ESTIMATING CANOPY HEIGHT OF MONTANE COOL TEMPERATE FOREST USING LARGE-FOOTPRINT SPACEBORNE LIDAR

Yanqiu Xing [a, *], Junjie Zhang [b]

[a] Center for Forest Operations and Environment, Northeast Forestry University, Harbin, China - yanqiuxing@yahoo.com.cn
[b] School of Resource and Environmental Science, Wuhan Univeristy, Wuhan, China

KEY WORDS: LiDAR, Forest, Canopy height, Estimation, Model

ABSTRACT:

Estimation of forest carbon storage is a critical challenge for understanding the global carbon cycle because it dominates the dynamics of the terrestrial carbon cycle. Light Detection and Ranging (LiDAR) system has a unique capability for estimating accurately forest canopy height, which has a direct relationship and can provide better understood to the aboveground carbon storage. To test the capacity of the large-footprint LiDAR for estimating canopy height in the montane cool temperate forest, the full waveform data of the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat) was used to extract forest canopy height in Wangqing of China. In this study, the maximum forest canopy height was regressed as a function of waveform extent and the elevation change in terms of terrain slope ranges. Final regression model for slope range of 0 to <5° explained 81% variance of maximum canopy height. With the increasing of slope ranges, the model accuracies significantly declined. The regression model could explain 61% and 47% of variance at the plots of which the terrain slope within 0 to <10° and 0 to <15° respectively. When the terrain slope is beyond 16°, the regression models became not so reliable anymore with less than 36% agreement. The results showed that the GLAS waveform data provides reasonable prediction for the maximum canopy height in the cool temperate forest of Northeast China.

1. INTRODUCTION

Forests play an important role in global carbon cycling for they are large pools of carbon as well as potential carbon sinks and sources to the atmosphere, and the accurate estimation of forest biomass is necessary for inventorying greenhouse gas and accounting terrestrial carbon (Muukkonen and Heiskanen, 2007).

The accurate estimation of forest biomass depends on a series of factors, in which the forest type and canopy height show more important roles. Passive optical remote sensing and active radar techniques have been widely applied in classifying forest type (e.g. Hagner and Reese, 2007; Lela and Said, 1993; Saatchi and Rignot, 1997). They, however, hold limitations in predicting the forest canopy height, which hampers the accurate estimation of forest biomass.

As a relatively new active remote sensing technique, the Light Detection And Ranging (LiDAR) system has a unique capability for estimating accurately forest canopy height (Hyde et al., 2005; Streutker and Glenn, 2006). LiDAR systems are typically classified into two types, depending upon the size of the “footprint” for the laser pulse. Large-footprint LiDAR systems digitize the full returned energy waveforms that cover relatively large areas (typically greater than 5 m in diameter), while small-footprint LiDAR systems typically record the range of one or more discrete reflections from laser pulses over small areas (typically less than 1–2 m in diameter)(Wehr and Lohr, 1999).Large-footprint LiDAR systems are better for getting canopy height compared to small-footprint LiDAR systems for their laser energy consistently reaches the ground even in dense forests due to their larger footprint sizes and do not miss tree tops (Dubayah and Drake, 2000).

* Corresponding author.

For the large-footprint LiDAR systems, the measure of maximum canopy height is obtained from the travel time between canopy top (first return) and ground (last return) reflections (Harding et al., 2001; Hudak et al., 2002). The waveform data from large-footprint LiDAR instruments have been successfully used to estimate forest canopy heights, for example Hyde et al. (2005; 2006) used Laser Vegetation Imaging Sensor (LVIS) to predict the forest canopy height in the Sierra Nevada mountains of California, and found that the metrics derived from LVIS waveform data could explained around 60-85% of the variation of maximum or mean canopy height; Anderson et al. (2006) employed LVIS to measure maximum canopy height in the Bartlett Experimental Forest (BEF) in central New Hampshire (USA), and the results showed that the LVIS metrics explained up to 80% of the variation in maximum canopy height; and Sun et al. (2007) applied the waveform data from the Geoscience Laser Altimeter System (GLAS) instrument aboard the Ice, Cloud, and land Elevation (ICESat) satellite to estimate the forest canopy height in the forests around Tahe and Changbai Mountain areas in Northern China, and found the maximum tree height measured in the field was well correlated to that predicted from GLAS indices ($R^2 = 0.75$).

Currently, most of these studies, however, focus on the forests on relatively flat terrains (Lefsky et al., 2005b; Lefsky et al., 2005c) . For the regions in which the footprint topographic relief is larger compared to canopy height, the interpretation of waveforms is complex because the acting attributes to the waveform shape are not as obviously differentiated as those from flat terrains (Harding and Carabajal, 2005) . In these cases, the canopy and ground reflections are mixed together, which makes the interpretation of waveforms more difficult (Harding and Carabajal, 2005). Lefsky et al. (2005a) combined ICESat-GLAS waveforms and ancillary topography to estimate maximum forest canopy height in three ecosystems over sloped terrains: tropical broadleaved forests in Brazil, temperate broadleaved forests in Tennessee, and temperate needle-leaved forests in Oregon. Final models for each site explained between 59% and 69% of the variation of the field-measured forest canopy height. For better understanding the effects of slopped terrain on LiDAR-based forest canopy height estimation, it is needed to define the terrain slope threshold for the performance of the regression model.

With ICESat-GLAS waveforms, this study aimed to examine the effects of different slope terrain on large-footprint LiDAR-based forest canopy height estimation, and to explore the potential of LiDAR in estimating forest canopy height in montane areas. Wangqing forest area, a cool temperate forest in Jilin Province, China, was selected as a case study, for the topography and forest type variations with this area make it an ideal candidate to achieve the aims of this study.

2. MATERIALS AND METHODS

2.1 Study area

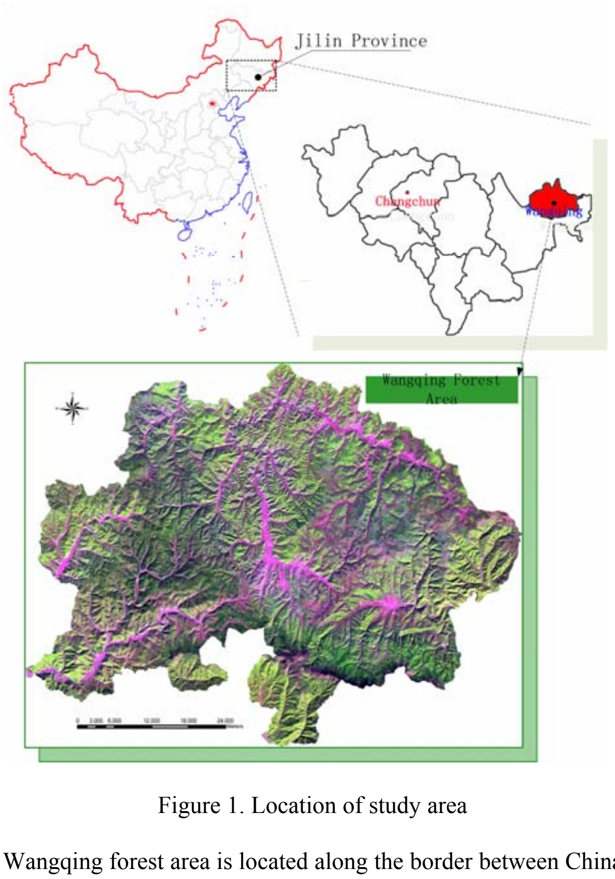

Figure 1. Location of study area

Wangqing forest area is located along the border between China and North Korea (43°05′~43°40′N, 129°56′~131°04′E), and its area is approximately 85 ×60 km (Figure 1). It belongs to Changbai mountain system, which is one of the most valuable reserves in China due to its rich gene pool of many plant species with the altitudinal vegetation zone in the Mountain. This region is 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. The mean annual temperature is 3.9 ℃, and the mean annual precipitation is 438 mm, about 80% of which falls between May and September. Elevation ranges from 360 to 1,477 m, and the terrain slopes range from 0 to 45° generally. The mixed conifer/broadleaved forest is the zonal vegetation between 500 and 1,100 m altitude. The dominant species are Korean pine (*Pinus koraiensis* Sieb. et Zucc.), Dahurian larch (*Larix gmelinii* Rupr.), Amur linden (*Tilia amurensis* Rupr.), Mongolian oak (*Quercus mongolica* Fisch.), Manchurian ash (*Fraxinus mandshurica* Rupr.) and Maple (*Acer mono* Maxim.). The mean forest canopy height is around 26 m and the soil type of this area is dark brown soil.

2.2 ICESat data

The full waveform data of the ICESat-GLAS were used to extract forest canopy height. GLAS is the first polar-orbiting LiDAR instrument for continuous global observations of the Earth. The laser footprint diameter on the Earth's surface is nominally 70 m, the space between footprints is about 175 meters, and the width of the transmitted pulse is 4 ns (nanosecond), equivalent to 60 cm in surface elevation. ICESat offers in total 15 products called GLA01 to GLA15. The waveforms were derived from GLA01 Global Altimetry Data Product, and geolocated footprint locations were obtained from GLA14 Elevation Data Product. The full waveform datasets with cloud-free profiles in the period from 2003-02-21 to 2006-10-27 over the study area were downloaded from the National Snow and Ice Data Center (NSIDC) (http://nsidc.org/data/icesat/, accessed on October 30 2007) for achieving the aim of this study.

2.3 Digital elevation model (DEM)

The fine resolution (20×20 m) DEM was obtained to (1) verify the geolocation accuracy of GLAS footprints; (2) calculate a terrain slope map that was used to stratify the study sites for footprint sampling plot locations; and (3) determine terrain index, which was defined as the range of ground surface elevations within one of three sampling windows (3×3, 5×5, and 7×7 DEM pixels) applied to a DEM at the GLAS footprint location (Lefsky et al., 2005a).

2.4 Field sampling

Although the forests in the study area, as a whole, are heterogeneous due to varied terrain conditions and the forest types, they are relatively homogeneous inside each forest management unit, which is a well defined and demarcated land area, predominantly covered by forests, managed on a long-term basis and having a set of clear objectives specified in a forest management plan (FAO, 2003). In this study, the plots in which the number of broadleaved/needleaved trees exceeds 70% were defined as broadleaved/needleaved forests, and plots with broadleaved or needleaved trees in number between 15% and 70% were then defined as mixed forest. It was found there were very few footprints located in the areas with terrain slope more than 30°, thus we classified the terrain slope into 4 categories: 0 to <5°, 5 to <15°, 15 to <25°, 25 to <30°. Plots were randomly selected within each slope category considering different forest types, and totally 166 sites within ICESat footprints were designed for sampling (see Figure 2).

The fieldwork was carried out in September-October 2007. A global position system (GPS) was used to localize each sampling site coinciding with the footprint centre. In each plot, a circular region with area of 500 m^2 in horizontal, which is a common plot size in forest investigation, was determined. The vegetation cover and ground cover charactered by tree species, percentages of broadleaved and needleaved tree and maximum tree height, were measured and recorded.

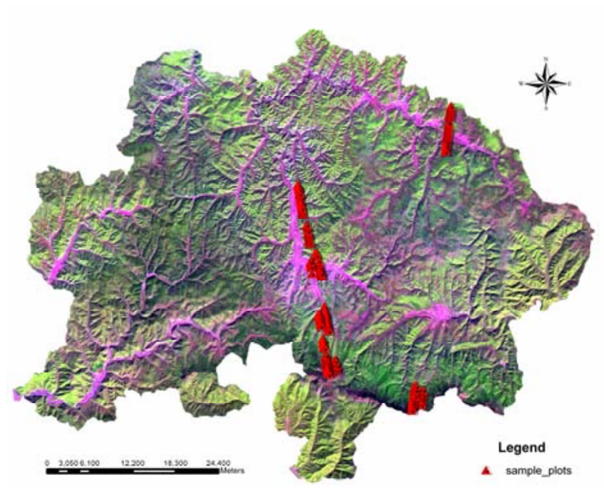

Figure 2. Field plots distribution

2.5 GLAS waveform processing

The binary data of GLA01 and GLA14 were firstly converted into ASCII format by the IDL program developed by the National Snow and Ice Data Center (NSIDC, 2006). Then, the waveform data, originally in counts from 0 to 255, were converted into voltage units, and the voltage waveform was normalized by dividing them by the total received energy to enable waveforms comparable, that means the area under any normalized waveform equals one. Next, the normalized waveforms were smoothed by a Gaussian filter. Finally, Gaussian components were fitted to the smoothed waveform. A detailed description of these processing steps was discussed by Duong et al. (2006).

2.6 Waveform extent extracting

LiDAR waveform extent was derived for predicting canopy height. Waveform extent is defined as the vertical distance between the first and last elevations at which the waveform energy exceeds a threshold level. The threshold was determined by fitting a Gaussian distribution to the peak of lowest energy in a histogram of waveform energy, which identifies the mode and standard deviation of background noise in each waveform. The threshold was set to the noise mode plus 4 times the standard deviation (Brenner et al., 2003).

2.7 Canopy height estimating

Considering different terrain slope ranges, the canopy height estimation method developed by Lefsky et al. (2005) was selected to predict the maximum forest canopy height, which relates field measured maximum canopy height to ICESat waveforms and DEM data:

$$H = b_0 (w - b_1 g)$$

Where H is the measured maximum canopy height, w is the waveform extent, g is the terrain index (ground extent) in meters, b_1 is the coefficient applied to the terrain index, and b_0 is the coefficient applied to the waveform height index (Lefsky et al., 2005a). In this study, we used the 20 m DEM data to extract the terrain index. The correlation between index and the difference between the GLAS waveform extent and the field observed maximum canopy height was investigated. The investigation showed that the terrain index derived from a square 3×3 matrix was a best choice to estimate the canopy height difference. Thus, regression was used to estimate maximum canopy height as a function of GLAS waveform extent and the 3×3 terrain index in the study. For the equation coefficients, b_0 and b_1, the examination was done to specify them for our study area. In terms of terrain slope ranges, all 166 plots were randomly divided into 2 groups for model calibration and validation.

3. RESULTS

For all GLAS footprints acquired from 2003 to 2006, the correlation of elevation derived from DEM and GLAS was 0.99 with RMSE equal to 0.016 m. According to the description of Carabajal and Harding (2005) and Sun et al. (2008), this result indicated that the error of GLAS geolocation is small.

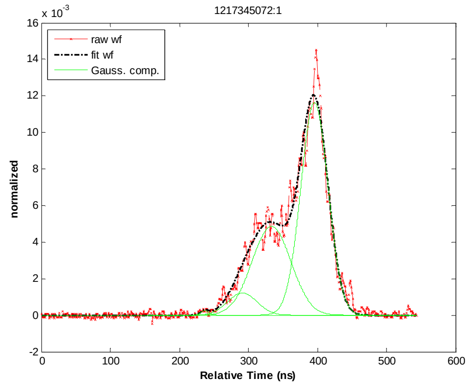

Figure 3. LiDAR waveform on track 121745072 with shot number 1: raw waveform-red, with the Gaussian component-green and fitted waveform-dashed black

Figure 3 shows the results of the fitting algorithm of LiDAR waveform. The leftmost Gaussian component referred to as the first mode corresponds to the first reflecting feature in the laser footprint, and it, over the forest areas, mostly originate from the reflection by the tree canopy. The rightmost Gaussian component corresponds to the energy reflected by the surface hit last, which corresponds to the last ground return below the tree in forest applications.

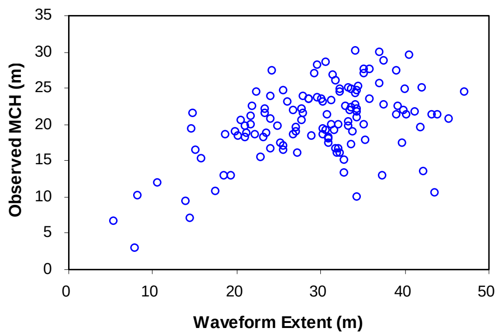

Figure 4. Correlation between waveform extent and observed maximum canopy height (MCH)

For all 166 forest plots, Figure 4 shows that there is almost no linear relationship between waveform extent and the observed maximum canopy height. To test the performance of model developed by Lefsky et al. (2005a), 140 plots were selected randomly for calibration, and the result shows that model was invalid with R^2 close to zero. In this study, the model was defined invalid when the residual sum of squares was higher than the corrected total sum of squares.

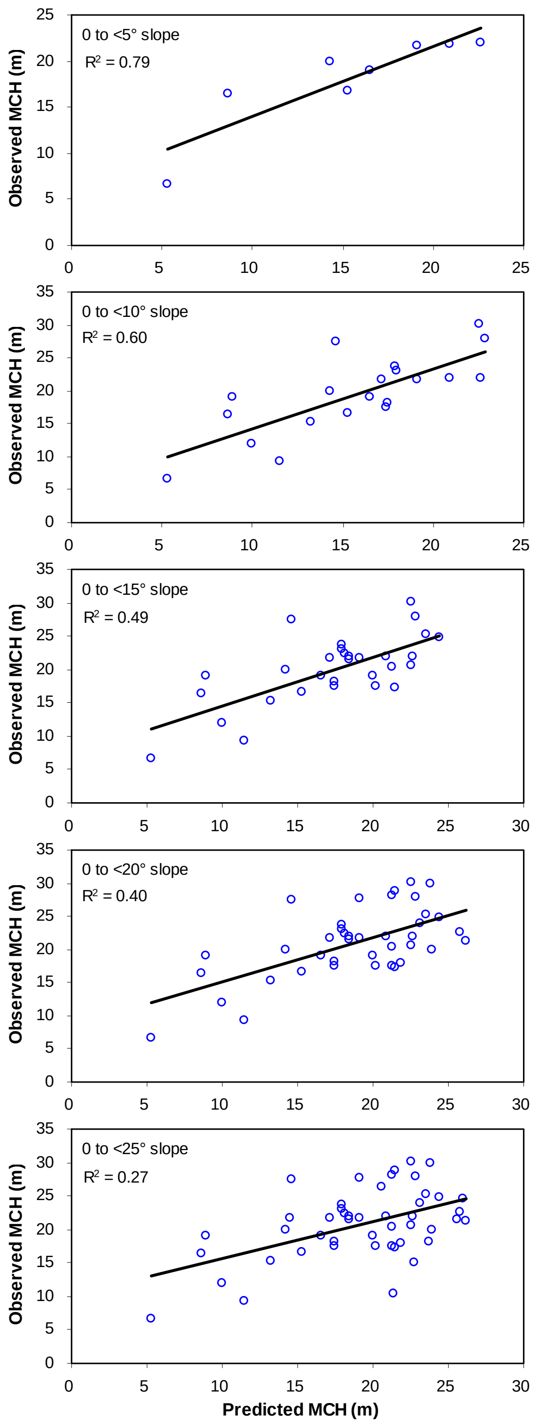

Figure 5. The correlation between the observed and predicted maximum canopy height (MCH) considering different terrain slope ranges

To evaluate the model performance according to terrain slope range, the data were then grouped into the cumulative slope categories. For each slope category, the model calibration and validation were done to determine how the model works in terms of the terrain slope. Table 1 shows the valid model fitting results of slope categories with *p*-level less than 0.05, and Figure 5 displays the model validation curve for each slope category. For the slope category of 0 to <5°, the regression model could explain 81% (RMSE=4.36 m) variance of maximum canopy height. With the increasing of slope ranges, the model accuracies significantly declined. The regression model could explain 61% (RMSE=4.20 m) and 47% (RMSE=4.31 m) of variance at the plots of which the terrain slope within 0 to <10° and 0 to <15° respectively. When the terrain slope is beyond 16°, the regression models became not so reliable anymore with less than 36% (RMSE=4.62 m) agreement. And, for the slope category of 0 to <25°, the model can only explain 7% of variance of the maximum canopy height. The independent model validation also showed the same trend. The model validation accuracy was 79% and 60% respectively for the slope range of 0 to <5° and 0 to <10°, and less than 50% when the terrain slope was beyond 15°.

Slope category	N	b_0±SE	b_1±SE	R^2	RMSE, m
0-<5°	12	0.707 ±0.104	-0.716 ±0.856	0.81	4.36
0-<10°	49	0.609 ±0.057	-1.204 ± 0.531	0.61	4.20
0-<15°	76	0.597 ±0.048	-1.021 ±0.446	0.47	4.31
0-<16°	80	0.585 ±0.050	-1.035 ±0.4711	0.36	4.62
0-<20°	90	0.601 ±0.048	-0.827 ±0.420	0.28	4.96
0-<25°	100	0.515 ±0.049	-1.435 ±0.551	0.07	5.77

Table 1. Model fitting results in terms of slope categories

4. DISCUSSION AND CONCLUSION

In this study, we found that the field observed and LiDAR predicted maximum canopy heights showed good agreement (R^2=0.81) on the flat terrain (0 to <5° slope). Such a result was comparable to the findings in other researches, for example, Sun et al. (2007) predicted maximum canopy height using GLAS data with agreement of 0.73 and RMSE of 3.4 m on the flat terrain.

We also found that, with the increasing terrain slope, the performance of regression models declined (Table 1.). Among all slope ranges, the regression models performed good up to 15° in this study. Both of the slope range for valid regression models and validation results were similar with those of Lefsky et al. (2005a), in which the regression equations explained between 59% and 68% of variance of maximum canopy height for the slope range between 0 to 15°. In our study area, the forest area in the 0 to 15° slope terrain accounts for more than 72% of total forest area (Zhao and Wang, 2007), we therefore could conclude that the regression model in the 0 to 15° terrain slope range was capable to be employed for estimating maximum canopy height in this study area.

As shown in Figure 5, the independent model validation indicated that the LiDAR predicted heights were lower than the field measurements in some plots. Most of those plots were corresponding to the GLAS footprints acquired in 2003 and 2004, which have 2 to 4 year interval to the field investigation carried out in 2006 and 2007 respectively. Some growth during the time interval may affect the accuracy of the validation. In addition, 2 times of field data collections were done both in the leaf-on seasons and some of LiDAR footprints were captured in the leaf-off seasons, and thus the seasonal effect could be another source of errors. It was possible that those leaf-off footprints with deciduous trees hit the tree top without leaves, and the differences between the leaf-on and leaf-off heights could influence the accuracy of the validation.

The study showed that the GLAS waveform data provides reasonable prediction for the maximum canopy height in the cool temperate forest of Northeast China, but further study areas are still needed to evaluate and reduce the terrain effect on canopy height estimation using GLAS waveform data, especially in the terrain with slope larger than 15°, which is the slope limitation for our study area.

REFERENCES

Anderson, J. et al., 2006. The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of Environment*, 105(3), pp. 248-261.

Brenner, A.C. 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.

Carabajal, C.C. and Harding, D.J., 2005. ICESat validation of SRTM C-band digital elevation models. *Geophysical Research Letters*, 32, L22S01, doi: 10.1029/2005GL023957.

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

Duong, H., Pfeifer, N. and Lindenbergh, R., 2006. Analysis of repeated ICESat full waveform data: methodology and leaf-on/leaf-off comaparison. In Proceedings: *Workshop on 3D Remote Sensing in Forestry*.

Hagner, O. and Reese, H., 2007. A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment*, 110(4), pp. 438-444.

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

Harding, D.J., Lefsky, M.A., Parker, G.G. and Blair, J.B., 2001. Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, 76(3), pp. 283-297.

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

Hyde, P. et al., 2005. Mapping forest structure for wildlife habitat analysis using waveform lidar: Validation of montane ecosystems. *Remote Sensing of Environment*, 96(3-4), pp. 427-437.

Hyde, P. et al., 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102(1-2), pp. 63-73.

Lefsky, M.A. et al., 2005a. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32, L22S02, doi: 10.1029/2005GL023971.

Lefsky, M.A., Hudak, A.T., Cohen, W.B. and Acker, S.A., 2005b. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. *Remote Sensing of Environment*, 95(4), pp. 532-548.

Lefsky, M.A., Hudak, A.T., Cohen, W.B. and Acker, S.A., 2005c. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment*, 95(4), pp. 517-531.

Lela, M. and Said, M., 1993. Forest inventory using satellite optical and radar remote sensing with ancillary data. *Buletin Ukur*, 4(2), pp. 56-64.

Muukkonen, P. and Heiskanen, J., 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, pp. 617-624.

Saatchi, S.S. and Rignot, E., 1997. Classification of boreal forest cover types using SAR images. *Remote Sensing of Environment*, 60(3), pp. 270-281.

Streutker, D.R. and Glenn, N.F., 2006. LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment*, 102(1-2), pp. 135-145.

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

Sun, G., Ranson, K.J., Masek, J., Fu, A. and Wang, D., 2007. Predicting tree height and biomass from GLAS data. In: *10th International Symposium on Physical Measurements and Signatures in Remote Sensing*, Davos, Switzerland.

Wehr, A. and Lohr, U., 1999. Airborne laser scanning--an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 68-82.

Zhao, Z. and Wang, L., 2007. A comparative study of the denoising methods of Thematic Mapper images for forest areas. *Joural of Forestry Research*, 18(2), pp. 123-127.