Within polygon grid based sampling for height estimation with LIDAR data

Mike Wulder[1], David Seemann[1], and Alain Bouchard[2]

[1] Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, 506 West Burnside Rd. Victoria, British Columbia, Canada, V8Z 1M5
[2] Geomatics Research Centre, Laval University, Quebec City, Quebec, Canada

Abstract

The expected canopy top height of a forest stand polygon is an important forest inventory variable. The use of LIDAR, or laser altimetry, to develop a remote estimate of canopy top height is accordingly desirable. Forest stand polygons are delineated from air photographs by interpreters to indicate regions of homogeneous forest conditions. The perceived homogeneity of forest stand polygons is a complex amalgam of forest conditions grouped using visual clues and the experience of the interpreter. An averaging of all LIDAR hits within a polygon is expected to give a biased result. Placing a sampling grid over a polygon within which maximum height may be used as representative, will factor in for crown openings and differing strata.

In this research, we investigated the development of tools for extraction of the within polygon LIDAR values, such as for the processing of LIDAR values for separation of overstorey from the understorey values, and differing grid sizes and input values. The goal of the research was to develop a system for remote estimates that would result in a representative canopy height from LIDAR data. To do so, we developed a within polygon grid based sampling system. The within polygon sampling system allowed for setting of a series of parameters to investigate the sensitivity of changes in height estimates from alterations to the system parameterisation. Additionally, we compare the grid based within polygon estimates to a non-gridded control. Our results indicate understorey improved the results, the number of within grid hits to average has a nominal impact in height estimates, and that the size of the sampling grid has a large impact upon final height estimates.

Introduction

Height of a forest stand is an important inventory attribute allowing for estimation of volume, biomass, future yields, and for determination of potential stand treatments. In the context of forest inventories utilising air photos, height is normally measured directly, through stereoscopic parallax, or shadow measurement. Interpretative measures of height from photos are often aided with ground validation data. Some examples of factors which affect the accuracy of height estimation in forest inventory are, film emulsion, scale, focal length, time of day, shape of tree, character of shadow, character of the forest, topography, observer skill, and measurement technique (Spurr, 1948).

Optical remotely sensed data, such as Landsat, typically provide a 2-dimensional representation of forests. A 2-dimensional representation of forests requires inference to be applied when estimating vertically distributed parameters (Wulder, 1998). LIDAR data provides for 3-dimensional representation of forest structure. This 3-dimensional view of forest structure allows for estimates of vertically distributed elements of the forest. LIDAR data is well suited to making measurements of individual trees (St-Onge, 1999), tree heights from canopy heights (Magnussen and Boudewyn, 1998), and canopy heights (Nelson, 1997).

Many LIDAR hits may be recorded in an individual forest polygon, yet these measured heights may indicate ground, understorey vegetation, or various strata of the overstorey. To take an average value to represent the polygon as the height an underestimate would result as many of the recorded hits do not relate canopy characteristics. To make use of all the LIDAR hits in a polygon, and still generate a valid estimate of the polygon height, we explore a method where each polygon is partitioned into grid sections. Within each grid element within a polygon we extract height values representative of that portion of the polygon which in-turn is used in the estimate of polygon height by averaging the grid values. Imposition of a grid allows for the estimation of a spatially weighted average height for the polygon to be estimated from the LIDAR data. This "grid" method follows examples by Nilsson (1996), Magnussen and Boudewyn (1998), Aldred and Bonner (1985), Næsset (1997), and Ritchie (1995).

In Figure 1 an example polygon is presented, where the variability in LIDAR hits is cartooned as well as a sample grid. The points indicated in the polygon represent a measured height. The forest characteristics for which the height is measured may be the ground, low shrubs, or any strata of the canopy. When extracting height estimates from the LIDAR within each grid element, several factors related to the grid may be altered. In this communication we investigate altering the grid cell size, the number of cells to average within the cell, and the minimum hit height for inclusion. A grid cell factor that we also assess is the impact of altering the number of cell values required for inclusion in the grid average (this factor is largely included to avoid the inclusion of zero values from cells with no LIDAR hits).

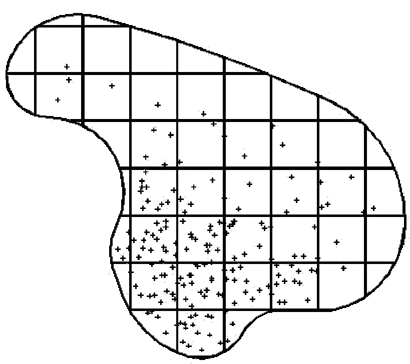

Figure 1. Example of variability of LIDAR hits within a polygon and a sample grid overlay

Methods

Study Area

The study area is located in central Saskatchewan near the southern limit of the boreal forest and is classified as mixed boreal forest. Mixed woods composed of aspen and white spruce which are common where the sites are well drained; whereas, Jack pine (*Pinus banksiana*, var. Lamb) and black spruce (*Picea mariana*, var. Mill.) are found with pure stands of jack pine on dry sites composed of coarse textured soils. In poorly drained areas, bogs support black spruce and small proportions of tamarack (*Larix laricina*, var. Du Roi) (Rowe, 1977; Lowe et al., 1996). Also present are fen areas, which are composed mostly of sedge vegetation with discontinuous cover of tree species such as tamarack. Forest disturbance is largely the result of localised logging operations and fire. Recent fires have generally been limited in areal extent and frequency through a comprehensive forest fire suppression program (Sellers et al., 1995).

Forest inventory data (GIS)

The forest inventory system in Saskatchewan is based on interpretation and digitisation of air photos on an approximate 15 year completion cycle (Gillis and Leckie, 1993). Inventory validation is undertaken through field visits and the establishment of temporary sample plots. The forest inventory data provided for this study is of variable vintage, with 82.7% of the inventory compiled in 1984; 3.8% compiled before and 13.5% after 1984.

SLICER LIDAR data

The SLICER was developed at the NASA Goddard Space Flight Centre as a scanning modification of a profiling laser altimeter (Blair, et al. 1994). The SLICER is a LIDAR system which digitises the backscattered return signal resulting in the capture of a full waveform representing the vertical distribution of illuminated surfaces within the laser footprint. In this study the footprint diameter was approximately 9m, varying by approximately ±5% due to laser divergence and changes in the distance from the aircraft to the ground.

As a component of the Boreal Ecosystem – Atmosphere Study (BOREAS) (Sellers et al., 1995) the SLICER data utilised in this study was collected in July of 1996[1] (Harding, 1998). The BOREAS LIDAR data was processed from the raw data into variables representing key components of the sensed waveform (Harding, 2000). For this study we primary utilised the "ground start" variable, which is the distance

[1] Detailed descriptions of the SLICER instrument and data utilised in this study may be found at: http://www-eosdis.ornl.gov/BOREAS/guides/SLICER.html and Harding et al. *In press*.

between detected laser returns from the canopy top and underlying ground, hereafter called height. A simple processing algorithm was used to determine the height from the full waveform data. The location of the footprint is referenced to the first detected reflection (i.e. the canopy top). Accordingly, the absolute geolocation accuracy of footprint locations is limited by the degree of elevation change within the footprint, the differential GPS positioning of the aircraft, and knowledge of the laser pointing established by means of an Inertial Navigation System and encoding of the scanning mirror angle. Tag time errors in the independently recorded data streams, where the range and angle are in one data stream and GPS information in another, may introduce occasional geolocation errors. As a result, the footprint location accuracy can be expected to be at the scale of the laser footprint, in this case, within 9m.

To compare the LIDAR data to **plot data** to calibrate the plot to LIDAR to polygon height estimates we extracted all the laser shots within 9m of a field measured plot. As a result we had available for analysis 22 LIDAR hits related to 6 plots that are found in 3 polygons. While the data set is smaller than optimal, we are limited to the material mined from the BOREAS study. We contend that the quality of the measures and positioning is high enough to counter sample size issues. We intend to use the results generated from this sample as exploratory and as an aid to development of future experiments. In the case of the **polygon based** estimates and comparison between LIDAR estimates and forest inventory estimates of height we used a robust 4104 samples.

Based upon the data available and the issues related to the impact of altering grid characteristics, we may estimate the height by:

- Using no grid and averaging all laser shots within a polygon
- in a grid while altering:
    - grid cell size
    - minimum number of hits to average to represent the grid cell in the averaging (i.e. non-zero value)
    - number of hits to average to represent a cell, and
    - minimum height.

Results and Discussion

For an initial calibration of the plot, LIDAR, and polygon data we compare the heights generated from each data source (Table 1). In Saskatchewan polygons are often placed into 5 meter height classes. The plot and polygon estimates of height compare favourably. There are some slight differences between plot and LIDAR height estimates, with a mean difference of –0.42m and a standard deviation of 1.45m.

Table 1. Comparison of height estimated for plots, LIDAR, and polygons (the difference between LIDAR and plot data is also compared)

Polygon	Plot	LIDAR	Comparison
Height	Height	Height	LIDAR-Plot
5	4.55	3.34	-1.21
10	10.47	9.17	-1.30
15	13.91	15.17	1.26
			Mean = -0.42 Standard deviation = 1.45

The estimates of polygon height from the LIDAR are first computed without a grid, with all LIDAR hits within a polygon averaged. Both the mean and median estimates are found to underestimate the polygon height (Table 2).

Table 2. Difference between average LIDAR height within a polygon to forest inventory height estimate ("LIDAR – polygon" denotes the difference in estimates)

Statistics	LIDAR-Polygon	Standard deviation
Mean	-2.00	2.75
Median	-1.10	2.61

The underestimate of height with the averaging of all LIDAR values was expected. By averaging values extracted from within grid cells superimposed upon a polygon we expect the difference between estimates to be reduced. In the following tables, the gird characteristics are denoted as:

[Grid_x_size] x [grid_y_size] – [minimum hits per cell] – [number of hits to average] – [minimum height]

Or

10x10-1-2-2

where, a 10 by 10 metre grid is created, where there must be at least one hit in the cell to be included in the average; where the average is based upon extracting from each cell the average of the two highest laser shots that are greater than a minimum height of 2 meters.

In Table 3 we examine the influence of grid cell size, finding that the grid size has a large influence on the height estimates. As expected, as the grid cell size increases the estimated height increases. As the grid cell size increases more hits are included in the cell creating opportunity for higher values to be included. In this example the number of hits to average within each cell was kept constant at 2.

Table 3. Height comparison between LIDAR and polygons when grid cells size is altered

Grid	LIDAR-Polygon	Standard Deviation
10x10-1-2-2	-0.40	3.07
15x15-1-2-2	-0.04	3.11
20x20-1-2-2	0.25	2.99
30x30-1-2-2	0.72	3.11
40x40-1-2-2	0.98	2.91

The minimum number of hits within a polygon had little influence on the resulting differences in height estimates as most polygons included in this study had many hits. This factor is largely included to not include zero values in the calculation of the grid based polygon average. Insufficient difference to require a table. The number of hits used to represent each cell did have a small impact upon resultant height estimates (Table 4). The use of at least two LIDAR hits to represent a cell, if possible, is recommended to counter the presence of unrepresentative large trees.

Table 4. Number of within cell laser hits used to represent a particular cell

Grid	LIDAR-Polygon	Standard deviation
40x40-1-1-2	1.44	2.95
40x40-1-2-2	0.98	2.91
40x40-1-3-2	0.74	2.94

The influence of the alteration of the minimum height required for inclusion of a point is illustrated in Table 5. The alteration of minimum height values appears to result in an increasing difference between estimates of height. As the minimum height required for inclusion as representative of a grid cell for the average calculation increases, the estimate for the polygon height also increases.

Table 5. Comparison of heights generated when the minimum height required to represent a cell is altered

Grid	LIDAR-Polygon	Standard deviation
40x40-1-2-0	0.12	2.77
40x40-1-2-1	0.27	2.78
40x40-1-2-2	0.98	2.91
40x40-1-2-3	1.03	2.91
40x40-1-2-4	1.12	2.98

Conclusions

When interpreting these results we caution that these findings are related to the characteristics of the sensor (such as the 9m footprint) and the forest structure present in central Saskatchewan. Height estimates, based upon an average of LIDAR hits within a polygon, are subject to a downward bias due to inclusion of non-canopy elements. The superposition of a grid over each polygon from within which estimates of height may be generalised results in an improvement of height estimates. Additional factors, such as minimum number of hits required, the number of hits to be average, and a minimum height requirement, also have an influence on resultant height estimates when averaging the grid values. Additional investigation is required to confirm these results.

Acknowledgements

Dr. David Harding of the NASA Goddard Space Flight Centre is thanked for data provision and valuable advice and encouragement.

References

Aldred, A. and M. Bonner, 1985; *Application of airborne lasers to forest surveys*, Canadian Forestry Service, Petawawa National Forestry Centre; Information Report PI-X-51. 62p.

Blair, J., Coyle, D., Bufton, J., and Harding, D., 1994; Optimization of an airborne laser altimeter for remote sensing of vegetation and tree canopies, *Proceedings of IGARSS '94*, Vol. II, pp. 939-941.

Gillis, M., and D., Leckie, 1993; *Forest inventory mapping procedures across Canada*, (Forestry Canada, PNFI, Information Report PI-X-114, 79p.).

Harding, D., 1998; Airborne LIDAR observations of canopy structure at the BOREAS tower flux sites, *Proceedings of IGARSS '98*, Seattle, pp. 1550-1552.

Harding, D., 2000; *BOREAS Scanning LIDAR Imager of Canopies by Echo Recovery: Level-3 Data*, Available online at [http://www-eosdis.ornl.gov/] from the ORNL Distributed Active Archive Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.

Lowe, J., K. Power, and M. Marsan, 1996; *Canada's forest inventory 1991: Summary by terrestrial ecozones and ecoregions*, Canadian Forest Service, Pacific Forestry Centre, Information Report BC-X-364E, 56 p.

Magnussen, S., and P. Boudewyn. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, Vol. 28, pp. 1016-1031.

Naesset (1997). Determination of mean tree height of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, Vol. 61

Nelson, R., 1997; Modeling forest canopy heights: The effects of canopy shape, *Remote Sensing of Environment*, Vol. 60, pp. 327-334.

Nilsson, M., 1996; Estimation of tree heights and stand volume using an airborne LIDAR system, *Remote Sensing of Environment*, Vol. 56, pp. 1-7.

Rowe, J., 1977; *Forest regions of Canada*, (Ottawa, Canadian Forest Service, 172 p.)

Sellers, P., F. Hall, H. Margolis, B. Kelly, D. Baldocchi, G. den Hartog, J. Cihlar, M. Ryan, B. Goodison, P. Crill, K. Ranson, D. Lettenmaier, and D. Wickland, 1995; The Boreal Ecosystem-Atmosphere Study (BOREAS): An Overview and Early Results from the 1994 Field Year, *Bulletin of the American Meteorological Society*, Vol. 76, No. 9, pp. 1549-1577.

Spurr, S., 1948; *Aerial photographs in forestry*, (Ronald Press Company, New York, 340p.)

St-Onge, B., 1999. Estimating individual tree heights of the boreal forest using airborne laser altimetry and digital videography. *Workshop on mapping surface structure and topography by airborne and spaceborne lasers.* November 9-11 1999, La Jolla (California), *ISPRS*.

Wulder, M., 1998; Optical remote sensing techniques for the assessment of forest inventory and biophysical parameters, *Progress in Physical Geography*, Vol. 22, No. 4, pp. 449-476.