Stability of surface LIDAR height estimates on a point and polygon basis

Mike Wulder [1*], Steen Magnussen [1], David Harding [2], Paul Boudewyn [1], and David Seemann [1],

[1] Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, BC
[2] Code 921, Geodynamics Branch, NASA Goddard Space Flight Centre, Greenbelt, MD, USA
[*] *Corresponding Author*

Abstract

LIDAR has been demonstrated as a tool for remotely sensing information on the vertical structure of forests. The Scanning LIDAR Imager of Canopies by Echo Recovery (SLICER) records data on canopy height, vertical structure, and ground elevation. Based upon the sensor configuration for this study, the vertical resolution of the SLICER is approximately 1m, with a horizontal resolution of approximately 9m, with five adjacent footprints resulting in an approximate 45m wide swath. Information on the height of trees within forest stands is an important attribute in forest inventories. The ability to remotely sense height information for forest inventory purposes may allow for procedures such as up-date, audit, calibration, and validation.

Prior to applying remote estimates of height in an inventory context the consistency of the estimates at locations and over areas is assessed. Locations which have more than one LIDAR observations from differing flight over-passes allow for an assessment of the stability of point height estimates. To assess the stability of area estimates, the height estimates from multiple flight lines through individual forest inventory polygons are compared.

For the boreal forest conditions present in our central Saskatchewan study area the following conclusions are made. On a point stability basis, LIDAR observations are found to vary little when separation distances between points are small. On a polygon basis, considering both between and within line standard deviations, the within polygon variability in LIDAR heights is well captured by collecting data over any portion of a polygon.

Introduction

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).

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).

In a previous study utilising SLICER, the relationship of remote height estimates to ground data was found to have an R^2 of 0.78. (Lefsky et al. 1999a; for additional background see Lefsky et al. 1999b and Means et al. 1999). The potential of LIDAR as a measurement tool for the collection of height samples for forest inventory usage is explored in this paper. The consistency of LIDAR estimates of tree height is a consideration for operational use of LIDAR data in forest inventory. Multiple LIDAR estimates at the same geographic location allow for assessing the stability of height estimates on a per point basis. Multiple LIDAR flight lines through individual polygons enables investigation of the stability of LIDAR estimates of polygon height. In this study we address the stability of large footprint SLICER data on a point and polygon basis.

[*]**Paper presented at: "*Remote Sensing and Spatial Data Integration: Measuring, Monitoring and Modeling*", 22nd Symposium of the Canadian Remote Sensing Society, Victoria, British Columbia, August 20th to 25th, 2000.**

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 localized 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. The GIS data provides the forest inventory polygon context within which we consider the consistency LIDAR observations from differing LIDAR data collection flight lines.

Remotely sensed data

To account for the differing vintage between the LIDAR data and the GIS data, we classified a Landsat TM image to provide an indication of current conditions within the polygons to aid in data stratification. The Landsat TM image, path 37, row 22, July 1994, was georectified to 30 x 30m pixel size using a first-order polynomial rectification, resulting in an RMS error of 0.80. We classified the Landsat TM imagery using a hyperclustering and labelling approach. An initial request of 241 clusters were merged down to 11 classes following the National Forest Inventory classification strategy, resulting in an accuracy sufficient for data stratification. The cover-types applied to each pixel are generalised to represent the current conditions of the co-georegistered forest inventory polygons.

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 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 geolocation of the height value for each LIDAR footprint, or hit, allows us to consider point to point overlap, and within polygon placement. 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.

---

[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*.

To undertake the point stability analysis, we recorded for all LIDAR hits the height of all points with a footprint location found within a 9m radius. The distance between the measured points from the reference point are also saved. The hits compared to the reference point are from different flight lines. 35,243 LIDAR hits from differing flight lines were found to overlap within the 9m distance. Tagging of pairs was applied to avoid double counting of individual pairs. The differences in the height measured from the differing over passes of the same location were then stratified by distance between locations to allow for observation of the relationship between height estimate and degree of overlap of LIDAR hits.

Graphical display of the point data, stratified by Landsat cover-type, enables visualisation of the relationship between height estimates and degree of hit overlap. Box plots, dividing the data into four areas of equal frequency, display the distribution of height differences by distance between footprint locations. The line in the centre of the box plots relates the median, a box encloses the middle 50 percent of the data, the lines extending from the box are drawn to values nearest 1.5 interquartile ranges from the quartile, with the points above indicating suspected outliers.

For the polygon stability analysis, 234,855 LIDAR hits are found to intersect with 513 individual polygons that have 2 or more flight-lines. For inclusion in a polygon, a LIDAR hit was required to be within a 9m buffer of the polygon boundary. The number of polygons with multiple flight lines decreases as the number of flight lines increases (Table 1). The results are presented only for polygons which have 2, 3, and 4 flight lines, due to the small number of polygons with greater than 4 co-occurring flight lines.

Table 1. Frequency of occurrence of flight lines collecting data on the same polygon

# flight lines	# polygons
2	366
3	89
4	32
5	6
6	10
7	5
8	3
9	1
10	1
Total polygons	513

An analysis of variance (ANOVA) was applied to all LIDAR height observations, stratified by polygon, to assess polygon height stability. The statistics of greatest interest for comparing the height estimates were the standard deviation between groups (lines) and the standard deviation within groups (lines). The standard deviations relate, in metres, the variability of the LIDAR heights, for each polygon. The within line results indicate the variability of height values that may be expected for an individual line within a polygon. Low values indicate little variety in height values, likely cover type related, such as for water or wetlands; while high values indicate height values representing tree tops, canopy openings, related to forested cover types. The between line results indicate the agreement of the variability of the lines passing through a polygon to one-another. Low values indicate (in meters) that while the lines are composed of a variety of heights the same variety of heights is captured by other lines through the polygon.

Results and Discussion

Point to point analysis

The box plots are used to illustrate the LIDAR stability when multiple measures are captured of the same geographic location. To stratify the LIDAR data by cover types we merged the LIDAR data with a classified remotely sensed image (Landsat TM). The polygons classified as water are flat, with no vertical structural component. In Figure 1a we illustrate that the between line variability is close to zero for the areas classed as water indicating that

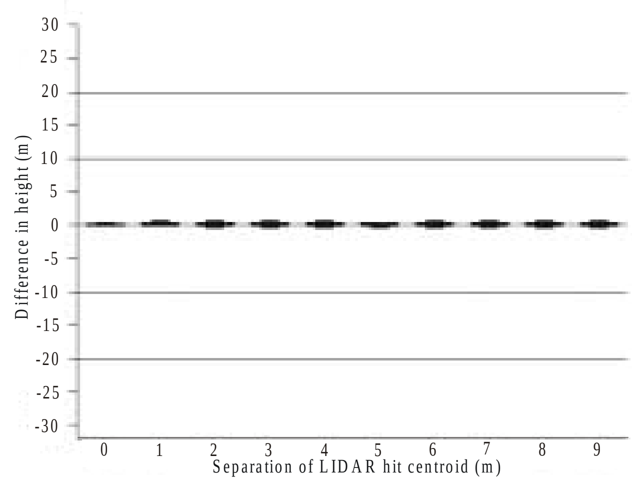

Figure 1a. Box plot illustrating variability of SLICER height estimates as a function of multiple measures of the same location and area, stratified as Landsat cover-type “Water”

multiple measures of the same location produce indistinguishable results. In Figures 1b to d we illustrate the variability in LIDAR heights as a function of forest cover-types. The compressed interquartile range of the coniferous observations,

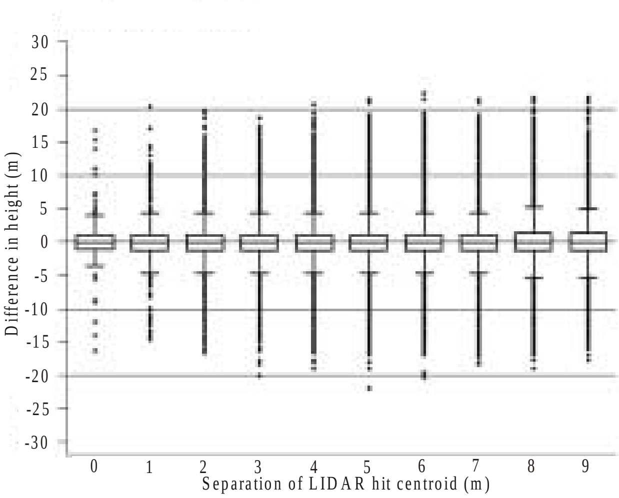

Figure 1b. Box plot illustrating variability of SLICER height estimates as a function of multiple measures of the same location and area, stratified as Landsat cover-type "Coniferous"

indicates low variability in LIDAR estimates of height from multiple measures of the same location (Figure 1b). Increased separation between coniferous observations result in uniform and constant box widths and interquartile ranges. In comparison, the observations for deciduous forest cover (Figure 1c) have less uniform and consistent box widths and interquartile ranges, likely indicating greater outer canopy structural variability. The mixed forest results are, as

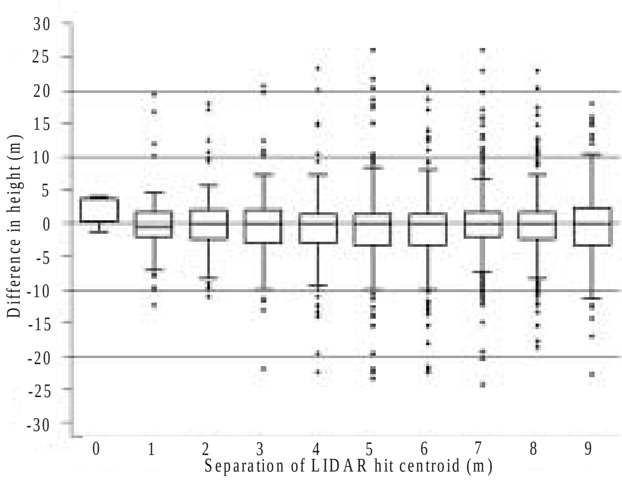

Figure 1c. Box plot illustrating variability of SLICER height estimates as a function of multiple measures of the same location and area, stratified as Landsat cover-type "Deciduous"

expected, intermediate between the conifer and deciduous results (Figure 1d). Any variability in LIDAR heights is likely due to surface cover variability and/or geolocating of the LIDAR hit.

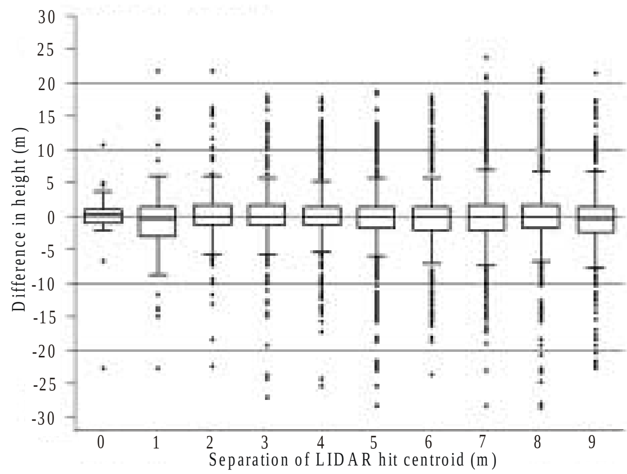

Figure 1d. Box plot illustrating variability of SLICER height estimates as a function of multiple measures of the same location and area, stratified as Landsat cover-type "Mixed wood"

Within polygon analysis

The within polygon stability analysis is undertaken through an analysis of variance generating the within line and among line height variability of polygons. With these data, we are able to investigate the stability of average height estimates for a polygon as a function of the number of lines characterising an individual polygon. The multiple lines through the polygon are not overlapping, as a result, multiple polygon measures are made from differing land cover within the polygons.

To act as a comparison, water, with a maximum of two flight lines going through three large polygons, having an average number of 5919 hits, a between line mean standard deviation of 0.15m and a within line mean standard deviation of 0.69m were found. The low values for both indicate that there is little to no variability seen between and within lines over the flat water surface. As a result, differences noted for vegetated surfaces are a function of variation sensed in the surface cover and how this cover is characterised by the LIDAR data. The tables illustrating within and between polygon standard deviation results are presented over all cover-types. Stratification of the data provides little additional explanatory information. Of the 513 polygons with multiple flight lines, 385 are classed as coniferous, 373 of which have 2, 3, or 4 flight lines.

Overall all classes, the between line results indicate, for example, that when 2 flight lines are flown over a polygon the lines are within 1 meter of agreement 50% of the time and within 2 metres 75 % of the time (Table 2). In Saskatchewan height is often measured to the nearest 5 meter class indicating the LIDAR accuracy as more than sufficient.

The within line variability indicates that there is a range of height values sensed for individual flight lines (Table 3). For example, the within line mean standard deviation for 2 lines through a polygon is 3.67m[2]. Therefore, if there is high within line variability and low between line variability, each LIDAR flight line through the polygon is sensing variable yet similar information. Collecting data over differing areas of individual polygons captures unique data which is appropriate to represent the entire polygon. The basic tenet indicated for operational data collection purposes is that if you wish to sample unique information and avoid wasting time and money on redundant data a single flight line is appropriate.

The stability of polygon height estimates allows for the use of LIDAR as a sampling, rather than as a mapping tool. In an inventory context, a lack of dependence upon actual area sensed within polygons to provide a valid estimate of height is useful. Sampling will allow for reduced data acquisition costs and time requirements.

Conclusions

In previous studies the estimation of height data has proven effective. The consistency of LIDAR data is of concern for forest managers considering the application of LIDAR data to forest inventory data collection. The consistency of the SLICER height data, on both a point and polygon basis, illustrate the utility of LIDAR in forest inventory surveys.

Recording and evaluating multiple LIDAR hits around the same location indicates that variability in LIDAR heights is limited when the separation between observations is small.

In consideration of the boreal forest cover of our central Saskatchewan study area, estimates of canopy height from multiple flight lines through individual polygons indicate an acceptable level of variability related to the path taken over the polygon. The between line standard deviations relate, in the case of 2 flight lines through a polygon, that height estimates are more often within 1m than not. The within line standard deviations indicate that there is variability of height values collected within each line. Considering both between and within line standard deviation results, it may be concluded that within each polygon there is variability in LIDAR heights and that this variability is well captured by collecting data over any portion of a polygon. Our results, on a point and polygon basis, indicate the utility of LIDAR data as a sampling tool in a forest inventory context.

[2] Note that the variances are likely biased downwards due to a positive spatial autocorrelation of within line canopy heights.

Acknowledgements

Xilin Fang of the Forestry Branch of the Saskatchewan Parks and Renewable Resources for provision of the forest inventory data.

References

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.

Harding, D., J. Blair, D. Rabine, and K. Still, *In Press*; *SLICER Airborne Laser Altimeter Characterisation of Canopy Structure and Sub-canopy Topography for the BOREAS Northern and Southern Study Regions: Instrument and Data Product Description*, NASA Technical Memorandum NASA/TM-2000-209891

Lefsky, M., D. Harding, G. Parker, and H. Shugart, 1999, Surface LIDAR remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA, Remote Sensing of Environment, Vol. 67, pp. 83-98.

Table 2. Relative frequencies of between line standard deviations of canopy heights in GIS polygons

		Between line standard deviation (m)							
		0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	> 5	count	mean
Number of lines per polygon	2	51.1%	24.6%	13.9%	6.8%	1.6%	1.9%	366	1.34m
	3	41.6%	30.3%	9.0%	12.4%	2.2%	4.5%	89	1.67m
	4	25.0%	50.0%	9.4%	9.4%	3.1%	3.1%	32	1.78m

Table 3. Relative frequencies of within line standard deviations of canopy heights in GIS polygons

		Within line standard deviation (m)							
		0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	> 5	count	mean
Number of lines per polygon	2	3.0%	10.1%	25.7%	24.9%	15.6%	20.8%	366	3.67m
	3	1.1%	4.5%	14.6%	24.7%	27.0%	28.1%	89	4.26m
	4	0.0%	15.6%	15.6%	21.9%	28.1%	18.8%	32	3.79m

Lefsky, M., Cohen, W., Acker, S., Parker, G., Spies, T., and Harding, D., 1999; LIDAR remote sensing of biophysical properties and canopy structure of forests of Douglas-fir and western hemlock, Remote Sensing of Environment, Vol. 70, pp. 339-361.

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.

Means, J., S. Acker, D. Harding, J. Blair, M. Lefsky, W. Cohen, M. Harmon, W. McKee, 1999; Use of a large-footprint scanning airborne LIDAR to estimate forest stand characteristics in the western Cascades of Oregon, *Remote Sensing of Environment*, Vol. 67, pp. 298-308.

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

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*, in press.

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.

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.