gaussian components of the large-footprint lidar waveforms, thus the hypothesis was accepted that the proposed metrics derived from ICESat-GLAS waveforms on low relief areas can be used to derive forest types information.

Research Question 1. For what slope range, can the waveform metrics proposed be successfully derived?

From visual interpretation of the waveforms for which corresponding field plots data were collected, it is concluded that from about 75.30% of the waveforms located on slope of less than 25% percent where canopy return and ground return could be clearly separated, in other words, from which the waveform metrics proposed could be successfully derived. What's more, theoretically, waveforms located on slopes larger than 25% were not suitable to derive the metrics from by using the proposed method, because more than half of the waveform range return from trees is convolved or blurred with that from the surrounding topography.

Research Question 2. How strong is the relationship between the metrics of canopy energy distribution (R25, R50, R75), which reflect canopy vertical structure distribution, and broadleaved and needleleaved forest?

In our study, the metrics of canopy energy distribution (R25, R50 and R75), which reflect canopy vertical structure distribution, showed little relationship with broadleaved and needleleaved forest. These metrics were used as input parameters in SVM classification of broadleaved and needleleaved forests, however, it ended up with the Kappa statistic value of 0, which means they are incapable of separating the two types of forest.

Research Question 3. How strong is the relationship between the metrics of statistical characteristics of decomposed Gaussian curve slopes (AGS, SGS), which reflect canopy vertical stratum characteristics, and broadleaved and needleleaved forest?

The metrics of statistical characteristics of canopy return part of the decomposed Gaussian curve slopes of the waveforms (AGS, SGS and Modified SGS, or MSGS), which reflect canopy vertical stratum characteristics, showed a strong relationship with broadleaved and needleleaved forest. The lowest accuracy and Kappa statistic achieved by the metrics combinations of the three metrics applied to SVM classification to classify these two types of forest was 88.68% and 0.7406 respectively (by

AGS, SGS and MSGS), and the highest accuracy and Kappa statistic of 90.57% and 0.7868 (by AGS and MSGS), respectively.

Research Question 4. Which combination of the derived metrics can best predict the presence of broadleaved and needleleaved forests? What is the accuracy of the classification?

The metrics combination of AGS and MSGS can best predict the presence of broadleaved and needleleaved forests. The accuracy of the classification is 90.57%.

Research Question 5. How strong is the relationship between the metrics and mixed broad-/needle-leaved forest?

The metrics combination, AGS and MSGS, that best predicted the presence of broadleaved and needleleaved forest failed to distinguish mixed broad-/needle-leaved forest from broadleaved and needleleaved forests.

6. RECOMMENDATIONS

6.1. Metrics Derivation Method

The relief correction model (Lefsky et al., 2005) is suggested for use in future work to be able to develop a more powerful and automatic metrics derivation method. It is an obstacle to separate canopy return from ground return so as to derive the proposed metrics in the study, which has also been for a long time a problem in other researches. By employing this relief correction model, the canopy height on the up-slope can be got, thus in turn the canopy return range in the waveforms can be calculated and identified. In this way, the metrics proposed in this study can be derived from waveforms on terrain with larger slopes to some extend, and investigation of the relationship between the metrics and different forest types on large slope areas will become possible. What's more important, it will finally eliminate the subjective in ground return determination exist in the many studies and help to identify canopy return automatically.

6.2. Validation Method

Other ancillary passive optical imagery or forestry survey map from which high accuracy forest type information can be retrieved are recommended for future studies. Forest type information of more ICESat-GLAS footprints need to be obtained, so that the SVM model developed can be used in a much larger validation set to better test the relationship between the metrics and different forest types.

6.3. Biomass Estimation with Additional Forest Type Information

The derived forest type information is suggested to be taken into consideration at the third and forth phases in forest AGBM estimation process (see Figure 1-3), especially for the studies that are not conducted in homogeneous forest areas or forests of consistent horizontal spatial pattern, due to the facts that forest biomass may vary considerably between different forest types, even when of similar height, density and age.

REFERENCE

Aboal, J. R., J. R. Arevalo, and A. Fernandez (2005). Allometric relationships of different tree species and stand above ground biomass in the Gomera laurel forest (Canary Islands). *Flora - Morphology, Distribution, Functional Ecology of Plants, 200*(3), 264-274.

Austin, J. M., B. G. Mackey, and K. P. V. Niel (2003). Estimating forest biomass using satellite radar: an exploratory study in a temperate Australian Eucalyptus forest. *Forest Ecology and Management, 176*(1-3), 575-583.

Balzter, H., L. Skinner, A. Luckman, and R. Brooke (2003). Estimation of tree growth in a conifer plantation over 19 years from multi-satellite L-band SAR. *Remote Sensing of Environment, 84*(2), 184-191.

Blair, J. B., D. L. Rabine, and M. A. Hofton (1999). The laser vegetation imaging sensor (LVIS): A medium-altitude, digitations-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*, 115-122.

Bortolot, Z. J., and R. H. Wynne (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.

Brenner, A. C., H. J. Zwally, C. R. Bentley, B. M. Csathó, D. J. Harding, M. A. Hofton, et al. (2003). Derivation of Range and Range Distributions From Laser Pulse Waveform Analysis for Surface Elevations, Roughness, Slope, and Vegetation Heights. Algorithm Theoretical Basis Document, Version 4.1.

Brown, S. (2002). Measuring carbon in forests: current status and future challenges. *Environmental Pollution, 116*(3), 363-372.

Brown, S., P. Schroeder, and J. S. Kern (1996). Mitigation of carbon emission to the atmosphere by forest management. *Commonwealth Forestry Review, 75*(1), 80-91.

Castel, T., F. Guerra, Y. Caraglio, and F. Houllier (2002). Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact on radar signature. *Remote Sensing of Environment, 79*(1), 30-41.

Chang, C. C., and C. J. Lin (2001). LIBSVM: a library for support vercter machines. Retrieved Dec. 20, 2007, from Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, detailed documentation can be found in http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.ps.gz

DeGier, A. (2000). Woody biomass assessment in woodlands and shrublands. In: Temu (Editor), Proceedings Off-forest tree resources of Africa, Arusha, Sunlitho, Nairobi.

DeGier, A. (2003). A new approach to woody biomass assessment in woodlands and shrublands. In: P.S. Roy (Editor), Geoinformatics for Tropical Ecosystems. Asian Association of Remote Sensing, Bishen Singh Mahendra Pal Singh, Dehra Dun, India.

Dixon, R. K., R. A. Houghton, A. M. Solomon, M. C. Trexler, and J. Wisniewski (1994). Carbon pools and flux of global forest ecosystems. *Science, 263*, 185-190.

Dobson, M. C., F. T. Ulaby, L. Thuy, A. Beaudoin, E. S. Kasischke, and N. Christensen (1992). Dependence of radar backscatter on coniferous forest biomass. *IEEE Trans. Geosci. Remote Sens., 30*(2), 412-415.

Dong, J., R. K. Kaufmann, R. B. Myneni, C. J. Tucker, P. E. Kauppi, J. Liski, et al. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sensing of Environment, 84*(3), 393-410.

Drake, J. B., R. O. Dubayah, D. B. Clark, R. G. Knox, J. B. Blair, M. A. Hofton, et al. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment, 79*(2-3), 305-319.

Drake, J. B., R. O. Dubayah, R. G. Knox, D. B. Clark, and J. B. Blair (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment, 81*(2-3), 378-392.

Dubayah, R. O., and J. B. Drake (2000). Lidar remote sensing for forestry. *Journal of Foresty, 98*(6), 44-46.

Duong, H., N. Pfeifer and R. Lindenbergh (2006a). *Analysis of repeated ICESat full waveform data: methodology and leaf-on / leaf-off comparison.* Paper presented at the Workshop on 3D Remote Sensing in Forestry, Vienna.

Duong, H., N. Pfeifer and R. Lindenbergh (2006b). Full Waveform Analysis: ICESat Laser Data for Land Cover Classification. *ISPRS, Commission VII.*

Evans, D. L., S. D. Roberts, and R. C. Parker (2006). LiDAR - A new tool for forest measurements? *The Forest Chronicle, 82*(2), 211-218.

FAO. (2007). FAO Forestry Department. Retrieved June 18, 2007, from http://www.fao.org/forestry/en/

Gunn, S. R. (1998). *Support Vector Machine for Classification and Regression*: University of Southampton.

Gwenzi, D. (2008). *Spaceborne Lidar canopy height estimation for aboveground forest biomass assessment in the cool montane area of North East China.* ITC, Enschede.

Harding, D., and C. C. Carabajal (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical research letters, 32*(L21S10, doi:10.1029/2005GL023471).

Hudak, A. T., M. A. Lefsky, W. B. Cohen, and M. Berterretche (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment, 82*(2-3), 397-416.

Hyde, P., R. Dubayah, B. Peterson, J. B. Blair, M. Hofton, C. Hunsaker, et al. (2005). Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. *Remote Sensing of Environment, 96*(3-4), 427-437.

IPCC. (2001). *Climate Change 2001: Working Group Ⅰ: The Scientific Basis.* New York: Combridge University Press.

IPCC. (2003). *Good Practice Guidance for Land Use, Land-Use Change and Forestry.* Hayama: Insititution for Global Enrivonmental Strategies (IGES).

ISPRS. (1999). Remote Sensing and the Kyoto Protocol: A Review of Available and Future Technology for monitoring Treaty Compliance, Ann Arbor, Michigan, USA.

Kiehl, J. T., and K. E. Treberth (1997). Earth's Annual Globle Mean Energy Budget. *Bulletin of the American Meteorological Society, 78*(2), 197-208.

Phat, N. K., W. Knorr, and S. Kim (2004). Appropriate measures for conservation of terrestrial carbon stocks--Analysis of trends of forest management in Southeast Asia. *Forest Ecology and Management, 191*(1-3), 283-299.

Kimes, D. S., K. J. Ranson, G. Sun, and J. B. Blair (2006). Predicting lidar measured forest vertical structure from multi-angle spectral data. *Remote Sensing of Environment, 100*, 503-511.

Kraenzel, M., A. Castillo, T. Moore, and C. Potvin (2003). Carbon storage of harvest-age teak (Tectona grandis) plantations, Panama. *Forest Ecology and Management, 173*(1-3), 213-225.

Kumar, A., S. K. Ghosh, and V. K. Dadhwal (2006). *Study of mixed kernel effect on classification accuracy using density estimation.* Paper presented at the ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processes", Enschede, the Netherlands.

Laclau, P. (2003). Biomass and carbon sequestration of ponderosa pine plantations and native cypress forests in northwest Patagonia. *Forest Ecology and Management, 180*(1-3), 317-333.

Lamotte, S., J. Gajaseni, and F. Malaisse (1998). Structure diversity in three forest types of north-eastern Thailand (Sakaerat Reserve, Pak Tong Chai). *Biotechnol. Agron. Soc. Environ., 2*(3), 192-202.

Lefsky, M. A. (1997). *Application of Lidar remote sensing to the estimation of forest canopy and stand structure.* Unpublished Dissertation Thesis, University of Virginia.

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

Lefsky, M. A., W. B. Cohen, D. Harding, G. Parker, S. A. Acker, and S. T. Gower (2001). Lidar remote sensing of above-ground biomass in three biomes. *International Archive of Photogrammetry and Remote Sensing, 34*, 22-24.

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

Lefsky, M. A., D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart (2001). An evaluation of alternative remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research, 31*(1), 78-87.

Lefsky, M. A., D. J. Harding, M. Keller, W. B. Cohen, C. C. Carabajal, F. Del Bom Espirito-Santo, et al. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical Research Letters, 32, L22S02, doi:10.1029/2005GL023971.

Lillesand, T. M., R. W. Kiefer, and J. W. Chipman (2004). *Remote sensing and image interpretation* (5th ed.). New York: John Wiley & Sons.

Losi, C. J., T. G. Siccama, R. Condit, and J. E. Morales (2003). Analysis of alternative methods for estimating carbon stock in young tropical plantations. *Forest Ecology and Management, 184*(1-3), 355-368.

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

Markowetz, F. (2003). Classification by Support Vector Machines. Retrieved Jan. 20, 2008, from www.bioconductor.org/workshops/2003/NGFN03/svm.pdf

Means, J. E., S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson, and C. J. Hendrix (2000). Predicting forest stand characteristics with airborne scanning LiDAR. *Photogramm. Eng. Remote Sens. , 66*, 1367-1371.

Means, J. E., S. A. Acker, D. J. Harding, J. B. Blair, M. A. Lefsky, W. B. Cohen, et al. (1999). Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment, 67*(3), 298-308.

Meyer, D. (2007). *Support Vector Machines*: Technische Universit at Wien, Austria.

Mickler, R. A., T. S. Earnhardt, and J. A. Moore (2002). Regional estimation of current and future forest biomass. *Environmental Pollution, 116*(Supplement 1), S7-S16.

NASA. (2007). ICESAT Home Page. Retrieved 20/06/2007, from http://icesat.gsfc.nasa.gov/

NSIDC. (2007a). ICESat/GLAS Data at NSIDC. Retrieved October, 20, 2007, from http://nsidc.org/data/icesat

NSIDC. (2007b). Tools for working with icesat/glas data. Retrieved Oct. 2, 2007, from ICESat/GLAS Data at NSIDC, http://nsidc.org/data/icesat/tools.html

Parresol, and R. Bernard (1999). Assessing Tree and Stand Biomass: A Review with Examples and Critical Comparisons. *Forest Science, 45*(4), 573-593.

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

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

Ranson, K. J., and G. Sun (1994). Mapping biomass of a northern forest using multifrequency SAR data. *IEEE Trans. Geosci. Remote Sens., 32*(2), 388-395.

Ranson, K. J., G. Sun, K. Kovacs, and V. I. Kharuk (2004). Landcover attributes from ICESat GLAS data in central Siberia. *IEEE, 0-7803-8742-20/04 (C)*, 753-756.

Rauste, Y. (2005). Multi-temporal JERS SAR data in boreal forest biomass mapping. *Remote Sensing of Environment, 97*(2), 263-275.

Rosenqvist, A., A. Milne, R. Lucas, M. Imhoff, and C. Dobson (2003). A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science & Policy, 6*(5), 441-455.

Scholkopf, B., and A. Smola (2002). *Learning with Kernels*: MIT Press, Cambridge MA.

Shawe-Taylor, J., and N. Cristianini (2000). *An Introduction to Support Vector Machines (and other kernel-based learning methods)*: Cambridge University Press.

Treitz, P. M., and P. J. Howarth (1999). Hyperspectral remote sensing for estimationg biophysical parameters of forest ecosystems. *Prog. Phys. Geogr., 23*, 359-390.

UNFCCC. (1997). Kyoto Protocol to the United Nation Framework Convention on Climate Change. Retrieved 18/06/2007, from http://unfccc.int/

UNFCCC. (2002). United Nations Framework Convention on Climate Change. Retrieved 18/06/2007, from http://unfccc.int/

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: NY: Springer-Velag.

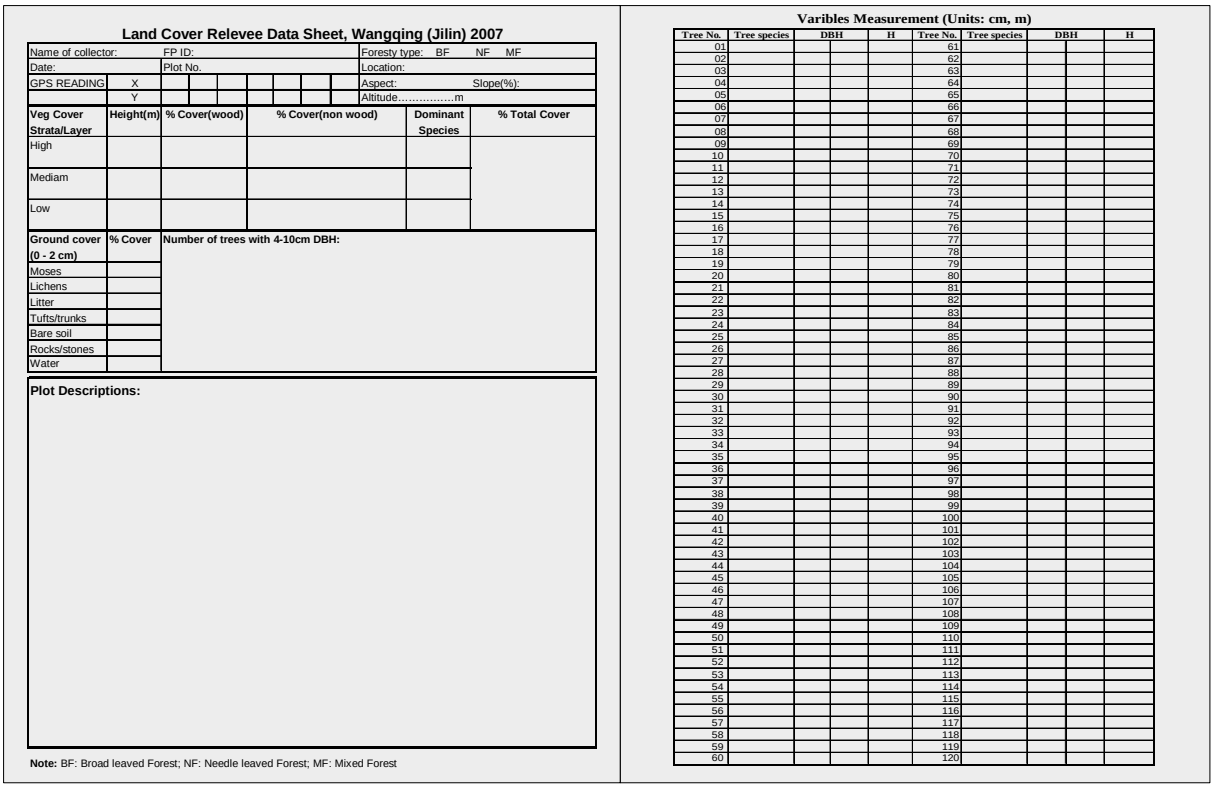
Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing, 60*(2), 100-112.

Weka. (2007). Weka 3: Data Mining with Open Source Machine Learning Software in Java. Retrieved Jan. 20, 2008, from http://www.cs.waikato.ac.nz/~ml/weka/index.html

Xiao, X., S. Boles, J. Liu, D. Zhuang, and M. Liu (2002). Characterization of forest types in Northeeastern China, using multi-temporal SPOT-4 VEGETATION sensor data *Remote Sensing of Environment, 82*(2-3), 335-348.

Yanqiu, X. (2006). Estimating Aboveground Forest Biomass Using Spaceborne Lidar Waveforms in the Cool-temperate Forest of Northeastern China.

Appendix I Data Collection



Land Cover Relevee Data Sheet, Wangqing (Jilin) 2007

Name of collector:		FP ID:					Foresty type: BF NF MF					
Date:		Plot No.					Location:					
GPS READING	X						Aspect: Slope(%):					
	Y						Altitude.....m					

Veg Cover Strata/Layer	Height(m)	% Cover(wood)	% Cover(non wood)	Dominant Species	% Total Cover
High					
Mediam					
Low					

Ground cover (0 - 2 cm)	% Cover	Number of trees with 4-10cm DBH:
Moses		
Lichens		
Litter		
Tufts/trunks		
Bare soil		
Rocks/stones		
Water		

Plot Descriptions:

Note: BF: Broad leaved Forest; NF: Needle leaved Forest; MF: Mixed Forest

Varibles Measurement (Units: cm, m)

Tree No.	Tree species	DBH		H	Tree No.	Tree species	DBH		H
01					61				
02					62				
03					63				
04					64				
05					65				
06					66				
07					67				
08					68				
09					69				
10					70				
11					71				
12					72				
13					73				
14					74				
15					75				
16					76				
17					77				
18					78				
19					79				
20					80				
21					81				
22					82				
23					83				
24					84				
25					85				
26					86				
27					87				
28					88				
29					89				
30					90				
31					91				
32					92				
33					93				
34					94				
35					95				
36					96				
37					97				
38					98				
39					99				
40					100				
41					101				
42					102				
43					103				
44					104				
45					105				
46					106				
47					107				
48					108				
49					109				
50					110				
51					111				
52					112				
53					113				
54					114				
55					115				
56					116				
57					117				
58					118				
59					119				
60					120				

Figure 0-1 Sheet for field data collection

Table 12 Slope correction table

SLOPE CORRECTION TABLE

Slope	Radius	Slope	Radius	Slope	Radius	Slope	Radius	Slope	Radius
0	12.62	*20*	12.74	*40*	13.10	*60*	13.63	*80*	14.28
1	12.62	*21*	12.76	*41*	13.12	*61*	13.66	*81*	14.32
2	12.62	*22*	12.77	*42*	13.14	*62*	13.69	*82*	14.35
3	12.62	*23*	12.78	*43*	13.17	*63*	13.72	*83*	14.39
4	12.63	*24*	12.80	*44*	13.19	*64*	13.75	*84*	14.42
5	12.63	*25*	12.81	*45*	13.22	*65*	13.78	*85*	14.46
6	12.63	*26*	12.83	*46*	13.24	*66*	13.81	*86*	14.49
7	12.64	*27*	12.84	*47*	13.27	*67*	13.85	*87*	14.53
8	12.64	*28*	12.86	*48*	13.29	*68*	13.88	*88*	14.57
9	12.65	*29*	12.88	*49*	13.32	*69*	13.91	*89*	14.60
10	12.65	*30*	12.89	*50*	13.34	*70*	13.94	*90*	14.64
11	12.66	*31*	12.91	*51*	13.37	*71*	13.98	*91*	14.67
12	12.67	*32*	12.93	*52*	13.40	*72*	14.01	*92*	14.71
13	12.67	*33*	12.95	*53*	13.43	*73*	14.04	*93*	14.75
14	12.68	*34*	12.97	*54*	13.45	*74*	14.08	*94*	14.78

15	12.69	*35*	12.99	*55*	13.48	*75*	14.11	*95*	14.82
16	12.70	*36*	13.01	*56*	13.51	*76*	14.14	*96*	14.86
17	12.71	*37*	13.03	*57*	13.54	*77*	14.18	*97*	14.90
18	12.72	*38*	13.05	*58*	13.57	*78*	14.21	*98*	14.93
19	12.73	*39*	13.07	*59*	13.60	*79*	14.25	*99*	14.97
Unit: Slope--percent; Radius--meter;								*100*	15.01

Table 13 Main part of the primary field data used in the study

Plot NO.	Date	Index	Shot	GLAS_X	GLAS_Y	DEM_slope	Forest_type*	Dominant Species	*Slope Range***
1	18/09/07	1217345092	29	617837	4802355	0.42	A		
2	18/09/07	1217345092	28	617865	4802184	1.34	A		
3	18/09/07	1217345092	27	617892	4802013	2.50	A		
4	18/09/07	1217345092	26	617920	4801841	18.18	M	*Larix ; Quercus*	
5	18/09/07	1217345092	21	618060	4800985	13.92	A		
6	18/09/07	1217345092	30	617809	4802526	42.14	B	*Quercus*	
7	20/09/07	1217345102	28	616749	4809049	3.86	A		
8	20/09/07	1217345102	29	616722	4809220	51.63	M	*Larix ; Pinus sylvestrus*	
9	20/09/07	1217345102	30	616694	4809391	49.96	B	*Quercus*	
10	20/09/07	1217345102	31	616667	4809563	38.99	M	*Quercus, Larix, Betula*	
11	20/09/07	1217345102	32	616639	4809735	15.25	B	*Quercus*	
12	21/09/07	1217345112	2	616361	4811454	6.73	N	*larix*	
13	21/09/07	1217345112	1	616389	4811282	18.66	B		*15-20*
14	21/09/07	1217345102	40	616417	4811110	27.28	M	*Pinus koraiensis*	
15	21/09/07	1217345102	39	616445	4810938	27.15	M	*Acer, Tilia*	
16	21/09/07	1217345102	38	616473	4810766	40.00	B	*Quercus*	
17	21/09/07	1217345102	37	616501	4810594	36.25	B	*Quercus*	
18	21/09/07	1217345102	36	616528	4810422	45.27	B	*Quercus*	
19	21/09/07	1217345102	35	616556	4810250	2.91	A		
20	21/09/07	1217345102	34	616584	4810078	10.90	N	*Larix*	
21	21/09/07	1217345102	33	616611	4809906	13.69	B	*Larix; Betula*	*10-15*
22	22/09/07	1217345112	12	616079	4813171	22.60	B	*Quercus*	
23	22/09/07	1217345112	11	616107	4813000	15.41	B	*Quercus; Larix ; populus*	*15-20*
24	22/09/07	1217345112	10	616135	4812828	41.84	B	*Quercus, Larix*	
25	22/09/07	1217345112	9	616163	4812657	36.77	M	*larix, quercus*	
26	22/09/07	1217345112	8	616191	4812485	36.83	M	*Larix ; Quercus*	
27	22/09/07	1217345112	7	616219	4812313	38.62	B	*quercus*	
28	22/09/07	1217345112	5	616276	4811970	30.15	B	*Betula ; Larix*	
29	22/09/07	1217345112	4	616304	4811797	37.63	B	*Quercus*	
30	23/09/07	1217345082	29	618953	4795489	8.44	B	*Betula*	*5-10*
31	23/09/07	1217345082	28	618981	4795318	29.62	B	*very diverse*	
32	23/09/07	1217345082	27	619010	4795146	21.34	B	*Betula*	*20-25*
33	23/09/07	1217345082	26	619038	4794975	24.20	B	*Betula ; Populus*	*20-25*
34	23/09/07	1217345082	25	619066	4794803	22.28	B	*Betula*	*20-25*
35	23/09/07	1217345082	24	619094	4794632	21.69	B	*Betula*	*20-25*
36	23/09/07	1217345082	23	619122	4794460	12.50	B	*Betula*	
37	23/09/07	1217345082	21	619179	4794117	33.50	B	*Betula*	

38	23/09/07	1217345082	20	619207	4793945	42.46	B	*Populus*	
39	23/09/07	1217345082	19	619236	4793773	47.33	B	*Populus*	
40	24/09/07	1217345082	9	619516	4792054	48.09	B	*Fraxinus. Tilia*	
41	24/09/07	1217345082	8	619544	4791882	47.46	B	*Populus,betula (BH)*	
42	24/09/07	1217345082	7	619572	4791710	56.54	B	*Betula ; Populus*	
43	24/09/07	1217345082	6	619599	4791538	41.85	B	*Betula*	
44	24/09/07	1217345082	5	619627	4791366	26.50	B	*Betula*	
45	24/09/07	1217345082	4	619654	4791194	33.53	B	*Betula , Populus*	
46	24/09/07	1217345082	3	619682	4791022	28.36	B	*Populus*	
47	24/09/07	1217345082	2	619709	4790850	22.27	B	*Tilia*	*20-25*
48	24/09/07	1217345082	1	619736	4790679	25.87	B	*Populus ; Betula*	
49	26/09/07	1217345072	26	620156	4788108	26.83	B	*Populus*	
50	26/09/07	1217345072	27	620128	4788280	23.07	M	*Betula ; Picea*	
51	26/09/07	1217345072	28	620099	4788451	27.11	M	*Abies ; Betula*	
52	26/09/07	1217345072	29	620071	4788623	37.34	M	*betula ; Picea*	
53	26/09/07	1217345072	30	620042	4788795	42.01	B	*Betula*	
54	26/09/07	1217345072	31	620014	4788966	40.57	N	*Picea ; Abies*	
55	26/09/07	1217345072	32	619986	4789137	20.13	M	*Betula ; Picea*	
56	26/09/07	1217345072	33	619958	4789309	35.13	M	*Betula ; Picea*	
57	26/09/07	1217345072	34	619930	4789480	31.89	B	*Betula*	
58	26/09/07	1217345072	35	619902	4789651	7.71	B	*Betula*	*5-10*
59	27/09/07	1217345072	25	620184	4787936	44.75	N	*Abies ; Picea*	
60	27/09/07	1217345072	24	620213	4787765	27.13	M	*Picea ; Betula*	
61	27/09/07	1217345072	23	620241	4787592	24.07	N	*Picea*	*20-25*
62	27/09/07	1217345072	22	620269	4787420	15.89	N	*Picea*	*15-20*
63	27/09/07	1217345072	21	620298	4787248	10.85	N	*Abies*	*10-15*
64	27/09/07	1217345072	20	620326	4787076	5.59	N	*Abies, Picea*	*5-10*
65	28/09/07	1221245452	30	636720	4819416	4.64	B	*Betula ; Shwitunga*	*0-5*
66	28/09/07	1221245452	31	636698	4819244	16.62	M	*Populus ; Larix*	*15-20*
67	28/09/07	1221245452	32	636677	4819071	25.09	B	*Populus*	
68	28/09/07	1221245452	33	636655	4818899	22.14	B	*Populus*	*20-25*
69	28/09/07	1221245452	34	636634	4818727	18.01	B	*Populus*	*15-20*
70	28/09/07	1221245452	35	636612	4818555	15.48	M	*larix ; Betula*	*15-20*
71	28/09/07	1221245452	36	636591	4818383	9.88	N	*Larix*	
72	28/09/07	1221245452	37	636570	4818210	8.50	M	*Larix ; Betula*	*5-10*
73	28/09/07	1221245452	38	636549	4818038	11.18	N	*Larix*	*10-15*
74	28/09/07	1221245452	39	636528	4817866	10.00	N	*Larix*	
75	28/09/07	1221245452	40	636507	4817694	11.18	BL		
76	29/09/07	1221245452	29	636741	4819589	34.85	BL		
77	29/09/07	1221245452	28	636763	4819762	22.72	B	*Quercus*	
78	29/09/07	1221245452	27	636785	4819934	17.83	B	*Quercus*	*15-20*
79	29/09/07	1221245452	26	636807	4820107	16.55	B	*Quercus ; betula*	*15-20*
80	29/09/07	1221245452	25	636829	4820279	22.73	B	*Quercus*	*20-25*
81	29/09/07	1221245452	24	636852	4820451	18.33	B	*Quercus, Acer*	*15-20*
82	29/09/07	1221245452	23	636874	4820623	18.00	B	*Quercus*	*15-20*
83	29/09/07	1221245452	22	636896	4820795	10.06	B	*Quercus*	*10-15*

84	29/09/07	1221245452	21	636917	4820968	9.00	B	*Quercus*	*5-10*
85	29/09/07	1221245452	20	636939	4821140	17.23	B	*Populus*	*15-20*
86	29/09/07	1221245452	19	636961	4821312	12.90	B	*Populus*	*10-15*
87	29/09/07	1221245452	18	636983	4821484	14.87	B	*Populus*	*10-15*
88	29/09/07	1221245452	17	637005	4821657	13.69	B	*Populus ; Quercus*	*10-15*
89	30/09/07	1221245452	13	637091	4822348	8.90	B	*Quercus ; Populus*	*5-10*
90	30/09/07	1221245452	12	637112	4822521	8.67	B	*Betula ; Quercus*	
91	30/09/07	1221245452	11	637133	4822693	8.14	M	*Betula ; larix*	*5-10*
92	30/09/07	1221245452	10	637154	4822866	10.81	N	*Larix*	*10-15*
93	30/09/07	1221245452	9	637176	4823038	10.98	B	*Quercus*	
94	30/09/07	1221245452	8	637197	4823211	8.25	B	*Populus*	*5-10*
95	30/09/07	1221245452	7	637218	4823383	8.53	M	*Populus ; larix*	*5-10*
96	30/09/07	1221245452	6	637239	4823556	6.82	B	*Quercus*	*5-10*
97	30/09/07	1221245452	5	637260	4823728	6.90	B	*Populus ; Betula*	*5-10*
98	30/09/07	1221245452	4	637281	4823900	10.60	B	*Quercus*	*10-15*
99	01/10/07	1221245502	32	632375	4784590	7.00	N	*Picea ; Abies*	*5-10*
100	01/10/07	1221245502	34	632332	4784245	8.74	N	*Abies*	*5-10*
101	01/10/07	1221245502	35	632311	4784073	12.63	N	*Picea*	*10-15*
102	01/10/07	1221245502	30	632417	4784935	16.69	N	*Abies ; Picea*	*15-20*
103	01/10/07	1221245502	27	632481	4785454	5.63	M	*Picea ; Populus*	

NOTE: *A--Agriculture land; B--Broadleaved forest; N--Needleleaved forest; M--Mixed forest; BL--Bare land;

**Only slope range of the waveforms selected validate for analyisis are displayed here.

Appendix II Softwares

A. Software used in the study

Table 14 Software used in the study

Software	Function
Microsoft Word, Visio	Word processing
Microsoft Excel	Spreadsheets preparation
ArcGIS	Map preparation, Map calculation and Data Visualization
Matlab, IDL	Waveform Pre-processing
Microsoft Visual Studio. Net	Program developing to derive waveform metrics
STATISTICA	Statistical Analysis
Weka	SVM classification of waveform metrics derived

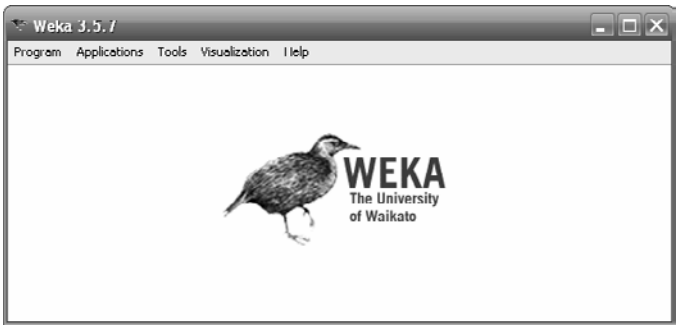
B. Brief Guide of Using Weka to classify the Derived Lidar Metrics

This guide is just a brief introduction of the process of using the Explorer Application in Weka to apply Support Vector Machine (SVM) classification to the lidar waveform metrics produced, so as to examine their relationship with different forest types. It does not explain the individual data preprocessing tools and learning algorithms in Weka. More information about Weka is available on the WekaWiki (http://weka.sourceforge.net/wiki/).

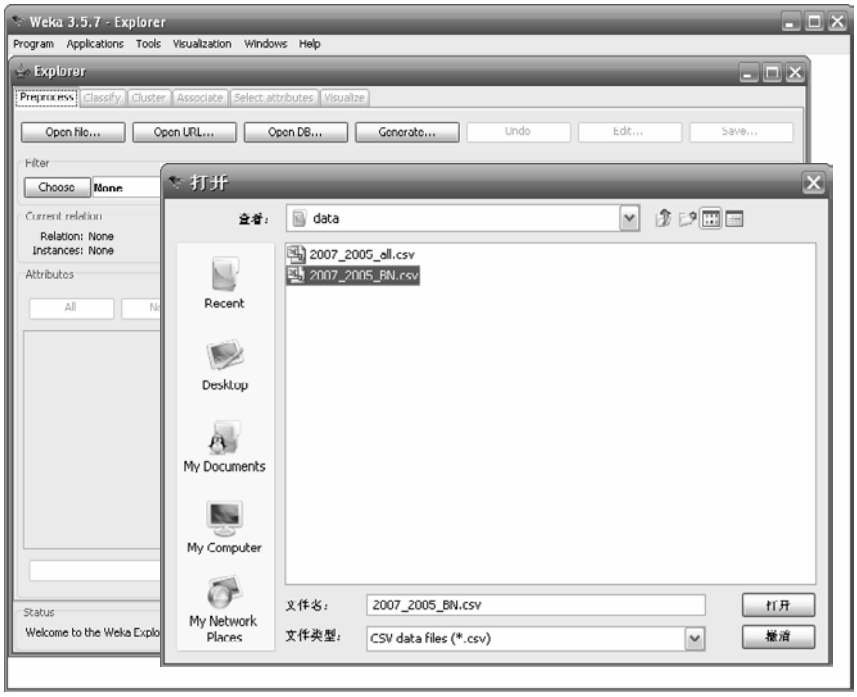
1. Launching Weka and loading data

1) Download and install the latest version of Weka from the website: http://sourceforge.net/project/showfiles.php?group_id=5091

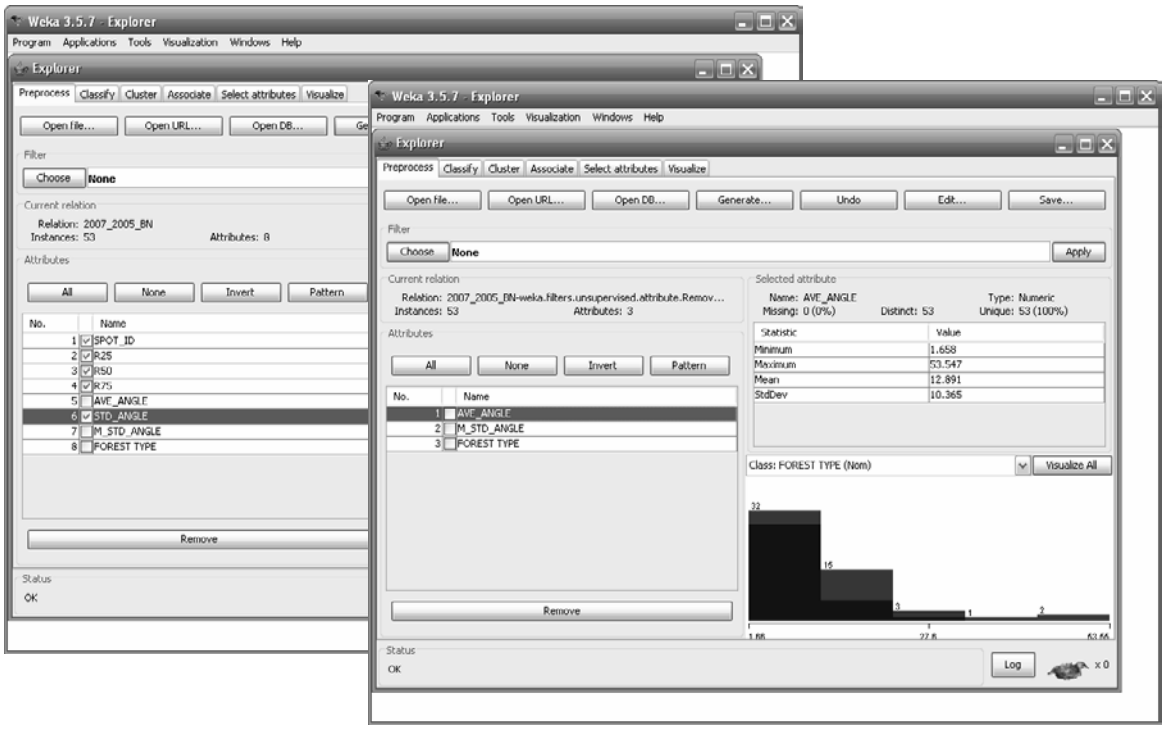
2) Launch Weka and select in the menu “Application” -> “Explorer”;



3) Click on “**Open file…**” button, and select the prepared waveform metric file in CSV format “*2007_2005_BN.csv*”, which contains the metrics records of 35 broadleaved forest and 18 needleleaved forest;

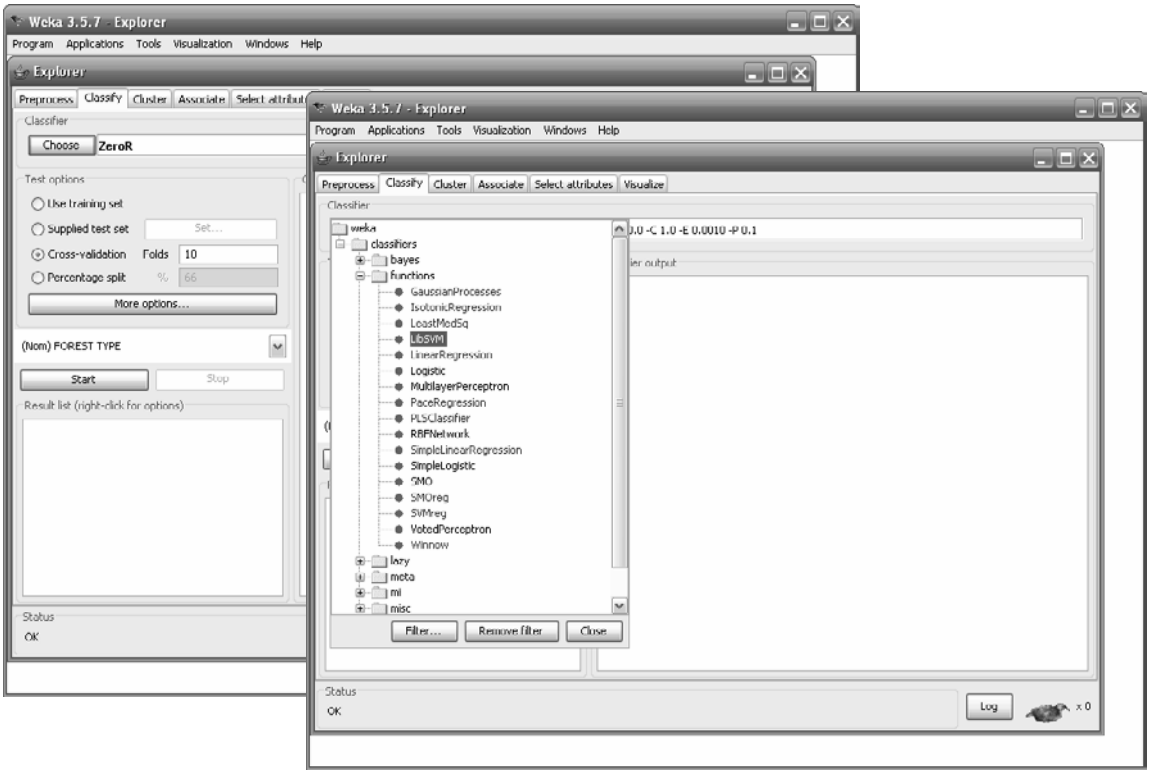


4) In this case, the relationship between metrics combination of AGS and MSGS (as *AVE_ANGLE* and *M_STD_ANGLE* shown in the graph respectively) and different forest types are to be examined. In the Attributes box, toggle on the tick boxes of the attributes are not needed, leaving only *AVE_ANGLE*, *M_STD_ANGLE* and *FOREST TYPE*, and click on "**Remove**" button. Then only the remaining metrics AGS and MSGS will be used as input parameter in the SVM classification in the next step;

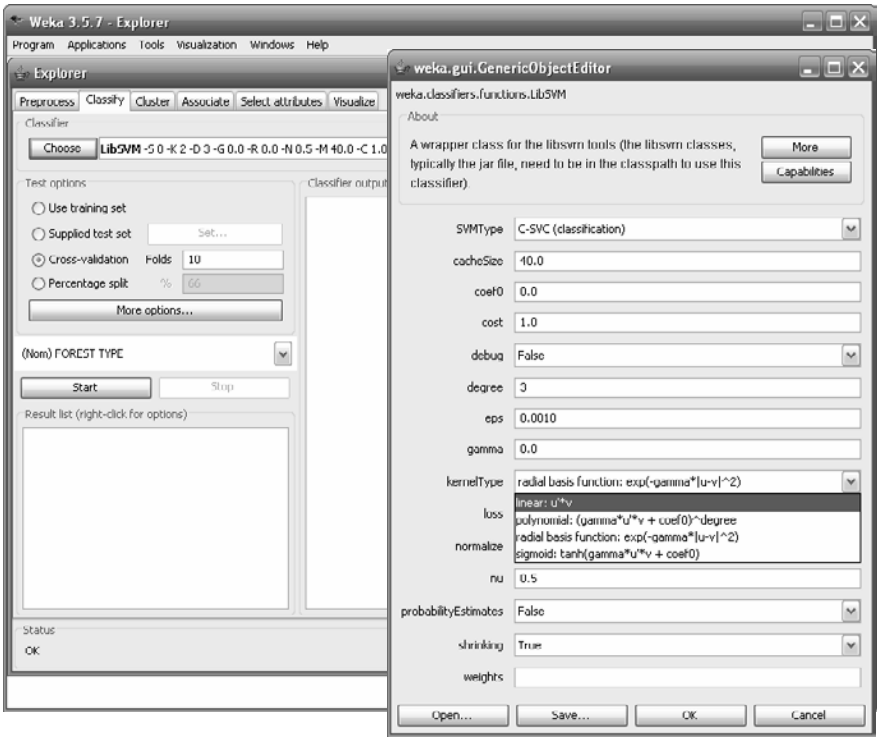


2. SVM classification of the metrics

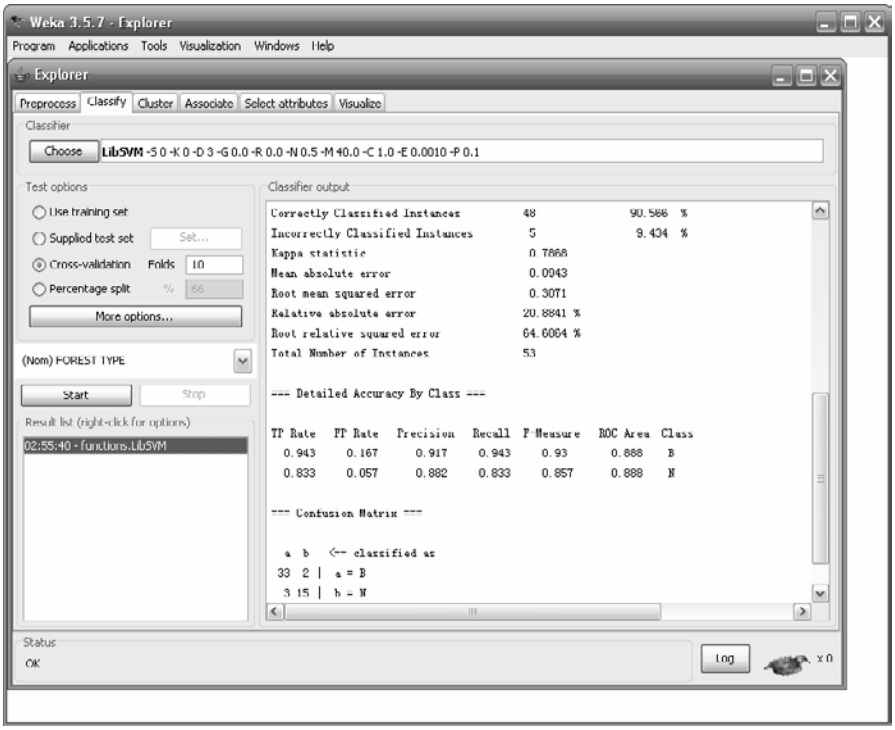
1) Click on "**Choose**" button to select a classifier, in our case, "*LibSVM*" classifier is selected. Then the name and initiate parameters of the currently selected classifier will appear in the text field of the **Classifier box**;



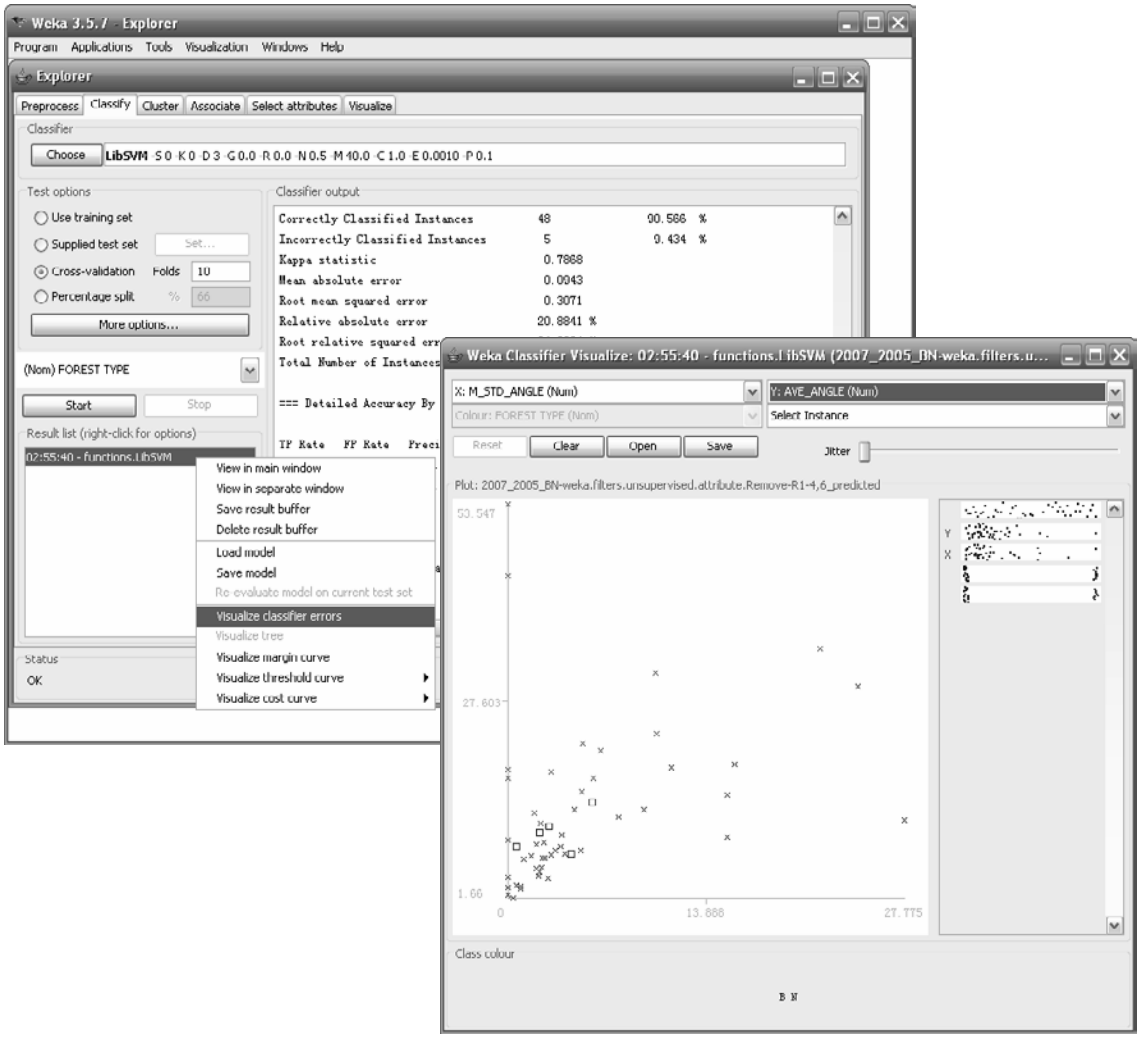
2) Click on the text box with the left mouse button to bring up the **GenericObjectEditor** dialog box. Then we can configure the options of the current classifier. In our case, we only change the **kernelType** as "*linear:u'*v*", and leave other value as default;



3) Leave the Test options as "*Cross-validation, Folds 10*"; Click on "**Start**" button to begin the classification and validation process. A moment later, classification result, such as "*Detail Accuracy By Class*" and "*Confusion Matrix*" can be obtained form the **Classifier output box**; As we can see in the graph, an overall accuracy of 90.566% and Kappa Statistic of 0.7868 were produced;



4) A visualization of the result can be get by right click on the result shown in the **Result list box**;



Appendix III Results

Table 15 Metrics derived from all waveforms used in the study

Dataset	SPOT_ID	R25	R50	R75	AGS*5E-5	SGS*5E-5	MSGS*5E-5	FOREST TYPE
First-phase data	1217345072:35	0.288959676	0.507441382	0.718875292	10.99067274	2.369458478	2.889337497	B
	1217345082:02	0.395572659	0.593358988	0.698432976	4.212606651	2.738648869	2.822671432	B
	1217345082:24	0.290492503	0.518736613	0.726231259	3.232321839	1.251577667	0.638835442	B
	1217345082:25	0.36023678	0.586147981	0.738790684	5.046910735	2.573344937	2.224879205	B
	1217345082:26	0.446528117	0.704140492	0.807185443	15.1962489	17.50871357	15.37442495	B
	1217345082:27	0.311479997	0.523852723	0.707909085	19.25008052	21.61984398	15.91360243	B
	1217345082:29	0.219220217	0.473054152	0.784577618	9.958856999	3.795149557	3.795149557	B
	1217345102:33	0.594251127	0.772526465	0.842756144	9.697007347	11.45149911	15.42125022	B
	1217345112:01	0.551772335	0.698911624	0.787195197	13.18247287	9.541229334	9.541229334	B
	1217345112:11	0.345817462	0.514508907	0.666331208	7.802646466	3.458531465	3.339440112	B
	1221245452:04	0.248523508	0.506605612	0.669101752	12.24469129	9.416340829	7.815434122	B
	1221245452:05	0.262433499	0.541798191	0.90581885	5.42710312	2.005976365	2.005976365	B
	1221245452:06	0.236277986	0.60053988	0.836817866	7.095613649	1.670256401	1.670256401	B
	1221245452:08	0.214713015	0.459041618	0.85885206	6.922015441	2.440526344	2.637323065	B
	1221245452:13	0.262145754	0.63078822	0.8519737	5.675412612	4.00219634	2.428431241	B
	1221245452:17	0.25834265	0.516685301	0.783101159	6.701583821	1.222806175	1.147619951	B
	1221245452:18	0.32606162	0.53619022	0.789793703	2.982973646	0	0	B
	1221245452:19	0.348446685	0.689634065	0.849338796	8.693760695	2.163173249	2.001830736	B
	1221245452:20	0.304562013	0.622967754	0.823701808	7.358339422	3.629250793	4.053791977	B
	1221245452:21	0.3776743	0.635598212	0.81061801	1.960307343	0	0	B
	1221245452:22	0.273617302	0.628306397	0.729646139	6.802798509	2.48617384	2.48617384	B
	1221245452:23	0.311292661	0.499954879	0.707483319	7.874707982	9.210924962	5.110428808	B
	1221245452:24	0.428336316	0.696046513	0.82607718	1.658447072	0.485781368	0.400167125	B
	1221245452:25	0.350640574	0.552933213	0.734996588	3.122331315	1.4067764	0.895597737	B
	1221245452:26	0.256582048	0.466512814	0.730870076	10.21777279	2.381730632	2.223335062	B
	1221245452:27	0.686130864	0.816259131	0.887238186	11.89228838	15.13061129	27.77505385	B
	1221245452:30	0.486157761	0.688723494	0.820391221	4.489337445	2.194804723	2.194804723	B
	1221245452:33	0.302114052	0.59109271	0.801259006	13.18084804	4.666780456	4.666780456	B
	1221245452:34	0.389867284	0.621931143	0.779734567	8.345477414	3.758550887	3.758550887	B
	1217345072:20	0.23368839	0.314270593	0.402911017	53.54705054	0	0	N
	1217345072:21	0.283599662	0.501753249	0.778081125	34.45971977	22.09986644	21.90864684	N
	1217345072:22	0.351303709	0.507438691	0.692848981	31.25247292	10.79051226	10.34309685	N
	1217345072:23	0.405970498	0.592317612	0.772009472	14.27121296	6.750906884	5.944202	N
	1221245452:10	0.258891423	0.420698563	0.663409272	11.48212624	3.811824337	2.303872601	N
	1221245452:38	0.285909804	0.571819609	0.791750228	8.348432934	0.985313034	0.649753712	N
	1221245502:30	0.311878087	0.511799937	0.735712409	21.06217889	6.474697364	6.514760176	N

	1221245502:32	0.287065794	0.414650592	0.566157539	22.02701424	4.465007667	5.314263491	N
	1221245502:34	0.323707753	0.462439647	0.601171541	12.81564164	1.895976254	1.895976254	N
	1221245502:35	0.345123875	0.517685813	0.678743621	23.18301739	11.3251505	10.42741775	N
	1221245452:07	0.228041708	0.367809207	0.89009828	6.882931375	0	0	M
	1221245452:11	0.36469878	0.644042527	0.907867177	3.916292596	0.400427312	0.400427312	M
	1221245452:31	0.420481072	0.592496055	0.745398263	3.726224211	0.416752517	0.416752517	M
	1221245452:35	0.468597779	0.744243532	0.840719546	9.570022321	6.159308128	8.147275581	M
	1221245452:37	0.381454715	0.689079485	0.830586879	7.011478137	3.776424812	2.833242973	M
Second-phase data	263028083:25	0.040257287	0.120771861	0.161029148	8.874359011	2.562289603	2.562289603	B
	898951817:18	0.541184097	0.702917275	0.814886398	18.82050585	12.40095443	11.45707319	B
	898951817:27	0.308951823	0.475310497	0.633747329	4.320527918	0	0	B
	898951847:03	0.401740646	0.626715408	0.779376853	29.54178821	26.46667341	24.53977785	B
	898951847:10	0.336741546	0.498377488	0.633074107	2.846666687	0.89740654	0.89740654	B
	902852257:01	0.367702188	0.543892819	0.712422989	7.218349281	4.196263908	3.100654659	B
	898951817:16	0.404509639	0.543559827	0.695250942	44.09826394	0	0	N
	902852247:35	0.284087207	0.403702873	0.530794518	18.57340949	0	0	N
	902852247:36	0.351615832	0.480780832	0.609945831	17.39152312	0	0	N
	902852247:37	0.334687115	0.514275323	0.66121113	18.17677964	3.031656809	3.031656809	N
	902852247:38	0.375071585	0.526542803	0.670801105	9.211816833	0	0	N
	902852257:02	0.354613782	0.531920673	0.691496875	17.39818238	5.462299961	6.00468424	N
	902852257:03	0.345055135	0.495624649	0.633646703	7.416474942	3.876105138	4.50140375	N
	902852257:04	0.378483137	0.53296605	0.672000671	15.64781592	5.051469521	5.24246925	N
	898951817:19	0.329487152	0.461282013	0.593076874	13.70344574	0	0	M
	898951817:20	0.38695352	0.555805965	0.682445299	18.1066889	10.30999007	7.410920384	M
	898951817:21	0.37100331	0.513089684	0.64728237	15.62535497	0	0	M
	898951817:22	0.401529822	0.590485033	0.755820842	34.14792605	32.583067	24.8891153	M
	898951847:05	0.324479911	0.530399854	0.792479782	4.872914843	2.363627878	2.446884995	M
	902852247:40	0.360520442	0.530766207	0.721040885	10.80138324	0	0	M

Table 16 Confusion matrix of SVM classification of broadleaved and needleleaved forest using AGS and SGS as input parameters

	Reference data		
Classified data	Broadleaved	Needleleaved	**Row Total**
Broadleaved	34	4	38
Needleleaved	1	14	15
Column Total	35	18	53

Table 17 Classification results of broadleaved and needleleaved forest using AGS and SGS as input parameters

Class name	**r**	**Classified total**	**Number correct**	**Accuracy (%)**		**Kappa**
				Producer	**User**	
Broadleaved	35	38	34	97.14%	89.47%	
Needleleaved	18	15	14	77.78%	93.33%	
Total	53	53	48			

Overall Classification Accuracy =	90.57%		0.7808

Table 18 Confusion matrix of SVM classification of broadleaved and needleleaved forest using AGS, SGS and MSGS as input parameters

	Reference data		
Classified data	Broadleaved	Needleleaved	**Row Total**
Broadleaved	33	4	37
Needleleaved	2	14	16
Column Total	35	18	53

Table 19 Classification results of broadleaved and needleleaved forest using AGS, SGS and MSGS as input parameters

Class name	**r**	**Classified total**	**Number correct**	**Accuracy (%)**		**Kappa**
				Producer	**User**	
Broadleaved	35	37	33	94.29%	89.19%	
Needleleaved	18	16	14	77.78%	87.50%	
Total	53	53	47			
Overall Classification Accuracy =			88.68%			0.7406

Table 20 Confusion matrix of SVM classification of broadleaved and needleleaved forest using all metrics (R25, R50, R75, AGS, SGS and MSGS) as input parameters

	Reference data		
Classified data	Broadleaved	Needleleaved	**Row Total**
Broadleaved	33	4	37
Needleleaved	2	14	16
Column Total	35	18	53

Table 21 Classification results of broadleaved and needleleaved forest using all metrics (R25, R50, R75, AGS, SGS and MSGS) as input parameters

Class name	**r**	**Classified total**	**Number correct**	**Accuracy (%)**		**Kappa**
				Producer	**User**	
Broadleaved	35	37	33	94.29%	89.19%	
Needleleaved	18	16	14	77.78%	87.50%	
Total	53	53	47			
Overall Classification Accuracy =			88.68%			0.7406