Full Waveform-based Analysis for Forest Type Information Derivation from Large Footprint Spaceborne Lidar Data

Junjie Zhang, Alfred de Gier, Yanqiu Xing, and Gunho Sohn

Abstract

This study developed a new method to derive forest type information from large-footprint lidar data based on full waveform analysis. For this purpose, the raw waveform was decomposed into Gaussian components, and canopy return and ground return of the waveforms were separated. Two types of metrics hypothesized to have relationship with forest types were derived from the canopy return part of the waveform. The first type of metrics is quantile-based metrics reflecting the vertical distribution of canopy return energy, and the second type is statistical characteristics of the Gaussian components of canopy return part. Support Vector Machine classification was applied to different combinations of the metrics to find their relationship with different forest types. The results showed that the second type of metrics, indicating the canopy stratum characteristics, showed great promise in separating broad-leaved and needle-leaved forests with the accuracy ranging from 88.68 percent to 90.57 percent and Kappa statistic from 0.7406 to 0.7868.

Introduction

Forest ecosystems play an important role in global carbon budgets because they are recognized as one of the largest reserves of terrestrial carbon (IPCC, 2001). They can influence climate change by affecting the amount of carbon dioxide in the atmosphere through nature (e.g., photosynthesis and respiration of plants) and anthropogenic (e.g., deforestation and afforestation) processes (Brown *et al.*, 1996; Dixon *et al.*, 1994; Patenaude *et al.*, 2005). Generally, Carbon sequestration in forest is studied through the assessment of forest aboveground biomass as it accounts for the majority of total biomass in forest ecosystems and carbon is approximately 50 percent of all dry wood weight (Drake *et al.*, 2002; de Gier, 2003). The accurate estimation the forest aboveground biomass at broad spatial scale is of increasing importance to gain a better understanding of the global carbon cycle (Muukkonen and Heiskanen, 2007).

Light Detection and Ranging (lidar) has shown great potential in providing more accurate forest aboveground biomass estimation over conventional optical remote sensing and radar remote sensing (Drake *et al.*, 2002; Patenaude *et al.*, 2004). The large-footprint lidar fully records the returning signals from the illuminated areas, which in the forests provide enhanced vertical architecture information of the forest canopy (Dubayah and Drake, 2000; Drake *et al.*, 2002; Patenaude *et al.*, 2004). The full waveform profile offers large-footprint lidar the unique capability of direct extraction of forest vertical information such as canopy height and canopy height profile, which in turn are used to estimate other canopy structure characteristics such as basal area, quadratic mean stem diameter (QMSD) and aboveground biomass (AGBM) (Drake *et al.*, 2002; Lefsky *et al.*, 1999, 2002 and 2005; Means *et al.*, 2000; Sun *et al.*, 2008; Xing *et al.*, 2010).

However, for spatially consistent biomass assessment, horizontal canopy structure (i.e., forest types) information is also essential (Patenaude *et al.*, 2004; Lefsky *et al.*, 2005; Patenaude *et al.*, 2005; Gwenzi, 2008), due to the fact that when studies are conducted in heterogeneous forest areas or forests of inconsistent horizontal spatial pattern, the forest biomass may vary considerably between different forest types, even when those forests are of similar canopy height, tree density, and age.

Nevertheless, there have been no studies that tried to develop an effective way to derive forest type information from large-footprint lidar. Ranson *et al.* (2004) first tried to examine the relationships between lidar waveform metrics and different forest types. Two types of metrics, front slope angle and centroid, were extracted from ICE-GLAS lidar waveforms on terrain with slope less than 5 degrees and grouped in eight categories according to land-cover types in central Siberia. However, the two types of metrics showed great overlapping between different forest types and couldn't separate them. Another study (Duong *et al.*, 2006b), conducting land-cover 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.

Junjie Zhang and Gunho Sohn are with the GeoICT Lab, Department of Earth and Space Science and Engineering, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada.

Junjie Zhang and Alfred de Gier are Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Hengelosestraat 99, 7500 AA Enschede, The Netherlands, (zhang18098@itc.nl).

Yanqiu Xing is with the Center for Forest Operations and Environment, Northeast Forestry University, No. 26 Hexing Road, 150040 Harbin, China.

Photogrammetric Engineering & Remote Sensing
Vol. 77, No. 3, March 2011, pp. 281–290.

0099-1112/11/7703–0281/$3.00/0

Research Problem

Figure 1 shows a typical lidar waveform profile and some common waveform-derived metrics. Drake *et al.* (2002) extracted lidar canopy height (LHT), the height of median energy (HOME), the height/median ratio (HTRT), and ground-return ratio (GRND) and used them in a stepwise regression to predict field measured basal area, QMSD and field estimated AGBM. Ranson *et al.* (2004) proposed metrics front slope angle and tried to examine the relationship between this metrics and forest stand characteristics in Central Siberia area.

However, the five existing large-footprint lidar waveform-derived metrics (Figure 1): (a) *Lidar Canopy Height (LHT)*, (b) *the height of median energy (HOME)*, (c) *the Height/Median Ratio (HTRT)*, (d) *Ground Return Ratio (GRND)*, and (e) *Front Slope Angle*, which are commonly used to describe vertical canopy structure characteristics (Brenner *et al.*, 2003; Drake *et al.*, 2002; Ranson *et al.*, 2004), are all incapable of distinguishing different forest types, because even for the same forest type, those metrics can vary a lot with the influence of factors such as canopy height, tree density, sub-canopy structure, and slope.

- ***Lidar Canopy Height (LHT)*** is calculated as the distance between the signal start and the center of the last Gaussian pulse. This metric can be regarded as a direct measure of canopy height within the waveform footprint, especially on flat terrain; but, it can be affected by slope.
- ***Height of Median Energy (HOME)*** is the height at which half of the return energy of the waveform is above and half below. It is not only affected by the tree height and the vertical canopy structure distribution, but may also sensitive to tree density and ground cover type.
- ***HOME/LHT Ratio (HTRT)*** is the HOME divided by canopy height. Compared with HOME, HTRT normalizes tree height. However, it is still influenced by some other factors affecting HOME, such as vertical canopy structure distribution and tree density, and is therefore not an ideal metric for distinguishing forest types either.
- ***Ground Return Ratio (GRND)*** is the ratio of ground return energy and the total energy contained in the waveform. GRND only provides an approximation of the degree of canopy closure with certain canopy assumption. It may also be influenced by ground cover type. As a result, it is also unsuitable for use in deriving forest type information.
- ***The Front Angle*** is the angle between the vector from the signal start to the first peak of canopy return energy and the vertical direction. To a large extend, this metrics is determined by both vertical structure distribution and canopy density of the top-most canopy. Consequently, it is not possible to derive forest type information from this single metric.

For good indicators of different forest types, the metrics should be able to distinguish between different forest types, while remaining insensitive to other forest structure characteristics and terrain factors. In order to derive forest type information from large-footprint lidar waveforms, new metrics that can reflect the different vertical canopy structures between different forest types without much bias of other factors, such as forest height, density, and slope have to be found.

For this purpose, it is intuitive to focus more on the canopy return part of the waveforms. As different forest types are expected to have different vertical distributions of canopy structures (Lamotte *et al.*, 1998), the differences are expected to be reflected in the canopy return energy and result in typical waveform characteristics in canopy return part. So in our proposed method, the canopy return part and ground return part will be first separated from the large-footprint lidar waveform. Then, two types of metrics hypothesized to have relationship with forest types will be extracted from the canopy return part of the waveforms.

The first group of metrics, R25, R50, and R75, which will be tried to derive from the lidar waveform, are some quantile-based metrics based on the vertical distribution of canopy return energy as shown in Figure 2. 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 percent quartile height calculated by subtracting the elevation at which 25 percent of the return energy occurs from the ground elevation (Kimes *et al.*, 2006). In our study, CH25, CH50, CH75 are proposed as heights of 25 percent, 50 percent, and 75 percent canopy return energy relative to signal start and distributed over the waveform interval between single start and the ground return. 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

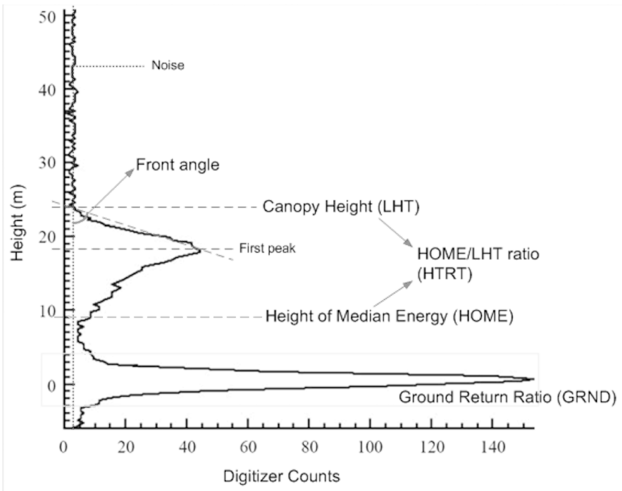

Figure 1. Typical lidar waveform profile and some common metrics derived (Drake *et al.*, 2002; Ranson *et al.*, 2004).

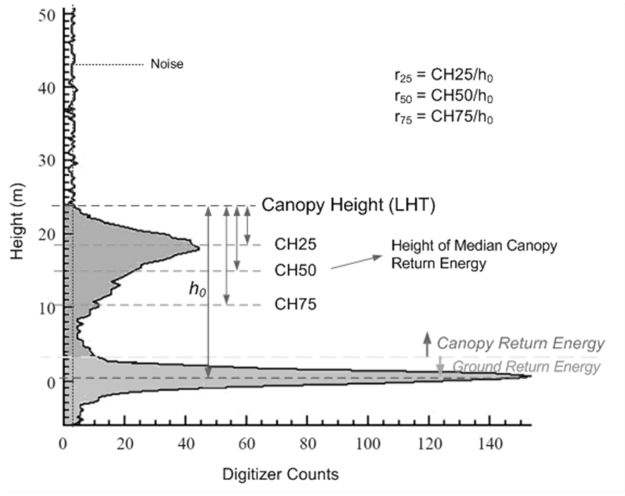

Figure 2. Proposed metrics R25, R50, and R75 to be derived from lidar waveform.

height to normalize the influence of different canopy heights and the waveform spreading effect caused by topographical slope. In this way, metrics R25, R50, and R75 can reflect the vertical canopy structure distribution of the forest, which is assumed to be different among different forest types.

In addition, the returned waveform can be regarded mathematically as a composition of several Gaussian curves and reflect Gaussian surfaces (Brenner *et al.*, 2003; Wagner *et al.*, 2006), in the scene of forests, the layers of branches and foliage. Different tree species, especially broad-leaved and needle-leaved trees usually have distinct vertical stratum characteristics, which are supposed to be reflected by the characteristics of the decomposed Gaussian curves, i.e., the Gaussian curve slope, which is calculated by dividing the amplitude of the Gaussian component by its standard deviation. When compared with broad-leaved trees, the needle-leaved trees usually have more obvious and regular canopy stratum patterns. It is then hypothesized that the average of the decomposed Gaussian curve slopes of the canopy return of waveforms of needle-leaved forests will be larger than that of broad-leaved trees, while the standard deviation of the decomposed Gaussian curve slopes of the canopy return of waveforms of needle-leaved forests smaller than that of broad-leaved forests. Therefore, the second group of metrics, the average of the decomposed Gaussian curve slope (AGS), the standard deviation of the decomposed Gaussian curve slope (SGS) and the Modified SGS (MSGS), which are statistical characteristics of decomposed Gaussian curves in the canopy return part of the waveform, are proposed and will be extracted from the decomposed waveforms as shown in Figure 3. After these two types of metrics are extracted from the lidar waveforms, their relationship with different forest types will be further explored.

In this way, the presented research aims to investigate the possibility of using full-waveform, spaceborne, large-footprint lidar to derive forest type information, thus extending the application of large-footprint lidar and potentially contributing to the global methodology of a more accurate biomass estimation using large-footprint waveform lidar. Full waveform date from ICESat-GLAS was used in our research. The footprint of GLAS on Earth's surface are nominally 70 m in diameter, and the space between footprints is about 175 m. The study area was selected in Wangqing forest, a cool temperate forest in Jilin Province, China because of its variable forest types and the availability of ICESat-GLAS data.

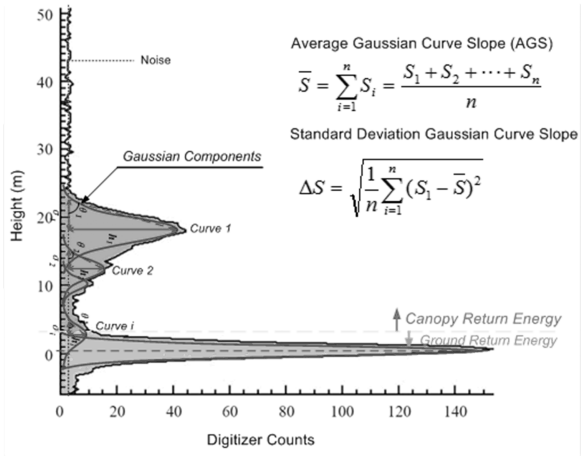

Figure 3. Proposed metrics AGS and SGS to be derived from lidar waveform.

Methods

Study Area

The study area is Wangqing forest area (43°05′N to 43°40′N, 129°56′E to 131°04′E), which is located in Jilin Province, China (Figure 4). Its area is approximately 85 km × 60 km 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 system. Wangqing forest is a cool temperate forest and dominated by a cool temperate continental climate. The elevations of the study area vary, ranging from 360 m to 1,477 m above sea level, and the steepest slopes are more than 100 percent (45 degrees). The natural vegetation in the study area is dominated by *Pinus koraiensis*, *Tilia amurensis*, *Quercus mongolica*, *Fraxinus mandshurica*, and *Acer mono*, and the mean forest canopy height is approximately 20 m (Xing *et al.*, 2010). The forests in the study area can be classified into three types: broad-leaved forest, needle-leaved forest, and mixed broad/needle-leaved forest.

ICESat Data

A full waveform dataset obtained from the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat) with cloud-free profile during the forest growth season of 27 October 2006 over the study area was acquired from the National Snow and Ice Data Center (NSIDC). ICESat provides 15 products from GLA01 to GLA15. In the study, the full waveform is derived from GLA01 (Global Altimetry Data Product) and precise geo-location of footprint centers from GLA14 (Global Land Surface Altimetry Data Product). A GLA01 waveform is linked to a GLA14 location by index and shot number.

Ancillary Data

A cloud-free Landsat TM satellite image acquired at the same time of year as when the fieldwork was conducted and a digital elevation model (DEM) of the study area were used in the fieldwork for visual interpretation of footprint location and to help design the sampling strategy. The Landsat TM image was geo-referenced and has a resolution of 30 m including seven bands of information. The horizontal resolution of the DEM is 20 m, and it was used to derive the slope map for footprint sampling plot design.

Field Sampling

Fieldwork was carried out in the Wangqing forest area in September to October 2007. A modified stratified random sampling method by segment slope was applied in the fieldwork due to the complex and rough terrain in the mountainous forest area, and the footprints in most cases are located in remote forest area far away from roads. A total of 103 circular plots centered on ICESat-GLAS lidar footprints with slope-corrected area of 500 m^2 were allocated in proportion to stratum size and measured in the footprints throughout the study area (Figure 4).

A Garmin eTrex GPS receiver with 10 m horizontal accuracy in dense forest conditions was used to locate the center of sampling plots coinciding with the center of ICESat footprints. The study tried to investigate the relationship between the forest type and lidar waveform profile, so in each plot, forest type, dominant tree species, and a maximum tree height were first recorded. Meanwhile, stem count and diameter at breast height (DBH) of broad-leaved and needle-leaved trees with DBH greater than 10 cm was measured, since these trees are the main contributors of the characteristics to the waveform profile. According to the data collected in the fieldwork, the footprint plots were classified into three categories: *needle-leaved forest* if the

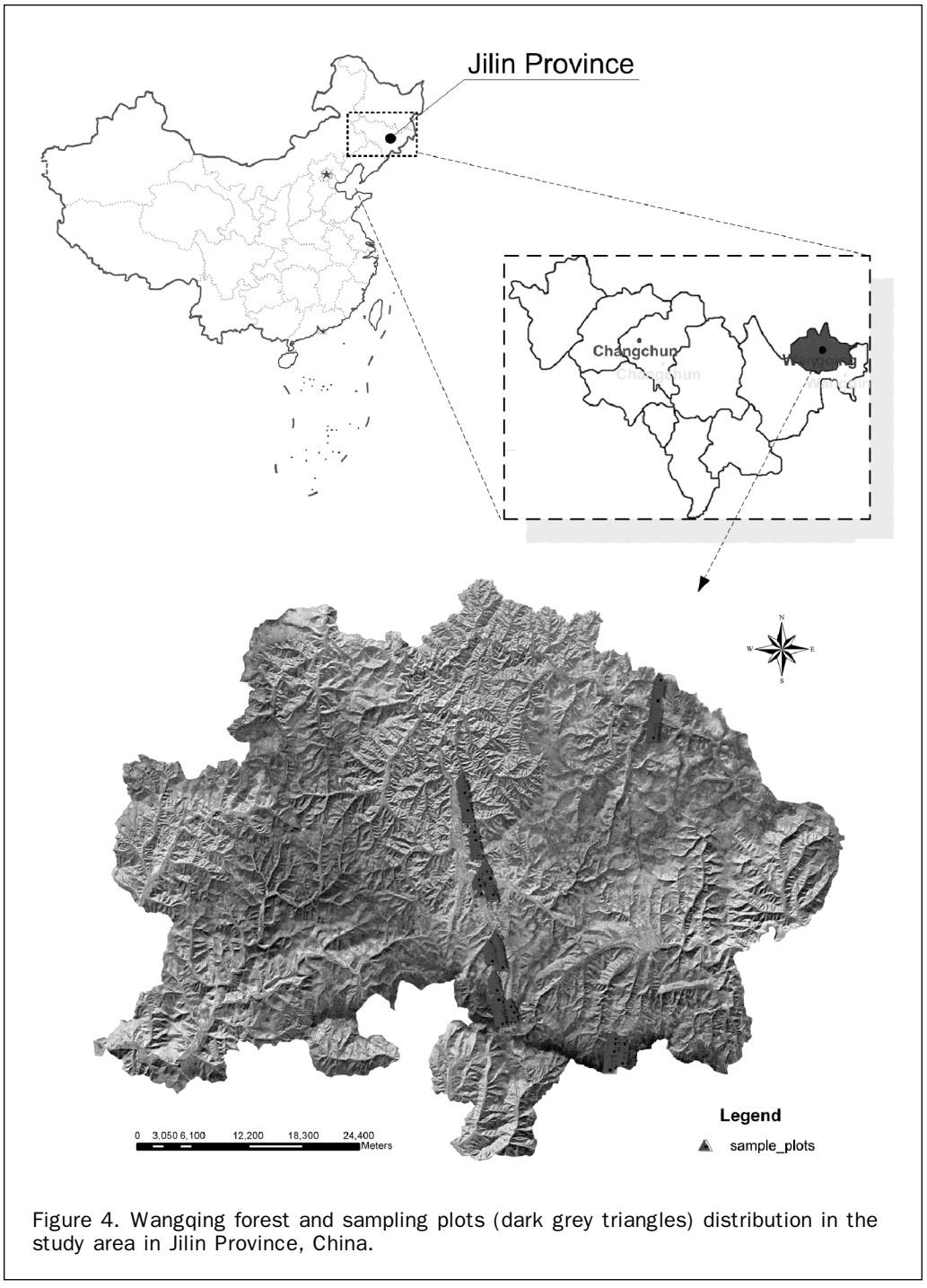

Figure 4. Wangqing forest and sampling plots (dark grey triangles) distribution in the study area in Jilin Province, China.

number of needle-leaved trees in the plot exceeded 70 percent, *broad-leaved forest* if the number of needle-leaved trees were under 15 percent, and otherwise *mixed broad-/needle-leaved forest.* Moreover, the actual slope of every sample plot was measured for comparison to the DEM derived slope, and to categorize the ICESat-GLAS lidar footprints in the data processing stage.

Waveform Processing

The ICESat-GLAS waveform data first goes through a pre-processing stage adopted from Duong *et al.* (2006a and 2006b), which consisted of waveform normalization, smoothing and Gaussian fitting. Then the proposed waveform metrics were derived from them.

The global full waveform data stored in ICESat's product GLA01 contains the voltage value of the waveforms. The voltage waveform was normalized by dividing amplitude V_i at each moment i by the total energy V_T of the waveform, where $V_T = \sum_{i=1}^{N} V_i$. The purpose of normalization was to make waveforms acquired in different epochs comparable, as different waveforms may be a result of laser pulses with different emitting energy, or captured under different atmospheric conditions. The normalized waveform can be described as $\left\{ V_N(i) = \frac{V_i}{V_T} \right\}_{i=1}^{N}$. In this study N equals 544.

Then, the normalized waveform was smoothed by a Gaussian filter with the Full Width at Half Maximum (FWHM) equals 3 to remove the noises in the waveform and help determine initial waveform parameters.

The smoothed waveform $w(t)$ can be modeled as a sum of Gaussian components $W_m(t)$ (Brenner *et al.*, 2003), as shown in Equation 1:

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

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 component 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 was used in the fitting processing to estimate the parameters in the model. The maximum number of decomposed Gaussian components was limited to 8. The quality of the fit was tested to be robust with 95 percent of the waveforms fitted well within 25 times noise standard deviation. Figure 5 gives an example of a raw waveform fitted with Gaussian components.

Waveform Metrics Derivation

After the waveforms were fitted with Gaussian components, the proposed waveform metrics were extracted from them using an automatic method detailed below.

Boundary Determination between Canopy Return and Ground Return

As stated in Research Problem Section, in order to derive the waveform metrics proposed, the canopy return and ground return of the waveform have to be separated. It has been noted by a number of studies that the laser beam of large-footprint lidar can always hit the ground below the tree canopy even in high dense forests. In research which was carried out in a study area with relatively flat terrain by Duong *et al.* (2006a), the rightmost Gaussian component of the waveform decomposition was simply regarded as the return from the ground below the trees. Another study by Rosette *et al.* (2008) used the last one or two Gaussian peaks whichever has greater amplitude as the ground return to extract maximum vegetation height from GLAS waveforms.

However, in our research, the ground was identified as the Gaussian component with the highest peak within the right half of the waveform range. This method was applied to waveforms located on slopes of less than 25 percent (14.04 degrees), which were selected in our research, for within this slope range, the height of slope would not exceed 17.5m ($\approx$ *footprint size* $\times$ *slope* $= 70 \times 25\% m$). Considering the average canopy height of 20 m in the study area, the ground return would only appear in the right half of the waveform. On steeper slopes, return from canopy could be seriously convolved or blurred with the surrounding topography (Hyde *et al.*, 2005).

After the ground return was identified, the boundary between canopy return and ground return was determined.

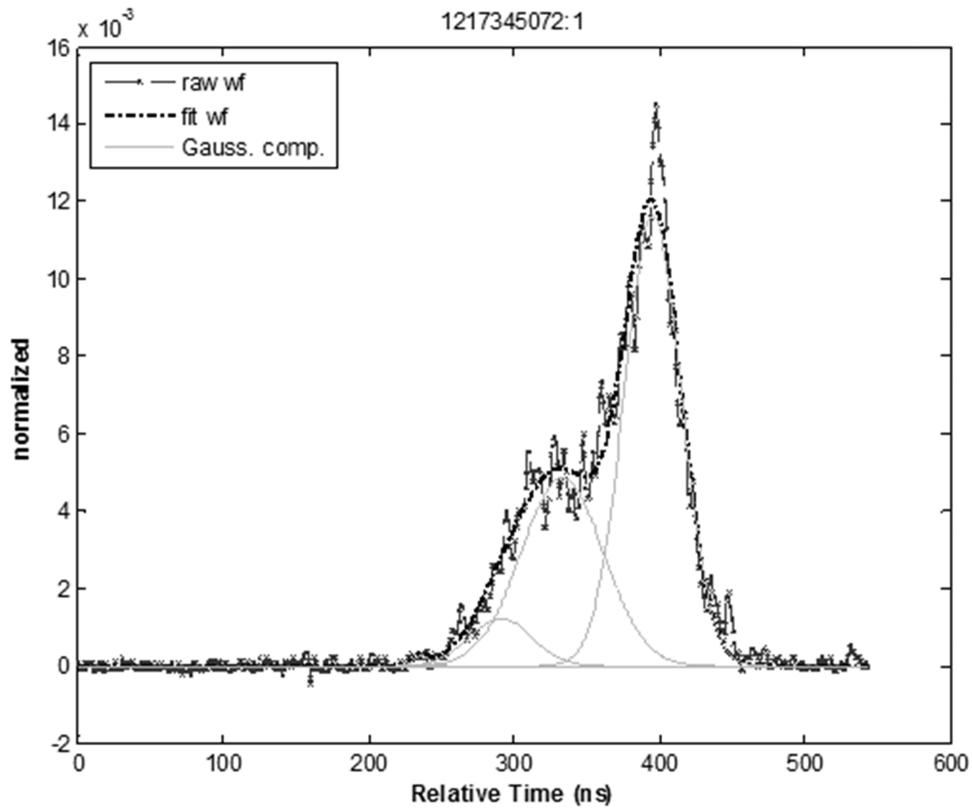

Figure 5. Lidar waveform on track 121745072 with shot number 1: raw waveform (black starred line), fitted waveform (black dash-dot line), and the Gaussian components (grey solid)

In our study, the position of the boundary was set to the left of the position of the Gaussian component identified as ground return, with a distance of 1.5 times its standard derivation as shown in Plate 1a. The waveform profile and Gaussian components to the left of the boundary were then defined as canopy return portion of the waveform. Lidar canopy height (LCH) was calculated as the distance from the signal start to the position of the Gaussian component of ground return.

Metrics Derivation

After the boundary between canopy return and ground return was determined, two sets of waveform metrics were derived from the canopy return part of the 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 components caused by canopy reflecting.

For the first set of metrics, CH25, CH50, CH75 were first measured as the heights of 25 percent, 50 percent, and 75 percent canopy return energy relative to the signal start, respectively (Plate 1b). Then, they were divided by lidar canopy height to product the ratios R25, R50, and R75, to normalize the influence of different canopy heights and the waveform spreading effect caused by topographical slope.

The slopes of the Gaussian components caused by canopy reflection are expected to have relationship with different forest types. The average of decomposed Gaussian curve slopes of the canopy return of waveforms of needle-leaved forests are assumed to be larger than that of broad-leaved forests, while the standard deviation of decomposed Gaussian curve slopes of the canopy return of waveforms of needle-leaved forests smaller than that of broad-leaved forests. In this case, the Average and Standard deviation of the decomposed Gaussian curve Slopes of the canopy return of each waveform, namely AGS and SGS (Figure 3; Plate 1c), were defined in Equations 2 and 3:

$$AGS = \bar{S} = \sum_{i=1}^{n} S_i \tag{2}$$

$$SGS = \Delta S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - \bar{S})^2} \; with \; S_i = \frac{A_i}{\sigma_i} \tag{3}$$

where n is the number of Gaussian components of canopy return part, A_i is the amplitude of the i^{th} Gaussian component, and σ_i its standard deviation.

A Modified Standard deviation of the decomposed Gaussian curve Slopes of the canopy return (MSGS) was also defined as Equation 4:

$$MSGS = \Delta S' = \sqrt{\sum_{i=1}^{n}\frac{E_i}{E_T}(S_i - \bar{S})^2} \; with \; E_T = \sum_{i=1}^{n} E_i \tag{4}$$

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 parts.

This Modified Standard deviation of the decomposed Gaussian curve Slopes of the canopy return (MSGS) is expected to amplify the characteristics of main stratum structures of the canopy. The main canopy strata are supposed to reflect a greater portion of lidar pulse energy, thus their corresponding Gaussian components are weighted higher in the calculation of MSGS.

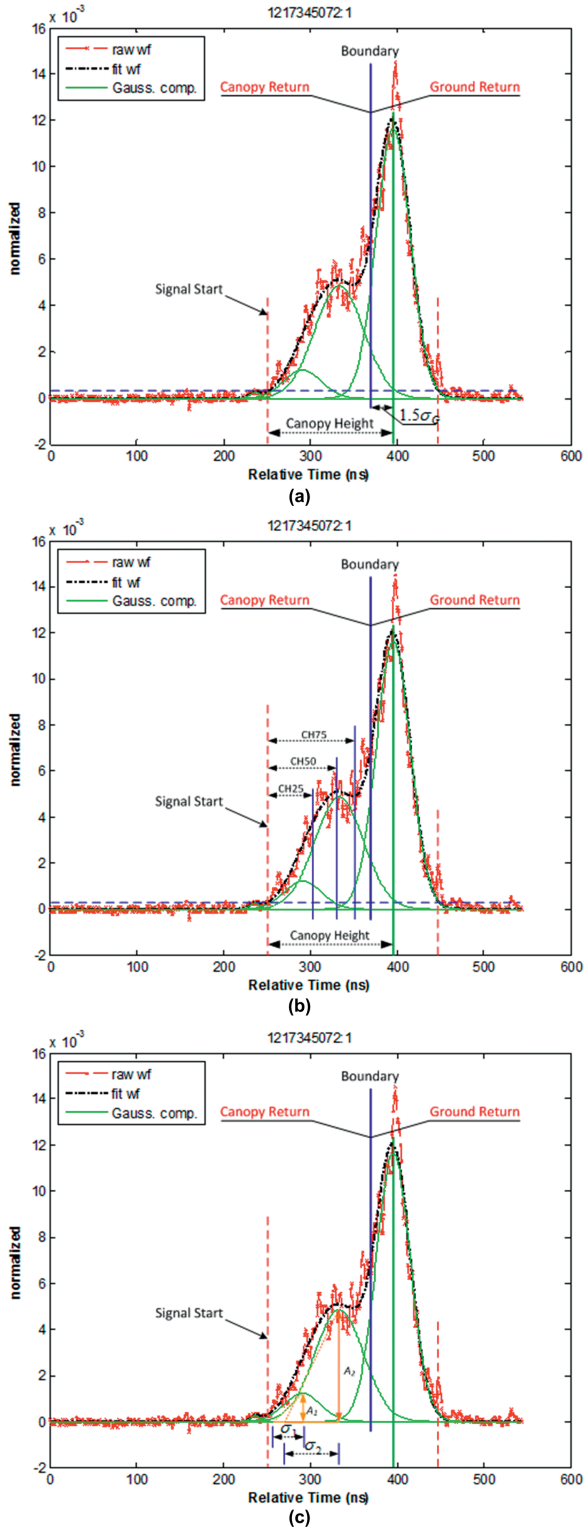

Plate 1. Waveform metrics derivation: (a) Boundary determined between canopy return and ground return in the waveform, (b) Height of 25 percent, 50 percent, and 75 percent canopy return energy, CH25, CH50, and CH75, and (c) The slopes of the Gaussian components contributed by the canopy reflection.

TABLE 1. METRICS COMBINATIONS THAT WERE TO BE TESTED IN SVM CLASSIFICATION

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

Waveform Metrics Classification

The two sets of derived waveform metrics, reflecting canopy vertical structure distribution and canopy vertical stratum characteristics, respectively, were then grouped according to their attributes into five combinations (Table 1), and used as input attributes in the Support Vector Machine (SVM) classification in Weka platform (Weka, 2007) to test which group of metrics can best predict different forest types. SVM has been proven as a powerful method for solving problem such as general (nonlinear) classification. SVM classification also has the advantage of having no restriction on the dimension of input data and could be trained to any size of training set (Gunn, 1998; Scholkopf and Smola, 2002; Vapnik, 1995). In the SVM classification, the LIBSVM algorithm (Chang and Lin, 2001) was used, and the linear Kernel function was selected. Classification was first applied to distinguish broad-leaved and needle-leaved forests, to find out the metric combinations with the highest classification accuracy. Then, classification using this group of metric combinations as input parameters was applied to the dataset where mixed forests were added in to check the extent to which the group of metrics can classify mixed broad-/needle-leaved forest. The classification accuracy of different metric combinations was assessed using confusion matrices, and performance metrics including overall accuracy, producer's and user's accuracies, and Kappa statistic were calculated (Congalton and Green, 2009).

Results and Discussion

Field Data Description

A total of 103 field plots of data were collected during the fieldwork. However, waveforms which were located on slope large than 25 percent (equal to 14.04 degrees), and those in which no clear ground return was discernable or where there was only ground return (no canopy return) were excluded from the later analysis, as the proposed metrics would not be able to be derived from those waveforms. Finally, 64 waveforms were left valid for the research, of which 35 were broad-leaved forest, 18 needle-leaved forest, and 11 mixed forest.

Metrics R25, R50, and R75 Classification Results

Table 2 showed the confusion matrix and SVM classification result of broad-leaved and needle-leaved forest using metrics R25, R50, and R75 as input attributes. From the table we can see that the waveforms of broad-leaved and needle-leaved forests cannot be separated by this group of metrics: all the 18 waveforms of needle-leaved forest were misclassified as broad-leaved forest. The classification result showed that the metrics R25, R50, and R75, which reflect canopy vertical structure distribution, failed to separate broad-leaved and needle-leaved forest.

This result may be due to the complexity of canopy vertical structure in natural cool temperate forests. The homogenous broad-leaved and needle-leaved forests with homogeneous tree species and canopy height usually have different typical vertical canopy structures. However, the natural forests in most cases were a mixture of different tree species and trees of different heights. The presence of understory was also common in the forest plots. Those factors made the canopy vertical structure differences between broad-leaved and needle-leaved forest not so distinct. Another reason may lie in the topographic effects on the vertical distribution of canopy return energy, which may also weaken the difference of vertical distribution of canopy return energy of different forest types.

Metrics AGS, SGS, and MSGS Classification Result

The metrics AGS, SGS, and MSGS, which reflect canopy vertical stratum characteristics performed better in distinguishing the two types of forests. Figure 6 illustrates a 2D plot of the SVM classification results of the two forest types using metrics AGS and MSGS. The pulse sign and cross symbols represent correctly-classified needle-leaved and broad-leaved forests, respectively. As we can see, a majority of the forests were correctly classified. Only five points were misclassified: three of them were misclassified needle-leaved forests and two were misclassified broad-leaved forests. Another noticeable point was that there was a clear pattern in the distribution of the two types of forests in the plot. The pulse signs (needle-leaved forests) were positioned at the upper-left part of the graph, while the cross symbols (broad-leaved forests) were located at the bottom-right part. With the same AGS value, the needle-leaved forests tended to have smaller MSGS than broad-leaved forests, which agrees with our previous assumption and indicates the hypothesis

TABLE 2. CONFUSION MATRIX AND CLASSIFICATION RESULT OF BROAD-LEAVED AND NEEDLE-LEAVED FOREST USING METRICS R25, R50, AND R75

	Reference Data			Classification Result			
				Percentage-based accuracy		Number-based accuracy	
Classification Data	*Broad*	*Needle*	*Classified Total*	Producer	User	Producer	User
Broad	35	18	53	100	64	35/35	35/53
Needle	0	0	0	0	0	0/18	0
Total	35	18	53				
Overall classification accuracy = 66.04% and kappa = 0							

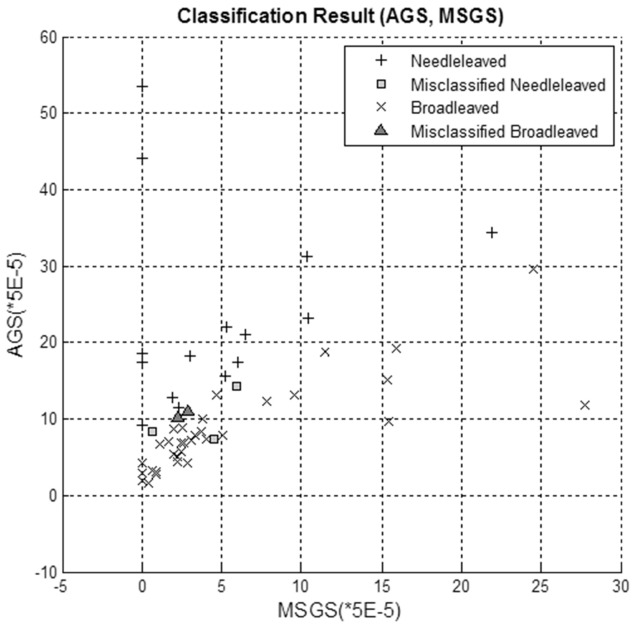

Figure 6. SVM classification result of broad-leaved and needle-leaved forests using metrics AGS and MSGS.

can be accepted: canopy stratum characteristics can be well represented by the proposed new metrics, and that these new metrics can be utilized to distinguish broad-leaved and needle-leaved forests.

Table 3 and Table 4 further present the classification results using different combinations of this group of metrics. Among the three combinations, AGS and MSGS achieved the highest overall accuracy of 90.57 percent and a Kappa statistic of 0.7868 and best predict broad-leaved and needle-leaved forests. The metric combination AGS and SGS, though it has the same overall accuracy as that of AGS and MSGS, got a lower Kappa value of 0.7808. When SGS is added into the combination of AGS and MSGS, the overall accuracy and Kappa decreases to 88.68 percent and 0.7406, respectively, rather than obtaining a better classification result, which proved MSGS was an improved version of SGS.

Comparison of Classification Results of Different Combination of Metrics

From Table 4 showing the classification results using all the five different combinations of metrics, we can conclude that the second type of metrics, AGS, SGS, and MSGS, representing the canopy stratum characteristics, were capable of distinguish broad-leaved and needle-leaved forests. The Kappa statistic value, together with high classification accuracy, indicated that this is a quite promising result for the classification based on full waveform analysis.

Classification Result when Mixed Forests Added In

The best indicators for broad-leaved and needle-leaved forests, metrics AGS, MSGS, identified from the previous classification, were then used to classify the dataset with mixed broad-/needle-leaved forests added in. Table 5 shows the preliminary results. The two metrics appear to be unable to distinguish mixed forests from broad-leaved and needle-leaved forests. Among the eleven mixed forests, six of them were misclassified as broad-leaved forests and the other five as needle-leaved forests. This may due to the fact that during a transition from broad-leaved forest to needle-leaved forest, waveforms returning from mixed forests will behave more similarly to needle-leaved forests, when needle-leaved tree canopy structures dominate the canopy of forest plots. This effect increases with the increased portion of needle-leaved trees, and vice versa. Another constraint of the

TABLE 3. CONFUSION MATRIX AND CLASSIFICATION RESULT OF BROAD-LEAVED AND NEEDLE-LEAVED FOREST USING AGS AND MSGS IN SVM CLASSIFICATION

	Reference Data			Classification Result			
				Percentage-based accuracy		Number-based accuracy	
Classification Data	*Broad*	*Needle*	*Classified Total*	Producer	User	Producer	User
Broad	33	3	36	94.29	91.67	33/35	33/36
Needle	2	15	17	83.33	88.24	15/18	15/17
Total	35	18	53				
Overall classification accuracy = 90.57% and kappa = 0.7868							

TABLE 4. COMPARISON OF SVM CLASSIFICATION RESULTS OF DIFFERENT COMBINATIONS OF METRICS

					Broadleaved		*Needleleaved*	
Metrics	Classified total	Number correct	Overall Accuracy	Kappa	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

*Classification Accuracy presented in percentage.

TABLE 5. CLASSIFICATION RESULTS OF BROAD-LEAVED, NEEDLE-LEAVED, AND MIXED FORESTS USING AGS AND MSGS

	Reference Data				Classification Result			
					Percentage-based accuracy		Number-based accuracy	
Classification Data	*Broad*	*Needle*	*Mixed*	*Total*	Producer	User	Producer	User
Broad	35	3	6	44	100.00	79.55	35/35	35/44
Needle	0	14	5	19	77.78	73.68	14/18	14/19
Mixed	0	1	0	0	0	0	0	0
Total	35	18	11	64				
Overall classification accuracy = 76.56% and kappa = 0.5642								

analysis was the small sample size of the mixed forest plots. More research will be needed to further explore how the portion of different tree types in the plot influence the characteristic of the waveform and proposed metrics, especially AGS and MSGS.

Conclusions and Future Work

In this study, a method for deriving forest type information from large-footprint spaceborne lidar waveform was developed. The research showed that the new proposed metrics derived from large-footprint lidar waveforms based on full waveform analysis can successfully separate broad-leaved and needle-leaved forests, thus further extending the application of large-footprint spaceborne lidar. The results also verified that the canopy stratum characteristics can be well-reflected by decomposed Gaussian components in the canopy return part of the waveforms and provided better understanding about how lidar pulse interact with forest canopy structures. This method has the potential to contribute to the global methodology of a more accurate biomass estimation using large-footprint spaceborne waveform lidar. It is also expected that this research could bring unique insight into the information extraction from large-footprint spaceborne lidar waveforms in forest application and inspire new ideas about better utilization of the products provided by current and next generation satellite lidar systems such as DESDynI (Deformation, Ecosystem Structure and Dynamics of ICE) and ICESat II in monitoring global forest and ecosystems.

The method in our study was applied and tested on ICESat-GLAS waveforms on low relief areas. A more automatic method to identify the canopy return part in the waveform on large slopes is suggested in the future research, so as to examine the relationship between the different forest types and metrics extracted from waveforms on large slopes. One possible solution 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, and in turn to obtain the canopy return range in the waveforms. In this way, the method can be applied to waveforms on terrain with larger slopes to some extent, and eliminates the subjective in ground return determination exists in many studies, to investigate the relationship between the metrics proposed and different forest types on large slope areas. Another problem that needs more investigation is how the portion of different tree types changes the behavior of the waveform and the characteristics of the proposed metrics. More field sampling is needed to conduct more comprehensive statistical analysis of the metrics and more scientific methods have to be employed to quantify the portions of different tree types in each forest plot. One possible improvement could be conducted using the portions of basal area of different type of trees, rather than the portion of different type of trees in number, to determine the field plot as broad-leaved or needle-leaved forest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant: 40871192), the Foundation of the Advanced Programs of the State Human Resource Ministry for Scientific and Technical Activities of Returned Overseas Chinese Scholars, and the Fundamental Research Funds for the Central Universities (Grant: DL09CA08). The authors would like to thank Ontario Centers of Excellence (OCE) and Natural Sciences and Engineering Research Council of Canada (NSERC) for providing funding to present this work on Silvilaser 2009 conference and publication. The National Snow and Ice Data Center (NSIDC) are thanked for their distribution of GLAS data. Many thanks are extended to Dr. Sorin Popescu, Dr. Ross Nelson, and the three anonymous reviewers for their valuable comments and suggestion. We also would like to thank Ms. Eleni Armenakis for her review of the English language.

References

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.

Brenner, A.C., H.J. Zwally, C.R. Bentley, B.M. Csathó, D.J. Harding, M.A. Hofton, J.-B. Minster, L.A. Roberts, and J.L. Saba, 2003. Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights, *Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document*, Version 4.1.

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

Chang, C.-C., and C.J. Lin, 2001. LIBSVM: A library for support vector machines, URL: *http://www.csie.ntu.edu.tw/~cjlin/libsvm/*, (last date accessed: 23 December 2010).

Congalton, R.G., and K Green, 2009. *Assessing the Accuracy of Remotely Sensed Data - Principles and Practices*, Second edition, CRC Press, Taylar & Francis Group, Boca Raton, Florida, 183p.

de Gier, A., 2003. A new approach to woody biomass assessment in woodlands and shrublands, *Geoinformatics for Tropical Ecosystems* (P.S. Roy, editor). Asian Association of Remote Sensing, pp. 161–198.

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.

Drake, J.B., R.O. Dubayah, D.B. Clark, R.G. Knox, J.B. Blair, M.A. Hofton, R.L. Chazdon J.F. Weishampel, and S. Prince, 2002. Estimation of tropical forest structural characteristics using large-footprint lidar, *Remote Sensing of Environment*, 79(2–3):305–319.

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

Duong, H., N. Pfeifer, N., and R. Lindenbergh, 2006a. Analysis of repeated ICESat full waveform data: Methodology and leaf-on / leaf-off comparison, *Proceedings of the Workshop in 3D Remote Sensing in Forestry*, 14–15 February 14–15, Vienna.

Duong, H., N. Pfeifer, and R. Lindenbergh, 2006b. Full waveform analysis: ICESat laser data for land cover classification, *International Archives on Photogrammetry, Remote Sensing and GIS*, Enschede, the Netherlands, XXXV(7).

Gunn, S.,1998. *Support Vector Machine for Classification and Regression* (Technical Report), 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*, M.Sc. thesis, ITC, Enschede, The Netherlands.

Hyde, P., R. Dubayah, B. Peterson, J.B. Blair, M. Hofton, C. Hunsaker, R. Knox, and W. Walker 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, Cambridge University Press.

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(4):503–511.

Lamotte, S., J. Gajaseni, and F. Malaisse, 1998. Structure diversity in three forest types of north-eastern Thailand (Sakaerat Reserve, Pak Tong Chai), *Biotechnology, Agronomy, Society, and Environment*, 2(3):192–202.

Lefsky, M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002. Lidar remote sensing of above-ground biomass in three biomes, *Global Ecology & Biogeography*, 11:393–399.

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., D. Harding, M. Keller, W. Cohen, C. Carabajal, F. Del Bom Espirito-Santo, M.O. Hunter, and R. de Oliveira Jr., 2005. Estimates of forest canopy height and aboveground biomass using ICESat, *Geophysical Research Letters*, 32(L22S02).

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, *Photogrammetric Engineering & Remote Sensing*, 66(11):1367–1371.

Muukkonen, P., and J. Heiskanen, 2007. Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories, *Remote Sensing of Environment*, 107:617–624.

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, aqnd 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., G. Sun, K. Kovacs, and V.I. Kharuk, 2004. Landcover attributes from ICESat GLAS data in central Siberia, *Proceedings of the Geoscience and Remote Sensing Symposium, 2004: IGARSS '04*, IEEE International, pp. 753–756.

Rosette, J., P. North, J. Suarez, and S. Los, 2008. Representation of vegetation and topography within satellite LiDAR waveforms for a mixed temperate forest, *Proceedings of SilviLaser 2008*, 17–19 September, Edinburgh, UK, unpaginated CD-ROM.

Scholkopf, B., and A. Smola, 2002. *Learning with Kernels*, MIT Press, Cambridge, Massechuttes.

Sun, G., K.J. Ranson, D.S. Kimes, J.B. Blair, and K. Kovacs, 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data, *Remote Sensing of Environment*, 112(1):107–117.

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

Wagner, W., A. Ullrich, V. Ducic, T. Melzer, 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, URL: *http://www.cs.waikato.ac.nz/~ml/weka/index.html*, Machine Learning Group at University of Waikato (last date accessed: 23 December 2010).

Xing, Y., A. de Gier, J. Zhang, and L. Wang, 2010. An improved method for estimating forest canopy height using ICESat-GLAS full waveform data over sloping terrain: A case study in Changbai mountains, China, *International Journal of Applied Earth Observation and Geoinformation*, 12(5):385–392.