Evaluation of Emerging Remote Sensing Technologies for the Assessment of Forest Resources.

Submitted by
K. Olaf Niemann
Department of Geography
University of Victoria
olaf@uvic.ca

Submitted to

B.C. Forest Sciences Programme,

Ministry of Forest and Range, Government of British Columbia,

and

Canfor Forest Products Ltd.



Acknowledgments

This project has matured over the past 3 years and has taken on a life of its own. It has consumed the hours / lives of a number of dedicated people all who are authors of this work and should be recognized as such. These include:

Dr. Gordon Frazer whose expertise on modelling is embodied in this work, and is co-author of this report;

Rafael Loos who developed the software allowing us to process the volumes of data that was collected;

Fabio Visintini, our spectroscopist, who worked out the nuances of the hyperspectral data;

Terra Remote Sensing Inc. who engineered and flew the "Multisensor Airborne Platform" (MAP);

Melanie Boyce, John Wakelin, and Xiaoping Yuan, BC Ministry of Forest and Range;

Dave Dobi (Canfor) and Paul Jeakins (formerly Canfor), financial and logistical support;

VRI field data support was provided by MoFR. Special thanks to Chris Butson;

Field work was completed by Gordon Frazer, Chris Butson, Roman Bilek, Bill Kovacs, Matt Makar, Martina Szabova, Laura Duncanson, and Rafael Loos; and

Funding through the BC Forest Sciences Programme (BC-FSP) is also gratefully acknowledged.

Executive Summary

This project has addressed a number of issues related to the potential operational use of two emerging remote sensing technologies in the area of forest resource inventory and sustainable forest management. These airborne technologies include Light Detection and Ranging (LiDAR) and hyperspectral imaging spectroscopy. The project was carried out on Canfor's Tree Farm License (TFL) 18, north of Clearwater, B.C .

The overall goal of the project was to evaluate LiDAR and hyperspectral technologies, independently and together, as potential contributors to the extraction of information relevant to forestry and forest sciences . The specific objectives were:

- To assess the state of the tools available for the processing of large volumes of LiDAR and hyperspectral data;
- To develop, where necessary, tools to fill processing voids;
- To apply these tools to the processing of the data collected for a large spatially contiguous block of data representing a forest management unit; and
- To evaluate the results within the context of current BC Provincial forest inventory standards and procedures (Vegetation Resource Inventory – VRI).

Data were acquired over a 4-day period using a multisensor instrument platform onboard a fixed wing aircraft. This is not an experimental design and has subsequently been used in numerous data acquisitions both in British Columbia and elsewhere.

A comprehensive processing framework and toolset was developed to address the needs of the project. Where possible we used commercially available software and developed processing tools to address any encountered deficiencies. The outcome of this assessment was that while a number of tools to preprocess data to prepare them for further analysis were available commercially, further processing to extract forest attributes required the development of a suite of custom tools.

The LiDAR data were processed to derive a number of initial products. These included a bare earth model representing a digital elevation model without the effects of vegetation, a canopy height model representing terrain-normalized vegetation height, and a series of metrics representing vertical and horizontal descriptors of the vegetation canopy. This set of metrics is based on the processing of the LiDAR vegetation returns

within a user specified grid cell on a quantile basis.

The hyperspectral imaging spectrometer data were initially processed to remove the effects of the atmosphere and illumination. The data were then orthorectified to accurately position the data, and remove the effects of relief/vertical displacement. For the latter a digital surface model representing the three dimensional (3-D) reflective surface was used. The result was that the hyperspectral data were positioned to within plus/minus 1 pixel.

Field data were collected for a single 1:20000 map sheet (BCGS 92P.069). These data were used as a training set to calibrate the LiDAR-based predictive models and hyperspectral information extraction. A novel technique to address the issue of sample selection based on using LiDAR canopy height data was developed. This technique is based on a stratified random sampling approach whereby landscape units (strata) are defined by differences in vertical canopy structure (height and density). This simple approach helps to insure that ground-reference plots cover the full range of structural variability found in the landscape. This approach provides the user with an oversight of the distribution of structural elements within the canopy, and ensures that the sample design optimizes for a both representativeness and sample design.

The LiDAR biometrics were modelled based on the attributes collected during the field work. The attributes included tree height, diameter, volume, and biomass. The results suggest that we can relate the field-measured attributes to the LiDAR biometrics with a high degree of correlation. A range of r^2 results from 0.95 to 0.86 was obtained, depending on the complexity of the attribute modelled. These models have been spatially extrapolated to include the entire study area.

The hyperspectral data were used to address the question of species composition. One of the drawbacks with traditional analysis of remotely sensed images is that when analyzing an image matrix environment we typically do not extract spatially explicit information. The approach used in this project derived an initial representation of individual treetops using the LiDAR data. Each of these treetops was stored in a tree list where the location and height of each of the dominant - codominant trees were retained. These treetops were subsequently mapped onto the hyperspectral data and a single spectrum representing the top of the tree was extracted. These individual spectra retained

the positional information of the treetop so that we could map them back into geographic space once the analysis was complete. The spectra were classified using a supervised Spectral Angle Mapper classification approach. An assessment of accuracy was performed using a *leave-one-out cross-validation* approach on the training data. The resultant overall classification accuracy was 76% (Cohen's Kappa = 0.65). An assessment of the specifics of the classification showed that the greatest confusion occurred between the spruce and fir species. The classified spectra were grouped into a 20 metre grid (based on the same grid structure that was applied to extract the LiDAR biometrics). The grouping that we applied was based on the leading (by frequency) species.

A comparison between the Vegetation Resource Inventory (VRI) surveys completed for the single 1:20000 map sheet suggests a general correspondence between the nature and range of attributes derived by both types of surveys. The advantage of using both the LiDAR and hyperspectral data, however, is that we can derive and map continuous variations in surface cover and related attributes.

While the work successfully identified a number of attributes that could be extracted, we also identified a number of issues that need further work. For the LiDAR data there is a need to further develop the metrics that describe the forest environment. Directions that are presently being considered relate to the modelling of site index integrating the very high information content of the bare earth model and potential evapotranspiration models and the application of the outputs to Predictive Ecological/Ecosystem Mapping (PEM). An additional model is one dealing with wood quality (based on wood density) from LiDAR/hyperspectral biometrics and site modelling. The hyperspectral data analysis is, in many ways, less mature and more complex than LiDAR. One of the issues addressed in this project is that of species classification. This area needs further elaboration, especially in addressing more effective methods to partition the feature space and addressing variations in intensity differences caused by influences not accommodated by atmospheric modelling. Other analysis includes the mapping of chlorophyll, and by extension foliar nitrogen. While the modelling processes of the spectral features to extract these attributes are important we need to address, more fundamentally, the noise characteristics of the data. This is crucial

as the forest environment is a relatively low reflector of energy, and as such the noise component of the data becomes increasingly important. If we can characterize the noise more effectively then it becomes less of an issue in low illumination environments.

The final area of research identified is that of newer technology and the potential for application to forest related problems. This is primarily focused on higher frequency LiDAR systems (currently up to 260 KHz.) and high spatial and spectral resolution imaging spectrometers. The increase in resolutions will increase the challenges in feature extraction, but also allow for a much higher potential to extract meaningful features. At the other end of the resolution spectrum are satellite-base systems. There are currently plans for global coverage of hyperspectral data with resolutions approaching .25 Ha. In addition, missions with large footprint full waveform systems are also planned. The integration of these two systems will allow us to accurately model regional to global changes in biomass quickly and relatively inexpensively.

Table of Contents

1. Introduction - Background, rationale, and objectives.1

2. Methods 1

2.1. Study Site Description 1

2.2. Remote Sensing Data Collection 2

2.3. Remote sensing data processing framework 8

2.4. LiDAR Data Processing9

2.4.1 Post-flight processing, calibration, and validation 9

2.5. Hyperspectral processing 22

2.5.1. Orthorectification of hyperspectral imagery 22

2.5.3. Image analysis and Classification 23

2.6 Grid-based, 2-stage inventory proceedure 28

2.6.1. Selection of ground-reference sites 31

2.6.2. Ground-reference sample 32

2.6.3. Predictive modelling and spatial extrapolation of inventory attributes 39

2.7. Extraction of forest inventory attributes 40

2.7.1. Comparison of VRI and remotely sensed datasets 40

3. Results & Discussion 41

3.1. LiDAR-based inventory attributes 41

3.1.1. Ground-reference sampling 41

3.1.3. Map sheet and TFL inventory statistics 53

3.2.1. Spectral-based species classification 69

3.3. Comparison of VRI and remotely sensed inventories 3.3.1. Structure-based inventory results 73

4. Conclusions and Management Implications 80

4.1. Summary of key findings 80

4.2. Future directions 81

4.2.1. Hyperspectral Data: 81

4.2.2. LiDAR 82

4.2.3. Sensor Development83

5. REFERENCES 85

1. Introduction - Background, rationale, and objectives.

Remote sensing technologies through the miniaturization of electronics have, over the past few years, allowed for the development of small relatively low cost instrumentation. Size and cost reductions have allowed us to develop sensor platforms that integrate a variety of sensors. Due to the reduced size and power requirements of the sensors, smaller and thereby less expensive, platforms can be utilized. This has resulted in unprecedented opportunities for data collection.

The present project was initiated to evaluate two emerging technologies, Light Detection and Ranging (LiDAR) and hyperspectral imaging, to collect data and extract information pertaining to forested environments. The term *emerging* is used in this case as the technologies are not new, but are now increasingly being considered technically and economically viable for more routine (i.e. non research) collection of data.

The overarching goal of this project was to evaluate hyperspectral - LiDAR data to address questions related to forest inventory in terms of compatibility with more traditional approaches to inventory. Given cost, time, and accuracy constraints associated with traditional methods of forest inventories, the question as to whether these technologies could act as alternative, or complimentary, information sources to alleviate these constraints was obvious. The research objectives for this project are focused on: 1) the development of data processing tools to allow us to evaluate the potential of these technologies for forest science and forestry applications, and 2) extraction of forest inventory attributes as a direct comparison to those data derived from more traditionally applied approaches. This report details the work carried out towards the realization of the primary goal of the project, focusing on the two principal objectives mentioned above.

2. Methods

2.1. Study Site Description

This project was conducted on Tree Farm License (TFL) 18, located in central British Columbia (North Thompson region) near Clearwater, BC (Figure 1). The TFL is held by Canadian Forest Products (Canfor) Limited, Vavenby Division, and administered

by the BC Forest Service, Headwaters Forest District Office, Southern Interior Forest Region, Clearwater, BC. TFL 18 is approximately 74,545 ha (745.45 km^2) in area, with approximately 90% of the land base classified as productive forest. The terrain is characterized by gently rolling hills and high-elevation plateaus, with elevations ranging between 516 and 1995 m.a.s.l (Figure 2). Numerous rivers, lakes, and wetland complexes are found throughout the TFL. Three different biogeoclimatic (BGC) zones occur within the TFL: the Engelmann Spruce-Subalpine Fir (ESSF) zone (approximately 51% of total area), Sub-Boreal Spruce (SBS) zone (32%), and the Interior Cedar-Hemlock (ICH) zone (16%). The main commercial tree species are *Picea engelmannii* (Engelmann spruce), *Picea glauca* (white spruce), *Pinus contorta* (lodgepole pine), *Abies lasiocarpa* (subalpine fir, also known as balsam), *Pseudotsuga menziesii* (Douglas-fir), *Tsuga heterophylla* (western hemlock), and *Thuja plicata* (western redcedar). Approximately 48% of the stands are spruce leading, followed respectively by pine (26%) and balsam (20%). Forest age classes within the TFL range from newly harvested to greater than 260 years: 0–40 yr. (31%); 40–80 yr. (12%); 80–140 yr. (24%); 140-260 yr. (27%); >260 yr. (6%). The TFL is currently experiencing a mountain pine beetle (MPB) epidemic, with approximately 3,460,640 m^3 of mature pine volume susceptible to attack (BC Ministry of Forests and Range, 2006).

2.2. Remote Sensing Data Collection

Airborne LiDAR and hyperspectral data were acquired over 647 km^2 of the total 745 km^2 area of the TFL between August 16 and 19, 2006[1], using the jointly developed UVic/TRSI[2] Multi-sensor Airborne Platform (MAP) (see Figure 3). MAP is designed to simultaneously acquire hyperspectral imagery, 3-D LiDAR points, and very high resolution colour digital photography. The sensor suite is installed on TRSI's multi-engine, fixed-wing (Piper Navajo) aircraft. The principal advantage of co-locating these

[1] Only partial remote sensing coverage of the TFL was acquired due to a persistent haze caused by forest fires in northern Washington State.
[2] Terra Remote Sensing Incorporated.

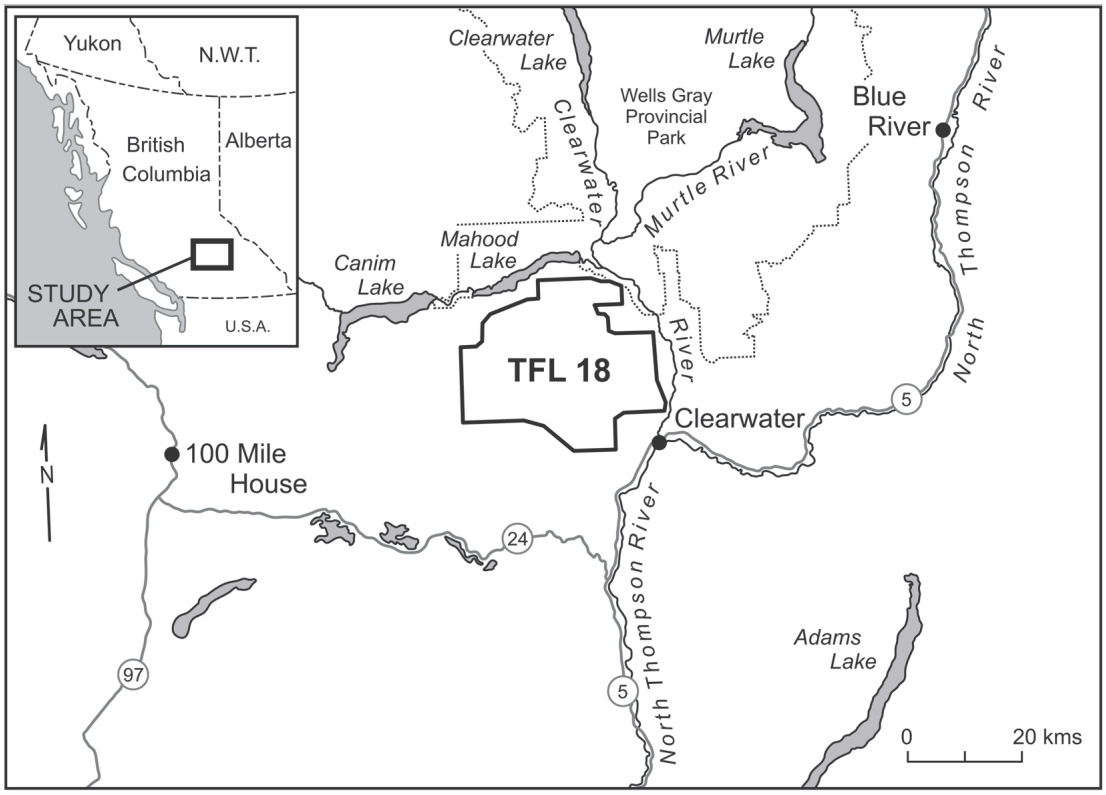


Figure 1. Location of TFL 18 used as a test site for this study.

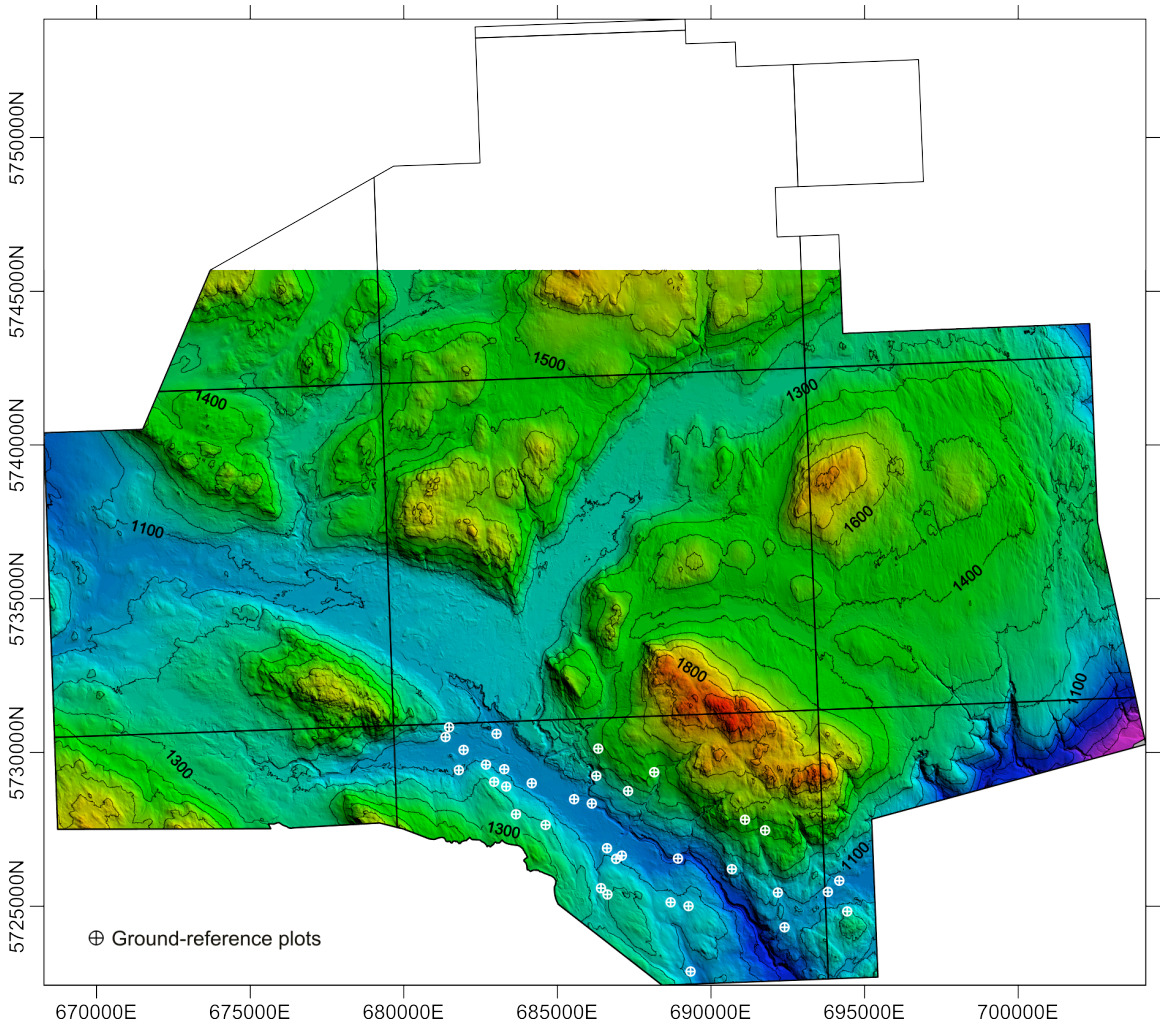


Figure 2. LiDAR generated digital elevation model (DEM) constructed from a triangulated irregular network (TIN) of LiDAR ground points interpolated onto a 1 m regularized grid. The primary purpose of the DEM is as a baseline elevation surface to compute LiDAR canopy heights and other derivative metrics; however, the DEM can serve many other purposes related to hydrological modelling, slope stability, road construction and decommissioning, orthorectification of optical imagery, as well as being the source of many other plant-relevant explanatory variables needed to support growth and yield, site index, ecosystem, wood quality, and habitat modelling.

sensors on the same airframe is that many of the post-flight data processing steps can be integrated and streamlined for improved efficiency. Additionally, fitting multiple sensors on a single platform reduces costs and simplifies mission planning.

The complete sensor suite for the airborne data acquisition includes a full-range imaging spectrometer, scanning LiDAR, and two small-format, true-colour (RGB) digital cameras. The hyperspectral instrument (Specim AISA Dual) is constructed of two separate spectrometers, one covering the visible/near infrared wavelengths (VNIR: 400-1000 nm), while the second images in the shortwave infrared (SWIR: 1000–2500 nm). Together, these two spectrometers produce 498 discrete spectral bands integrated into a single image cube. The LiDAR instrument is a scanning, small-footprint (< 1 m), discrete-return instrument (designed and built by TRSI), capable of recording the first and last return of each reflected laser pulse. The laser has an operational wavelength of 1060 nm and maximum pulse rate of 150 kHz (or 150,000 pulses per second). The dual small-format digital cameras collect colour photography at centimeter spatial resolution. An onboard Global Positioning System (GPS) receiver and an Inertial Measurement Unit (IMU) acquire all of the positional information, including attitude (i.e., roll, pitch, and heading) of the aircraft, during the survey flight. One of the metadata attributes that is recorded by all instruments is a very high resolution time stamp. This feature is critical for accurate georeferencing and spatial integration of the hyperspectral, LiDAR, and photographic datasets.

The spatial resolution of the multi-sensor remotely sensed data is determined by the electronics of the individual sensors, as well as by numerous flight and survey parameters (Table 1). The airborne mission for this current project required approximately 12 hours of total flying time over the span of four consecutive days to collect all of the remotely sensed data for this project. The sensor platform was flown at approximately 1600 m above the ground surface at airspeeds of 250 km/hr. The LiDAR system operated at a laser pulse rate of 60 kHz, scan speed of 37 Hz, and maximum scan angle of ±20 degrees. This yielded an average LiDAR point density of 1.22 points/m^2 (± 0.44 points/m^2 at one standard deviation), 2 m spatial resolution for the hyperspectral imagery, and colour photography with a 25 cm spatial resolution. Horizontal positional

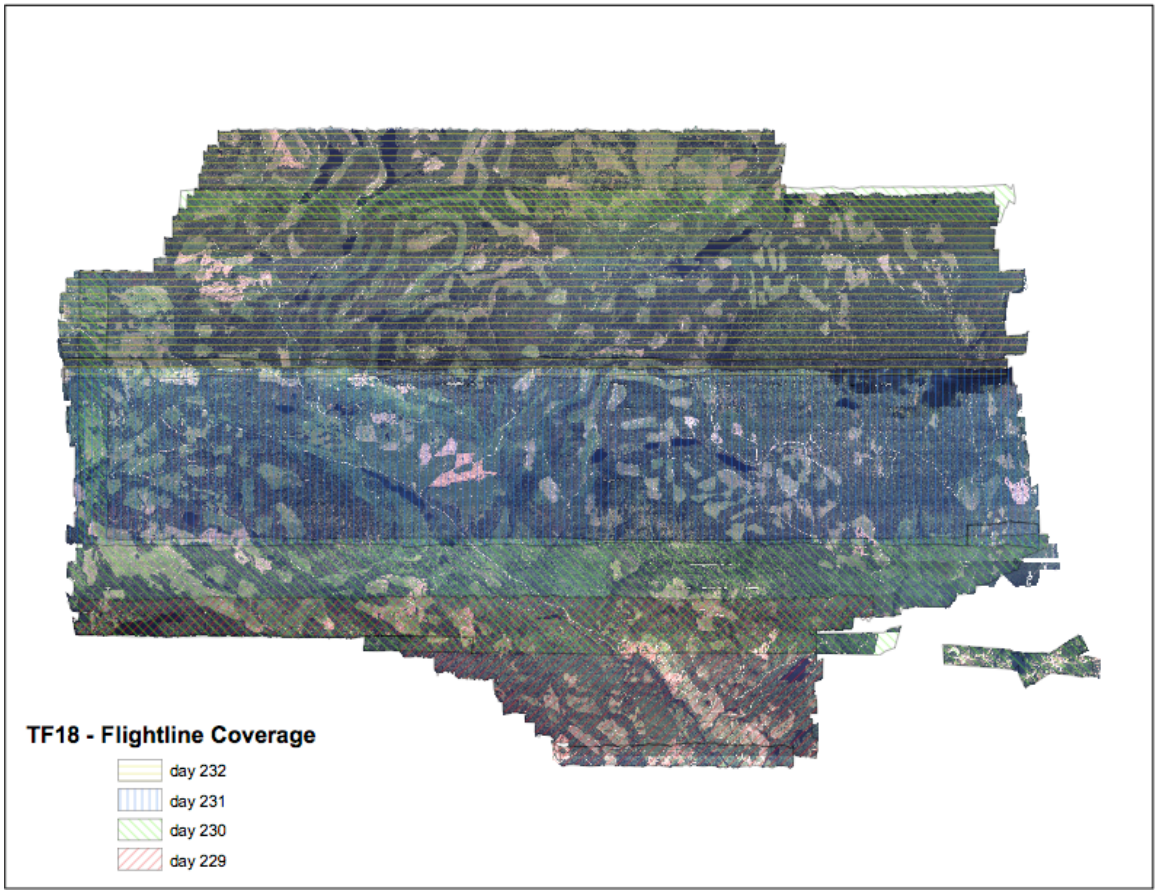


Figure 3. Data acquisition by day for TFL 18.

Table 1: Summary of instrument settings and survey parameters used for remote sensing data acquisition over TFL 18.

Flight Date	Aug. 16, 2006	Aug. 17, 2006	Aug., 18, 2006	Aug., 19, 2006
Start Time (Hr.)	20:38:57	17:13:29	17:33:48	16:48:35
End Time (Hr.)	22:38:03	20:10:16	20:27:42	20:54:39
Weather	Partial OC (<10) / Hazy	Partial OC (<10) / Hazy	Partial OC (<10) / Hazy	Partial OC (<10) / Hazy
Flying Height (m)	1600	1600	1600	1600
Air Speed (km/hr)	250	250	250	250
GPS PDOP*	< 2	< 2	< 2	< 2
Laser Pulse Repetition Rate (kHz)	60	60	60	60
Max. Laser Scan Angle (°)	±20	±20	±20	±20
Laser Scan Rate (Hz)	37	37	37	37
Average Laser Point Density (points/m^2)	1.22	1.22	1.22	1.22
Nominal laser footprint diameter (cm) @ 1600 m	72	72	72	72
Hyperspectral Spatial Resolution (m)	2	2	2	2
Hyperspectral Spectral Resolution (nm)	VNIR: 2.9 (244 bands) SWIR: 5.8 (254 bands)	VNIR: 2.9 (244 bands) SWIR: 5.8 (254 bands)	VNIR: 2.9 (244 bands) SWIR: 5.8 (254 bands)	VNIR: 2.9 (244 bands) SWIR: 5.8 (254 bands)
Digital Camera Spatial Resolution (cm)	25	25	25	25

*Positional dilution of precision (PDOP) indicates the accuracy of a 3D GPS position based on the number and geometry of satellites at the time of survey (4 or less required for survey grade accuracy).

accuracy was estimated by the service provider (TRSI) to exceed 1 metre, while relative and absolute vertical accuracies of the LiDAR data were 15 cm and 30 cm, respectively. All remotely sensed data were georeferenced using the UTM Zone 10 North projection, North American Datum 83 (NAD83), and HT 2.0 geoid.

2.3. Remote sensing data processing framework

A number of technical challenges arose related to the data processing requirements of this project. The first was simply the storage and handling of a remotely sensed dataset the size and complexity of the one produced by MAP. The hyperspectral sensor, alone, generated approximately 1 GB of data per minute of operation, while the LiDAR produced nearly 1 billion 3-D points plus associated attributes (e.g., GPS time stamp, horizontal and vertical position, intensity, return class, etc.) over the entire survey area. The challenge, therefore, was to develop a processing environment that could deal with these large data volumes using commonly available computing resources. The second challenge was to design and develop a data processing framework under which we could convert these large and disparate remotely sensed datasets into more meaningful forest measurements. Prior to starting this project there existed no commercially available data processing environment that allowed us to take advantage of the combined sensor payload of MAP. Consequently, much of our research effort has been directed towards the design and development of analytical procedures and software-based toolsets.

We chose ITT Visual Information Solutions (ITT VIS) Interactive Data Language (IDL) as our main software development environment, because of its long history[3] of use in array-oriented image processing, analysis, and visualization. IDL is also the development environment on which ITT VIS ENVI (a popular commercial remote sensing application) was built. Our software toolset has been developed for cross platform use with Linux, Windows, and Macintosh, and compiled as a suite of standalone run-time executables. Image output formats are in GeoTIFF and ENVI Standard formats, which allow direct access to more advanced ENVI functionality and other commercial

[3] IDL was originally developed in the 1970s at the Laboratory for Atmospheric and Space Physics (LASP) at the University of Colorado at Boulder, and was used by LASP scientists to interpret data from the Mars Mariner 7 and Mariner 9 missions.

CAD and GIS applications. The following sections summarize the processing framework developed for these simultaneously collected LiDAR and hyperspectral datasets (Figure 4). This processing framework was designed to be modular so that enhancements to the individual processing components can be made while not affecting others along the processing thread. In all cases the I/O is retained so that the flow is not compromised.

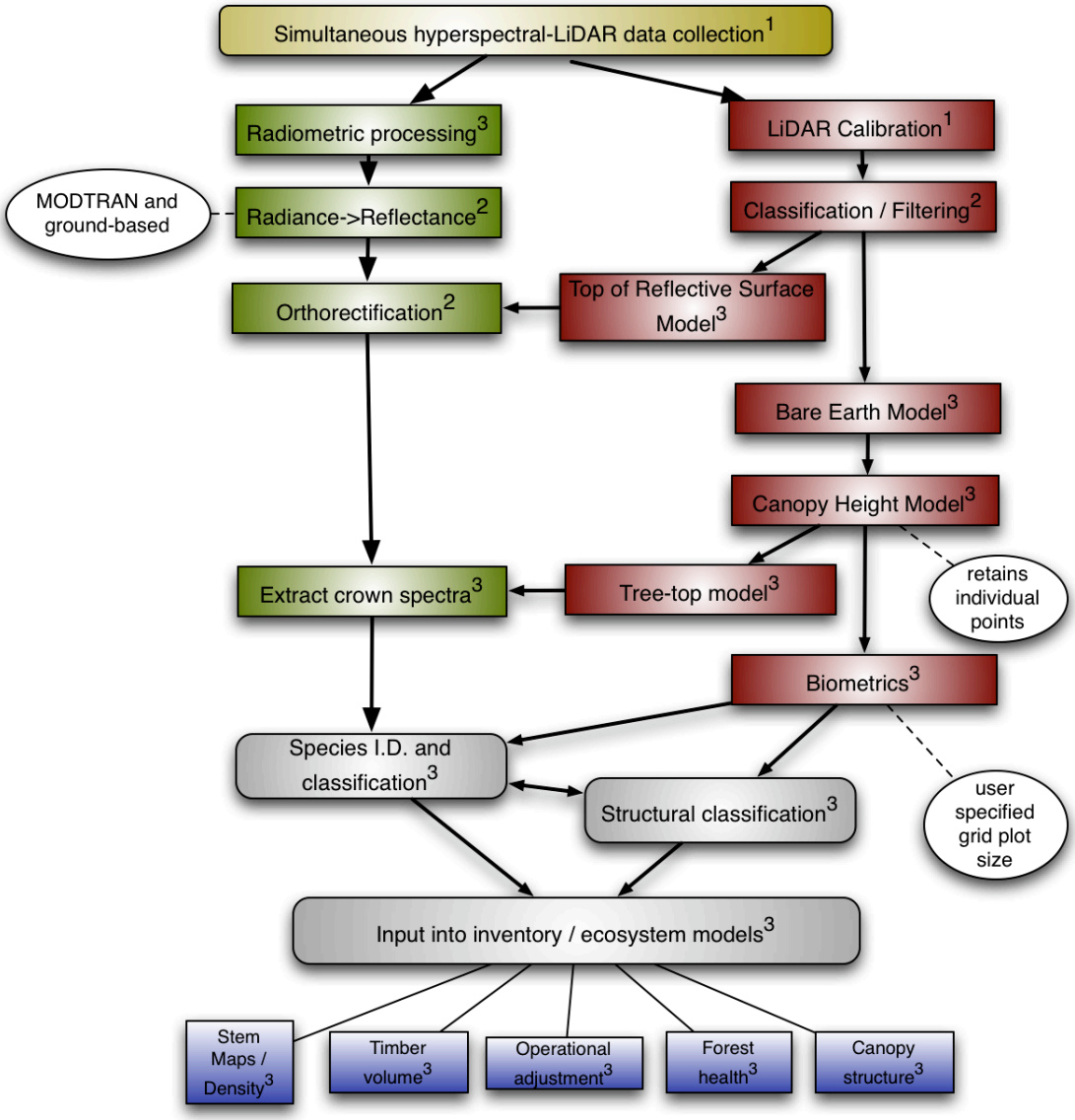
2.4. LiDAR Data Processing

2.4.1 Post-flight processing, calibration, and validation

Initial post-flight processing, calibration, and validation (i.e., quality assurance and quality control) of the LiDAR dataset were carried out by the data service provider (TRSI). Post-flight processing involved the conversion of raw laser range (distance), scan angle, GPS, and IMU measurements to georeferenced 3-D points (xyz) projected in a known coordinate space (UTM Zone 10 North, NAD83, HT 2.0). Georeferenced points were then assessed for survey-wide (gross) geometric distortions (i.e., positional inaccuracies and outliers), and any between-flight-line differences in point-cloud geometry (step errors) adjusted to allow precise 3-D spatial matching of LiDAR flight lines and points across the entire survey area. Once geometrically corrected and validated, the LiDAR dataset was released for further processing (i.e., classification, filtering, tiling, digital surface modelling, and attribute extraction).

2.4.1.2. Point classification and filtering

We relied on commercially available software, Terrasolid TerraScan (http://www.terrasolid.fi/), running on top of Bentley MicroStation CAD V8 (both Windows applications) to classify and filter the LiDAR dataset. Classification involves the separation of individual LiDAR points into ground (bare-earth) and non-ground (vegetation) classes using an automated slope-based filtering algorithm (Vosselman, 2000). The algorithm is iterative and begins by binning the entire LiDAR point cloud into a regularized grid of square cells of user-defined dimension (50 m). The lowest point



[1]Terra Remote Sensing;[2] Commercial software (TerraSolid- Terra Scan, ATCOR4, Caligeo); [3] Custom IDL run-time executables

Figure 4. LiDAR and hyperspectral data processing framework developed using a suite of commercially available software and custom applications written in Interactive Data Language (IDL). LiDAR data have been used for (1) modelling structural attributes, such as stand height, basal area, volume, and biomass; (2) orthorectification of the hyperspectral imagery; and (3) identification of individual tree tops for height-guided sampling of crown-scale radiance spectra. We relied on the continuous (400-2450 nm) spectra collected by the hyperspectral sensor for species classification at the individual tree scale.

(minimum elevation) in each grid cell is then used to construct an initial triangulated irregular network (TIN) of the ground surface. Each remaining unclassified point is subsequently evaluated for its angle and height above the triangle plane, and those within a certain threshold distance (height) and angle above the plane join the TIN. The main assumption here is that points exhibiting large increases in elevation over short horizontal distances are more likely due to variations in vegetation (or other feature) height rather than topography. The routine continues until all possible ground points are incorporated into the TIN and a complete bare-earth model has been formed. Finally, unusual points (outliers) isolated by elevation (i.e., those found well below or above the local ground surface) are filtered and removed from the LiDAR point cloud as error points.

2.4.1.3. Use of data tiles or blocks

The segmentation of LiDAR datasets into spatially discrete data tiles or blocks is a necessary part of LiDAR data processing, because of the excessive random access memory (RAM) requirements needed to read, store, and manipulate large point clouds. The logic we applied to tile the LiDAR dataset was based on a number of processing requirements. First, the tile origin and extent needed to be compatible with the origin and spatial scale (i.e., pixel size and extent) of the hyperspectral imagery. Second, the tile geometry and layout needed to conform to a 2-D Cartesian coordinate system, so that individual LiDAR points, cells (pixels), and tiles could be spatially indexed using a single pass through the dataset. Last, the tiling system needed to accommodate a multi-scale nested design that maintained the spatial integrity and continuity of individual pixels within and between tiles. We employed two tile sizes to accomplish this: (1) ground points were segmented into 4000 m (16 km^2) square tiles, and (2) all points (both ground and non-ground classes) were placed into 1000 m (1 km^2) square tiles. Both tile sizes shared the same origin point (i.e., outside edge of the lowest left-hand pixel in the hyperspectral image array), so that four 1000 m (all points) tiles fall neatly within the bounds of one 4000 m (ground points) tile, and 2.5×10^5 2 m hyperspectral image pixels fit exactly within a 1000 m tile. We added an arbitrary 50 m of overlap to both the 1000 m and 4000 m tiles prior to file generation, as a way to avoid any edge effect that might

hamper other downstream processes (i.e., construction of the bare-earth and canopy height models, and single-tree isolations).

2.4.1.3. Extraction of digital surface models

We built four different digital surface models (DSM) using the classified, filtered, and tiled LiDAR points, each with a different spatial resolution (grain rather than extent) and spatial data structure (i.e., raster, vector, and TIN). First, we constructed a gridded (raster) 2 m spatial resolution top-of-reflective-surface model (TORS) using the highest point (maximum elevation) found within each 2 m cell. The TORS shared the same pixel size and range rectangle (i.e., origin and extent) as the hyperspectral imagery so that each image set would overlay perfectly one upon the other. The TORS was used exclusively to orthorectify (planimetrically correct) the hyperspectral dataset (details of the orthorectification process are given below).

Second, we created a 1 m raster bare-earth or digital elevation model (DEM) for use as a standalone map product (Figure 5) and as a baseline surface to convert canopy elevations above mean sea level to local heights above the ground. This was a two step process: (1) we reconstructed the original TerraScan bare-earth TIN using all of the LiDAR points classified as ground; (2) we used an exact fitting, C_1 linear interpolator to extract a regularized 1 m DEM from the TIN. We used triangulation with a linear interpolation rather than other popular methods of interpolation (e.g., kriging and inverse distance weighting; Lloyd and Atkinson, 2002), because of its speed, efficiency, and direct compatibility with TerraScan's ground classification algorithm and TIN data structure. Third, we extracted a canopy height model (CHM) by (1) overlaying all of the classified LiDAR points on top of the 1 m DEM, (2) estimating the ground surface elevation (Z_{Ground}) located immediately beneath (or in the rare case above) each classified LiDAR point (Z_{LiDAR}) using a bilinear interpolator (IDL INTERPOLATE function), and finally (3) differencing the two elevations (Z_{LiDAR} minus Z_{Ground}) to compute a canopy height in metres above the ground surface for each classified point in the dataset (Figure 6). The CHM served two main purposes: (1) it provided the basic spatial information needed to compute other LiDAR biometrics and synthetic variables, and (2) it was the primary source of data used to produce a 2 m top-of-canopy-surface model as a precursor

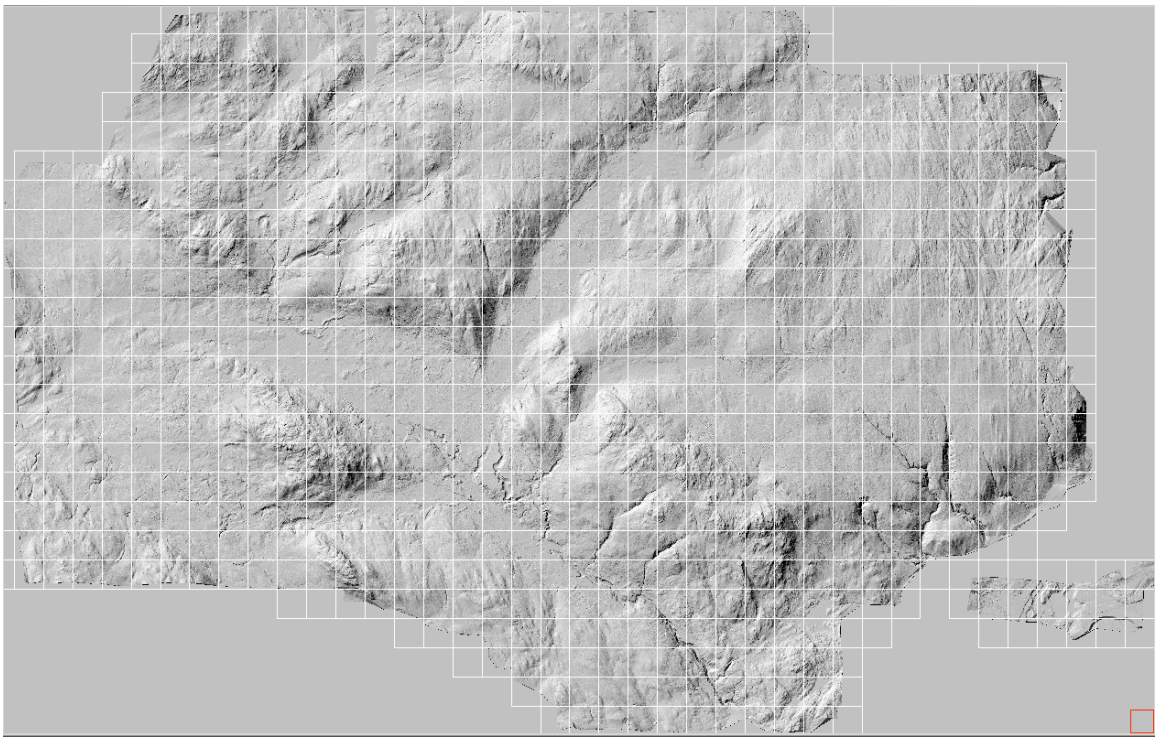


Figure 5. Hillshade representation of bare earth model with the 1 km^2 processing grid superimposed.

to the tree-top model. We constructed the 2 m top-of-canopy-surface model using the maximum LiDAR height found in each 2 m cell (Figure 7).

Last, we developed a tree-top model by moving a 3 × 3 pixel (36 m^2 in horizontal ground units) or 5 × 5 pixel (100 m^2 in horizontal ground units) 'local maximum (LM)' filter kernel (window) over the 2 m top-of-canopy model to isolate and locate the apex of each overstory crown in the image array. The LM algorithm was originally developed for high-spatial-resolution optical imagery, whereby the brightest pixel (maximum digital number) within the crown was assumed to represent its apex or highest point (Hay and Niemann, 1994; Niemann et al., 1999; Coops and Culvenor, 2000; Wulder et al., 2000). With LiDAR data, the LM rule is simplified somewhat by the assumption that the crown apex has been found when the centre cell of the filter kernel is coincident with the local maximum canopy height (i.e., all other surrounding grid cells have a lower surface canopy height). As expected, this technique is highly effective when the scale (grain and extent) of the LM filter optimally matches the scale of the image objects (individual crowns). However, in cases where there is a significant scale mismatch between these two spatial entities (i.e., when kernel size is markedly smaller or larger than crown size), errors of omission (missed trees) and commission (falsely identified trees) become more prevalent.

We relied on the tree-top model to support two critical processing steps. First, we used the location of each identified local maximum canopy height to sample continuous reflectance spectra (400–2500 nm) backscattered from each identified crown. These sampled spectra were subsequently classified by species, or species group, using a Spectral Angle Mapper (SAM) classification (details of SAM are given below). We developed this sampling approach based on two main assumptions: 1) individual species within and between taxonomic groups are most reliably identified using branch- and crown-scale reflectance spectra (Roberts et al., 2004), and 2) the least ambiguous spectral signatures tend to be associated with the sunlit portion of the crown (Leckie et al., 2005). Second, we arbitrarily adjusted the height associated with each tree-top location upward by 0.5 m to compensate for the potential negative bias associated with LiDAR tree heights (Rönnholm et al., 2004; Andersen et al., 2006).

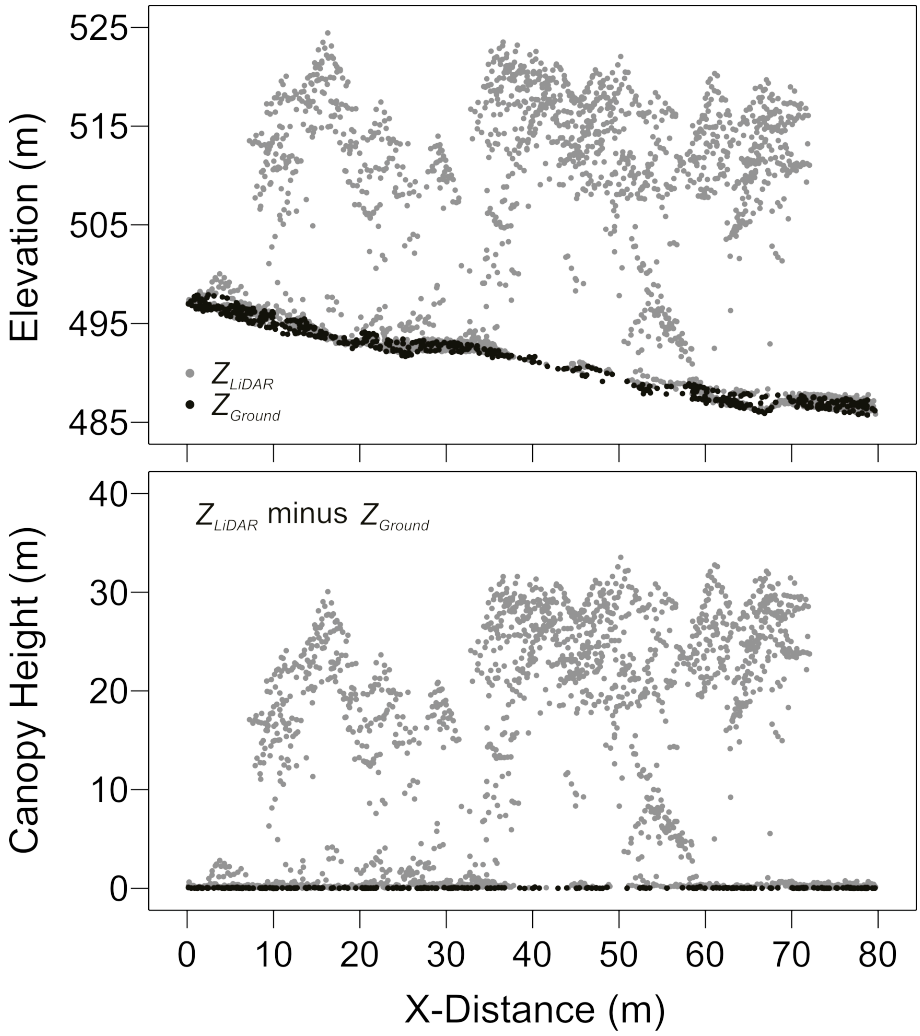


Figure 6. Transformation of LiDAR ground and vegetation elevations (metres above mean sea level or some other geodetic datum) to vertically projected heights above the ground surface. The conversion of LiDAR point elevations to heights is a two step process. First, a high-spatial-resolution (≤1 m) model of the ground-surface (DEM) is constructed from a triangulated irregular network (TIN) of LiDAR points classified as ground. Second, the original LiDAR point cloud (composed of ground and vegetation elevations) is superimposed on the DEM, and each LiDAR point in the dataset is then subtracted from an interpolated ground point located directly below on the DEM. We use these new, transformed LiDAR height data to compute canopy height and density metrics for predictive modelling of inventory attributes.

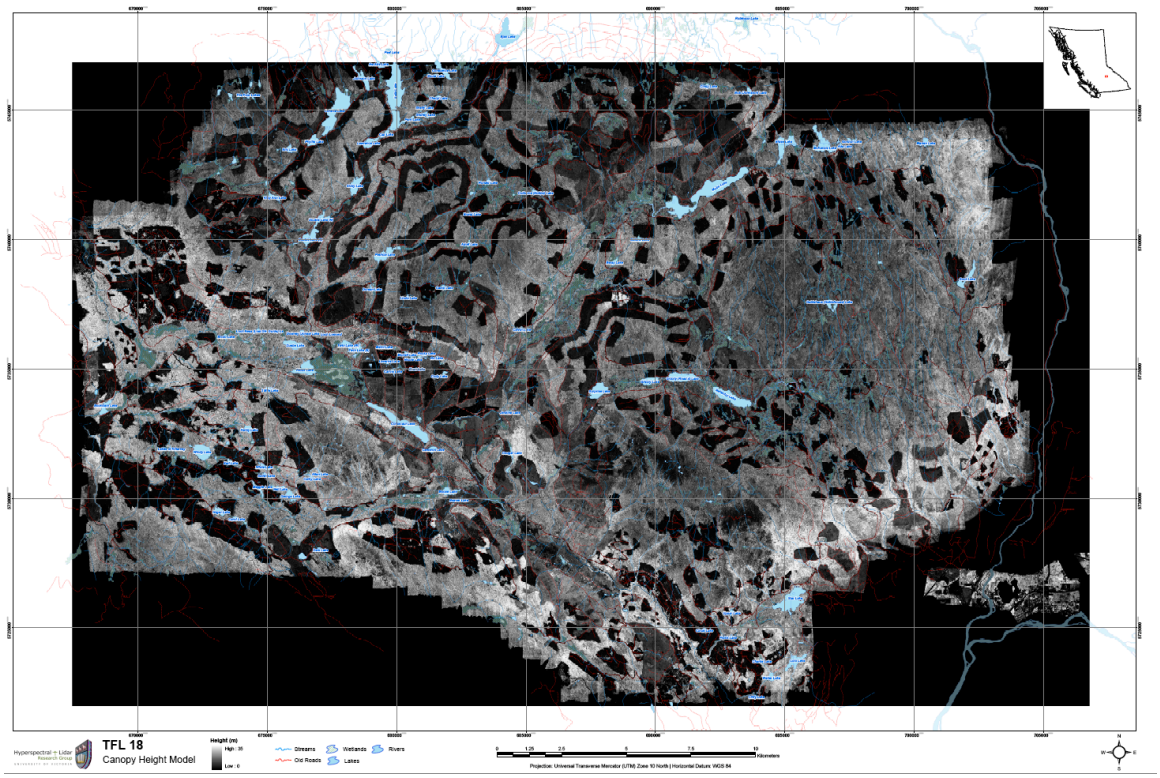


Figure 7. Terrain normalized canopy height model for TFL 18.

2.4.1.4. Grid-based LiDAR biometrics and synthetic variables

We constructed a regularized grid of 20 m (400 m^2 or 0.04 ha) cells, with an origin and extent (range rectangle) identical to the hyperspectral imagery and the four LiDAR-derived elevation/height models. The choice of grid-cell size was constrained not only by the spatial scale of these other data layers, but also by the need to match, as closely as possible, the footprint of our ground-reference inventory plots (i.e., 20 m diameter [314 m^2] circular plots). We applied the 20 m grid as a spatial template to extract a number of descriptive statistics (biometrics) from the normalized LiDAR point cloud composed of heights rather than elevations. We applied a fixed height threshold of 2 m to eliminate any of near-ground LiDAR points from the biometric calculations (Nilsson, 1996). Near-ground returns are associated with a much higher classification uncertainty and therefore considered to be a potential source of noise or error when modelling inventory attributes. We also found through experimentation that biometrics (particularly density-based metrics, see below) calculated with height thresholds of 1 m or less were highly sensitive to variations in LiDAR point density (Gobakken and Næsset, 2008).

We selected three separate classes of biometrics based on previous work by Magnussen and Boudewyn (1998), Means et al. (2000), Næsset (2002), and Frazer (2007). First, we estimated sample quantiles corresponding to the 0, 5, 10,…., 100th percentiles of the LiDAR canopy height (>2 m) distribution on a per grid cell basis following "*Definition 6*" (SAS and SPSS defaults) in Hyndman and Fan (1996):

$$Lh_q = (1 - g)x_i + gx_{i+1}$$

where Lh_q is a sample quantile of LiDAR canopy height corresponding to sample fraction or probability q ($0 \le q \le 1$); x_i is the sorted height or ith order statistic of a LiDAR canopy return with ascending rank order i (i.e., $x_1 \le x_2 \le \ldots x_n$); g is a weighting coefficient ($0 \le g \le 1$) given by $j = \lfloor qn + m \rfloor$ and $g = qn + m - j$, where $\lfloor u \rfloor$ denotes the largest integer not greater than u (i.e., floor function); n is the sample size, and m is a constant set to q. Estimates Lh_q at selected values of q (typically $0.5 \le q \le 0.9$) are often highly correlated with stand height and other inventory attributes (e.g., volume, biomass, etc.) dependent on height (Magnussen and Boudewyn, 1998; Lim and Treitz, 2004). Also, quantile plots

of Lh_q (x-axis) versus q (y-axis) reveal the overall pattern of vertical canopy structure (Figure 8), and therefore provide a useful signature to further quantify vertical canopy structure (Frazer, 2007).

Second, we implemented the method of L-moments developed by Hosking (1990) to quantify location, scale (dispersion), and shape (skewness and kurtosis) parameters of the LiDAR canopy height distribution within each cell. We adopted this method rather than the conventional method of moments (i.e., sample product moments given by any statistical package), because L-moments are markedly more robust (show less bias, variance, and ambiguity) in the presence of outliers, skewness, and small sample sizes (n < 100), and able to accurately identify and distinguish between a much broader range of distributional shapes (Hosking, 1992; Vogel and Fennessey, 1993; Guttman, 1994). L-moments are derived from linear combinations of order statistics, and can be estimated indirectly for any distribution using unbiased sample estimates of the probability weighted moments (Vogel and Fennessey, 1993):

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$b_1 = \sum_{i=1}^{n-1}\left[\frac{(n-i)}{n(n-1)}\right]x_i$$

$$b_2 = \sum_{i=1}^{n-2}\left[\frac{(n-i)(n-i-1)}{n(n-1)(n-2)}\right]x_i$$

$$b_3 = \sum_{i=1}^{n-3}\left[\frac{(n-i)(n-i-1)(n-i-2)}{n(n-1)(n-2)(n-3)}\right]x_i$$

where, b_0, b_1, b_2, and b_3 are unbiased sample estimates of the first four probability weighted moments (PWM); x_i is the sorted (ordered) height, or ith order statistic, of a LiDAR canopy return with descending rank order i (i.e., $x_1 \geq x_2 \geq,\ldots., x_n$), and n is the sample size. The first four L-moments, which are analogous to conventional central moments, can be estimated from PWMs using the following linear equations:

$$\lambda_1 = b_0$$

$$\lambda_2 = 2b_1 - b_0$$

$$\lambda_3 = 6b_2 - 6b_1 + b_0$$

$$\lambda_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

Finally, *L*-moment ratios, which are analogous to product moment ratios (i.e., coefficient of variation, skewness, and kurtosis) and independent of the units of measurement of x (m), can be estimated by (Hosking, 1990):

$$\tau = \lambda_2/\lambda_1, \qquad (0 < \tau < 1).$$

$$\tau_3 = \lambda_3/\lambda_2, \qquad (-1 < \tau_3 < 1).$$

$$\tau_4 = \lambda_4/\lambda_2, \qquad (0.25(5\,\tau_3^2 - 1) \leq \tau_4 < 1).$$

We therefore defined the mean (Lh_{Mean}), coefficient of variation (Lh_{COV}), skewness (Lh_{Skew}), and kurtosis (Lh_{Kurt}) of the LiDAR canopy height distribution as λ_1, τ, τ_3, and τ_4, respectively, in keeping with the definitions given by Hosking (1990).

Last, we computed canopy density or cover (CC_F) at 20 equal intervals of relative height F = 0, 0.05, 0.1,…,1 from the near bottom to the near top of the canopy following the methods of Gobakken and Næsset (2008):

$$CC_F = 100\%\left(\frac{n_f}{N}\right)$$

where n_f is the number of LiDAR points found above relative height F, and N is the total number of LiDAR points (ground and vegetation) within the cell. Relative height F ranged between 0 and 1, and was constrained between the absolute heights of 2 m (F = 0) and $Lh_{0.95}$ (F = 1) above the ground surface. Estimates of CC_F were restricted to within these two height limits, because we, and others (e.g., Gobakken and Næsset, 2008), have found that CC_F becomes increasingly sensitive to variations in LiDAR point density outside these limits. CC_F, especially when combined with Lh_q, has also been found to be a useful predictor of inventory variables that are dependent on stem density (e.g., basal area, volume, biomass, etc.). CC_F changes dramatically from a minimum value at the top of the canopy to a maximum at the bottom, and again provides a useful and distinctive signature of vertical canopy structure (Figure 8). CC_0 is often applied as an estimate of vertically projected canopy cover (Means et al., 2000; Hopkinson and Chasmer, 2008).

Previous research has shown that many of the LiDAR-derived height and density metrics can be highly intercorrelated, which is no surprise given that all of these variables

were derived from the same canopy height distribution (Næsset et al., 2005). To eliminate this redundancy, we constructed a cross-products (correlation) matrix using all of the original Lh_q, L-moment, and CC_F variables (p = 42), submitted it to a principal components analysis (PCA), and saved only the non-trivial components (axes) as new synthetic variables (Legendre and Legendre, 1998; Peres-Neto et al., 2005). We used these new uncorrelated (orthogonal), synthetic variables as more complex measures of vertical canopy structure by which to stratify the landscape into relatively homogeneous units (see below).

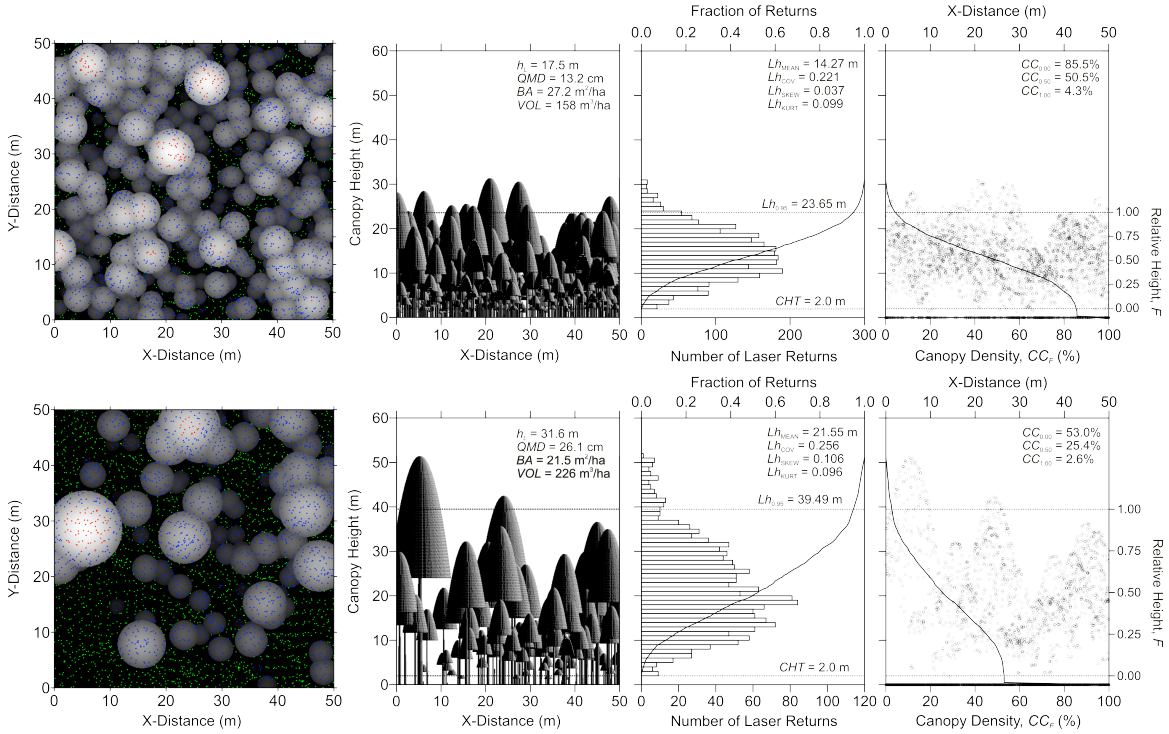


Figure 8. Graphical examples of LiDAR canopy height (Lh_q and L-moments: Lh_{MEAN}, Lh_{COV}, Lh_{SKEW}, and Lh_{KURT}) and density (CC_F) metrics extracted from two simulated forest canopies. The left most panel shows the view of the simulated canopy from above; the coloured points, which are coded by height (i.e., red ≥ $Lh_{0.95}$; 2 m < blue ≤ $Lh_{0.95}$; green ≤ 2 m), represent a sample of LiDAR points clipped randomly from the TFL 18 LiDAR dataset (n = 3192 per 0.25 ha or 1.28 points per m^2). All three panels to the right share the same vertical scale and show the vertical distribution of canopy surfaces and intersecting LiDAR points. The third and fourth panels (counting left to right) show the curves associated with Lh_q and CC_F datasets, respectively. These curves and their associated statistics provide unique signatures of the vertical canopy structure found within each discrete grid cell.

2.5. Hyperspectral processing

2.5.1. Orthorectification of hyperspectral imagery

The hyperspectral data collected by MAP is unique on a number of fronts. The first is that the spatial resolution of the imaged pixel is very fine (1 to 2 square metres). This yields an image database that that can be highly variable with respects to the signal to noise ratio (S/N), with reflectance intensity decreasing quickly into the canopy, or on sides of objects. A second area where the data are unique is in the simultaneous collection of both hyperspectral and LiDAR data. This allows us to exploit the positional accuracy and information of the LiDAR data to position the hyperspectral imagery, resulting in an image dataset that is positioned to within the accuracy of the DGPS of the INS (typically better than 1 metre) as well as othrocorrected to within a pixel of the LiDAR data.

As part of the preparation of the hyperspectral data for further analysis we *georectified* them to position the imagery accurately within geographic space, as well as to remove the effects of relief / vertical displacement. This coregistration is crucial in many of the subsequent elements of the processing. Our processing thread is based on a parametric *georectification* approach where the specifics relating to the absolute positioning and attitude of the platform is derived from the metadata supplied by the INS system. This allows us to position the imagery with respect to a flat surface without the need for ground control and tie points. Both the absolute location of the platform as well as the attitude are considered. The data are further *orthorectified* through the use of the LiDAR height data, which removes the effects of displacement caused by vertical deviations from a flat surface. With data having a very small spatial resolution they are prone to effects of displacement caused by relatively small vertical features, such as trees. To account for this displacement we needed to develop an elevation model that takes into account the effect of these features.

Traditional orthorectification employs a DEM that assumes a bare earth. We needed to develop a surface model that also included the features that were located on the bare earth. To this end we derived the TORS elevation model that provides us with the maximum height within a grid cell. The spatial resolution of the TORS model was set to the same as the hyperspectral imagery. The result of the orthorectification is presented in Figure 9a. In this sample we can examine the relationship between the orthorectified

hyperspectral and the LiDAR intensity data. The results of the orthorectification can be seen to be within a single pixel of the LiDAR image. A second example is presented in Figure 9b where the hyperspectral data has been draped onto corresponding LiDAR data demonstrating the accuracy of the positioning relative to sharp vertical boundaries.

2.5.2. Radiometric and Atmospheric Correction

The hyperspectral data were, in addition to the orthorectification, processed to extract ground reflectance. This allowed us to compare data collected over multiple days and in differing light conditions. Ground reflectance was modelled using a *radiative transfer model* (MODTRAN). This allowed us to account for variations in incoming radiation caused by geographic location, time of day and time of year. The resulting output, *apparent reflectance*, did not take into account the atmosphere between the instrument and the ground. The noise introduced by the atmosphere here was largely compensated for by applying an *empirical line calibration* (ELC) using ground reflectance spectra collected on the day of the data collection. This calibration adjusts the apparent reflection to the ground calibration ground spectra. The ELC compensates for differences between measured ground reflectance and modeled reflectance at the sensor at the time of measurement. It does not however compensate for changes in the opacity of the atmosphere throughout the day, or for spatial variations encountered.

2.5.3. Image analysis and Classification

Hyperspectral data were used in this project for the identification of forest species. The spatial format of the hyperspectral data, like other forms of passive optical imagery, is one that is reliant on gridded matrices. These data matrices, while positionally consistent, have no spatially explicit topological structures. In other words, they are essentially flat files with a minimum of explicit spatial referencing, limited to an origin and pixel spacing. The objective of this work was to integrate the hyperspectral with the LiDAR data so a new approach to addressing the data was envisioned. This new approach used the LiDAR data to guide the sampling of training/calibration of the hyperspectral data. The reasons for this were twofold. The first was that it allowed us to

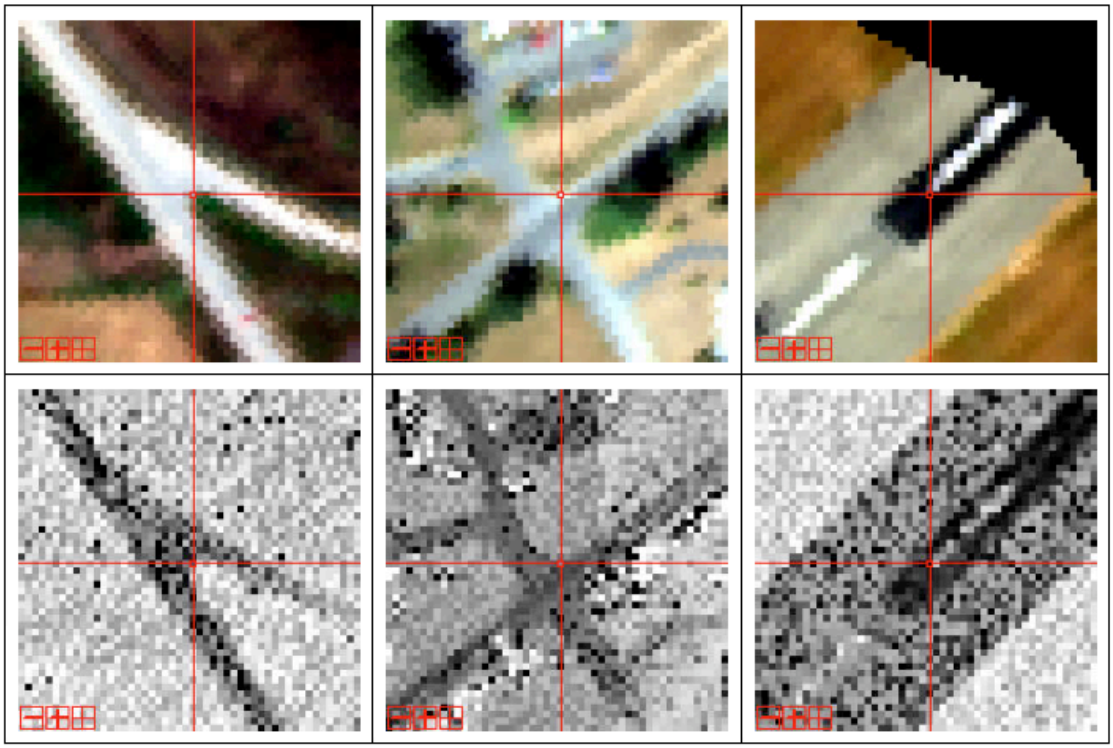


Figure 9a. Spatial accuracy of the orthorectification process employed in the processing of the hyperspectral data. The top images represent the hyperspectral data while the bottom is LiDAR intensity. The two sets are spatially linked so that the cross hair is placed on the same geographic coordinate.

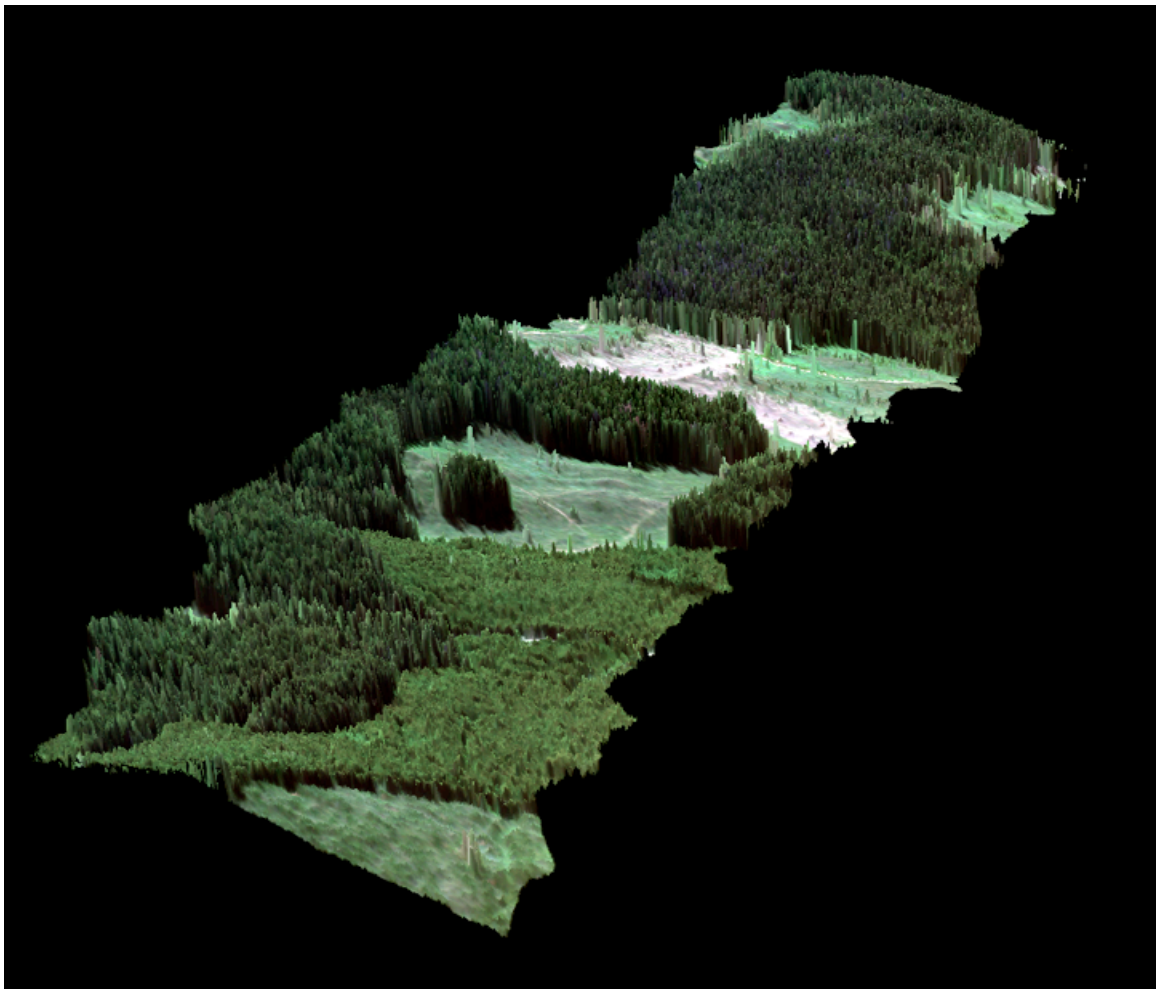


Figure 9b. A sample of a portion of a flightline with the hyperspectral (3 bands) draped onto the LiDAR canopy height model.

focus only on the portions of the image that were relevant to the analysis. Second, and more important, was that with very high spatial resolution images the influence of shadowing, and a general decrease in illumination with depth into the canopy, results in a decrease in the signal to noise, often very rapid, as we move away from the top of the canopy. This lowering of S/N often results in decreasing classification, or other analytical, accuracies. To accommodate for the signal to noise problem we adopted two separate strategies: 1) a guided sampling, and 2) a spectrum smoothing / polishing. The first strategy uses the LiDAR data to help filter out pixels that do not represent the areas with the highest illumination. This can be achieved in two ways. The first is to use the LiDAR canopy height surface to eliminate all of the areas below the dominant-codominant canopy. While this approach helps to remove unwanted pixels it does not resolve the first of the issues described above, which is that the image remains in a georeferenced flat file format.

To over come this we adapted a method based on the tree-top isolation methodology developed in the 1990's (Hay and Niemann, 1994, Niemann, Adams et al., 1999; Coops and Culvenor, 2000; Wulder, Niemann et al., 2000), based on a relatively simple *local maximum filter* (LM) approach, to define the locations of the trees. The LM approach iteratively scans a height matrix (in this case the LiDAR canopy height model) and searches for a local maximum within a 3x3 or a 5x5 submatrix . The LM rule is that the grid cell in the centre of the submatrix must be higher than all of the surrounding grid cells. While this approach proved to have some utility in 2-dimensional digital imagery, where the assumption that the brightest pixel within a tree crown is also the highest is commonly met, it is ideally suited for LiDAR data where this assumption is not necessary. That is, the highest LiDAR value represents the highest measured point on a tree crown. There are a number of drawbacks to using this technique. The first relates to the stand density, where overlapping crowns may be missed as separate entities. This method is more accurate in managed stands where the crowns are separate, and more regularly spaced. High success rates for extracting the locations of individuals have, however, also been achieved for old-growth coastal stands, although there are increased omission and commission errors due to large and broken crowns . (Our independent assessment suggests that when addressing coastal old-growth Douglas-fir stands

approximately 85% of the individuals were isolated using this approach). The second is that the highest point on the crown may not be represented by the LiDAR data. The LiDAR footprint (or the area of the laser beam that intersects with the reflecting surface), is typically 20 to 30 cm for the data flown in this mission, and may, or may not, intersect with the top of the tree. The result is an underestimation of the tree height, which has repeatedly been reported in the literature (Gaveau and Hill, 2003; Rönnholm et al., 2004; Andersen et al., 2006).

The individual tree-top locations were spatially registered with the hyperspectral data. A single spectrum corresponding to the top of each tree was sampled and retained for future analysis. Once the spectra were collected then the second portion of the noise reduction strategy was implemented. This involved the use of a *polishing,* or smoothing, mechanism to remove the white noise component imbedded in the signal, while retaining valid variations in the spectrum. A number of methods have been devised for noise reduction (see for example Tsai and Philpot, 1998). The nature of the noise is random, and any averaging (be it along the length of the spectrum, or using multiple spectra) will reduce its effects. One of the issues associated with applying a filter such as this is that there is a threshold between not removing enough of the noise and too much, so that the actual information content of the spectrum is affected. For this project we implemented a relatively simple smoothing algorithm based on a moving average. Through iterative analysis we found the optimum number of points to average along the spectrum to be 7. This allowed for the removal of the noise while not smoothing the spectrum. A sample of the before and after smoothing of a spectrum is presented in Figure 10.

Along with the spectrum we retained the geographic coordinates and LiDAR-based tree height. The advantage of using this approach is that we can apply any appropriate analytical technique, and then map the tree locations with desired attribute(s) back into geographic space where they can be integrated with other attributes. This allows us to address the attributes on an object basis and upscale to uniform polygons or continuous grid representing, for example, dominant species, volume/ha, or biomass.

The resultant spectra were analyzed with the same techniques available to those working within the image domain. The species classification for this project adopted the *Spectral Angle Mapper* approach (Kruse et al., 1993). This approach relates the

accumulated differences between spectra of known reference spectra from a library with unknowns. The reference spectrum that comes closest to the unknown shares its label with that unknown sample (see Figure 11). Other approaches to classify the spectra, based on multivariate statistics, and borrowing from the more traditional classification approaches in remote sensing, are also being explored, although they were not implemented for this project.

The SAM classifier relies on a supervised classification approach. It is imperative therefore that the spectra used to calibrate the SAM classifier (or training spectra) be representative of the species that are on the ground. For our work we built the calibration library incrementally, with initial input only from identified pure, single species, stands. As these stands were typically not abundant we needed to also locate others, from the existing forest inventory information, where a single species constituted 80 to 90 percent of the composition. In these cases the extracted spectra were sorted, based on the characteristics of the initial pure spectra selection, and separated into their representative species. This allowed us to build, iteratively, reference libraries using the most reliable samples and using them to help identify subsequent samples with a lower likelihood of being pure. In this project we developed a reference library of 105 spectra representing the main species found within the study area. We chose multiple samples of each species. As there were always minor variations in the shape of common species, we chose a range of samples to capture the majority of the variance. The remaining spectra sampled from the project dataset were then compared to the reference library spectra and assigned a specie label.

2.6 Grid-Based, Two-Stage Inventory Procedure

We adopted the grid-based, two-stage inventory procedure of Means et al. (2000) and Næsset (2002) to model structure-related forest inventory variables across the TFL. The first stage of the procedure relied on a network of ground-reference plots with field inventory measurements and spatially coincident LiDAR-derived canopy height and density metrics (biometrics). Multiple linear regression of a single Y (inventory) on multiple Xs (LiDAR biometrics) was then used to statistically relate one dataset to the

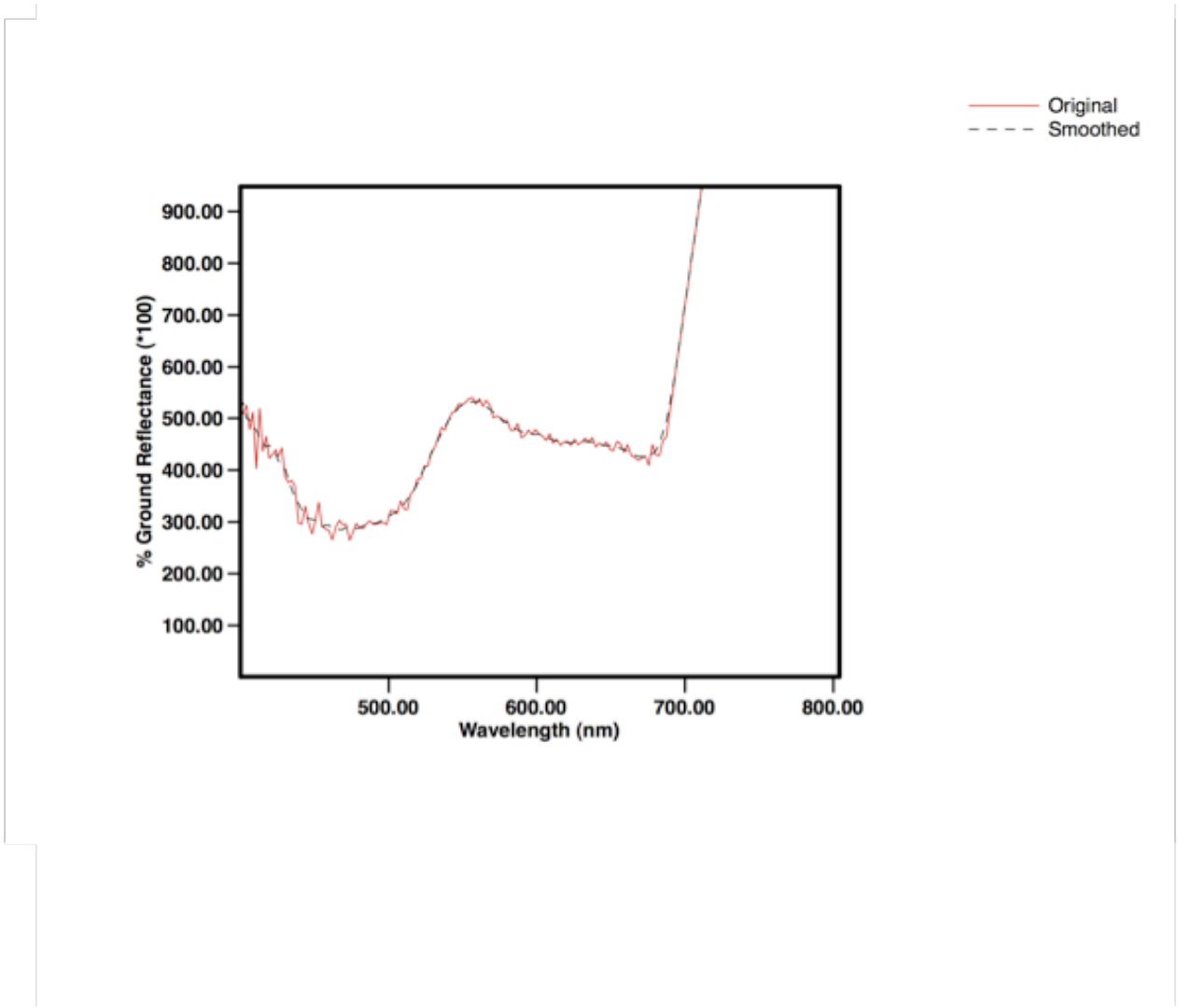
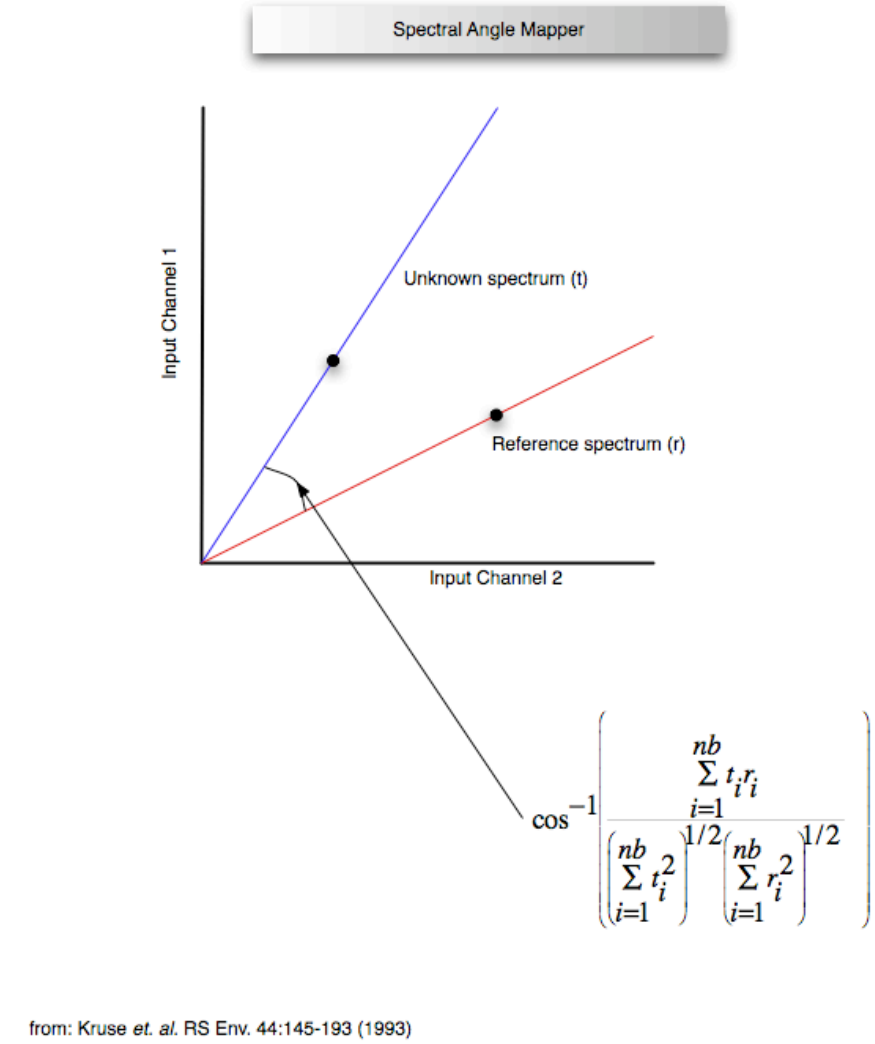


Figure 10. A sample of a portion of a spectrum showing a raw and smoothed data.



from: Kruse *et. al.* RS Env. 44:145-193 (1993)

Figure 11. Schematic of the Spectral Angle Mapper (SAM) algorithm.

other for the purpose of prediction (Næsset et al., 2005). In the second stage, inventory attributes were spatially extended across the survey area on a per cell basis using the fitted regression models that were developed in the first stage.

2.6.1. Selection of ground-reference sites

One critically important step in this two-stage procedure involves the selection of an unbiased sample of ground-reference sites for model building. We developed a novel approach for sample selection based on the assumption that model bias could be minimized by choosing field plots that were representative of the full range of canopy structures present within the TFL. The reasoning behind this assumption was based on three known facts: (1) correlations between canopy structure and stand structure are sufficiently strong that LiDAR-derived attributes of canopy structure (primarily height and cover) can be considered reliable proxy for measures of stand structure (Lefsky et al., 2005); (2) a single estimate of basal area or volume, etc. can be associated with an infinite variety of canopy height distributions (i.e., top-loaded, bottom-loaded, uniform, bimodal, multimodal, or Gaussian canopy height profiles) (Frazer, 2007); (3) canopy structure reflects the integrated influence of age, species, disturbance, site productivity, and climate, and is therefore an important and powerful unifying variable (Parker, 1995).

Our methodology for site selection was based on the following steps. First, we saved only the non-trivial synthetic principal components obtained from a correlation matrix of the original 42 LiDAR biometrics, and then standardized each of them by the mean absolute deviate (Legendre et al., 2002; Kaufmann and Rousseeuw, 2005). Second, we arbitrarily selected a random sample of 30,000 cells from the total 1,438,414 cells comprising the TFL and constructed an n (rows) × p (columns) rectangular matrix, where n is the sample size and p is the number of standardized synthetic variables. Third, we used hierarchical agglomerative clustering by Ward's minimum variance method (Legendre and Legendre, 1998), and three global stopping rules to identify a smaller, optimal number (k) of statistically unique and relatively homogeneous groups (clusters) of vertical canopy structures within the dataset (Caliński and Harabasz, 1974; Hartigan,

1985; Krzanowski and Lai, 1988). Fourth, we refined Ward's cluster membership using *k*-means partitioning (i.e., by minimizing the total Euclidean sum of squares), with Ward's group centres as the initial starting configuration for the reallocation of samples (Legendre and Legendre, 1998). Fifth, we used the final *k*-means solution to train a maximum likelihood classifier to assign class membership to all other unclassified cells in the dataset. Last, we obtained a proportionate stratified random sample of cells (i.e., potential field plots), and selected only those locations that were well within obvious stand boundaries, and not more than 300 m from the nearest access road. Due to the limited resources available for this project, we restricted our stratified random sample and modelling effort to a single 1:20,000 BCGS map sheet (92P.069).

2.6.2. Ground-reference sample

We established a total of 35 ground-reference plots in different forest stands across BCGS map sheet 92P.069 in early October 2008 (approximately two years after the initial airborne remote sensing survey). Three of these plots were located within adjacent map sheet 92P.070 (Figure 12). The majority of the plots (30 plots) occupied the BGC ICH zone, while the remainder were found in the SBS (4 plots) and ESSF (1 plot) zones. Plot elevations ranged between 1009 and 1516 m.a.s.l., although most (28 plots) were located at elevations below 1200 m.a.s.l. The centre position of each 10 m, fixed-radius (314 m^2 or 0.0314 ha) plot was surveyed using differentially corrected GPS (UTM Zone 10 N, NAD83). No formal accuracy assessment was completed on the post-processed GPS positions; however, we expect a horizontal error of less than 5 m based on dead-reckoning with the aid of high-resolution aerial photographs and LiDAR imagery. Our field data collection and subsequent data processing included the following (summarized in Table 2 a and b):

- We censused all live and dead trees greater than 2 m in height for total stem height, diameter at breast height (DBH), species, and crown class and condition.
- We computed plot-level estimates of top height (m), stem density (stems/ha), basal area (BA; m^2/ha), quadratic mean diameter (QMD; cm), gross volume (VOL; m^3/ha), total aboveground biomass (TAGB; Mg/ha, dry weight), species

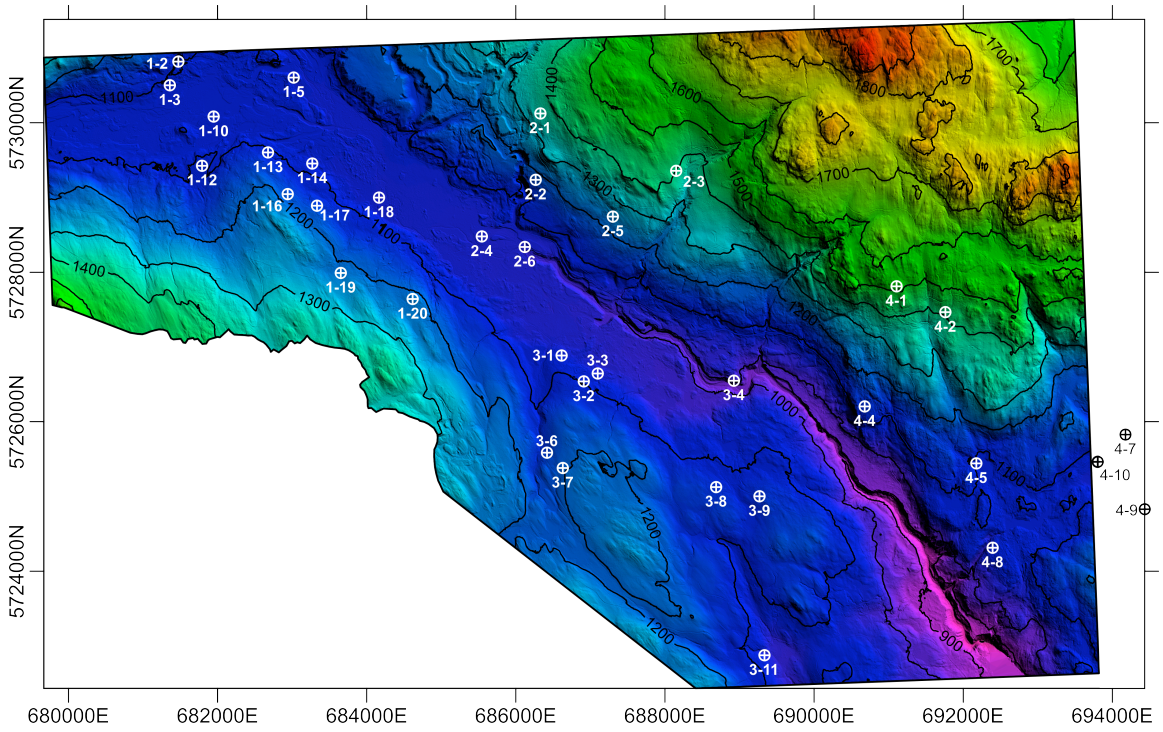


Figure 12. Locations of ground-reference plots within BCGS map sheet 92P.069 (see Figure 2 for expanded view). Plots were identified using a stratified random sample of points, and by selecting only those found within obvious stand boundaries and not more than 300 m from an access road. Sample strata were identified by a *k*-means partition (k = 8) of the first 5 principal components extracted from a correlation matrix of LiDAR biometrics. Plots 4-7, 4-9, and 4-10 were located in adjacent map sheet 92P.070.

proportion by BA (%), and leading species by BA based on the live tree component only.

- We counted all MPB red-attack and some gray-attack pine (those still retaining finer structural elements, such as defoliated shoots and twigs) as live, based on the assumption that these trees would have been alive (healthy or green-attack) in the year of the airborne remote sensing survey (August 2006).
- We implemented three different measures of top height (Sharma et al., 2002): Lorey's (basal-area-weighted) mean height (H_L); mean height of the tallest 100 trees per hectare (H_{100T};), and the mean height of the largest diameter trees per hectare (H_{100D}). H_{100T} and H_{100D} were both estimated using the average height of the largest 4 trees per plot. QMD is defined as the diameter of the tree of mean basal area, and like H_L gives greater weight to the larger canopy trees (Curtis and Marshall, 2000). Whole stem volume (m^3) was estimated for each tree using species- and region-specific taper equations implemented in BC Ministry of Forests and Range, Tree Volume Compiler (TREEVOL) software. Finally, we calculated TAGB (kg, dry weight) on a per stem basis using species-group-specific, DBH-based allometric equations (Jenkins et al., 2003).
- We used the GPS plot centres and 10 m fixed-radius boundary to sample the normalized LiDAR point cloud at each of the 35 plot locations. From these LiDAR points we extracted a total of 46 biometrics (see subsection 2.3.1.4.) from each of the 35 LiDAR canopy height distributions: Lh_q (n = 21), Lh_{Mean}, Lh_{COV}, Lh_{Skew}, Lh_{Kurt} (L-moments), and CC_F (n = 21) (Table 3a–b).

All biometrics were treated as potential candidate predictors (explanatory variables) in multiple linear regression analysis of field-measured attributes (response) against LiDAR biometrics (predictors).

Table 2a. Description of ground-reference plot locations and plot-level inventory statistics.

Plot ID	Plot Centre		Elevation (m.a.s.l.)	BEC Subzone	Struct. Class	Age (yr)	MPB Mortality (%)	Dead/ Damaged (%)	Live Stems (n)	H_L (m)	H_{DC} (m)	H_{100T} (m)	H_{100D} (m)	QMD (cm)	Mean DBH (cm)	SD (n/ha)	BA m^2/ha)	VOL (m^3/ha)	TAGB (Mg/ha)
	Easting (m)	Northing (m)																	
1-2	681473.5	5730818.6	1109.8	SBSmm	1	120	10.71	17.46	56	14.3	14.3	20.8	18.7	15.4	13.1	1783	33.4	195.0	139.1
1-3	681359.1	5730501.8	1080.2	SBSmm	7	120	1.85	10.34	54	20.7	22.3	26.5	24.7	18.2	14.2	1719	44.9	376.6	215.8
1-5	683020.0	5730604.8	1087.7	SBSmm	1	160	8.06	19.44	62	19.1	19.2	24.9	24.4	17.1	14.3	1974	45.5	362.8	203.5
1-10	681946.6	5730082.9	1072.1	SBSmm	7	–	0.00	3.85	100	20.1	24.1	28.3	28.3	14.7	11.7	3183	54.2	433.9	252.4
1-12	681791.4	5729424.2	1105.5	ICHmk2	3	–	0.00	4.48	65	7.7	7.3	9.7	8.5	9.9	9.3	2069	16.0	50.7	49.4
1-13	682675.9	5729602.9	1109.8	ICHmk2	5	218	0.00	32.14	43	33.4	35.9	40.2	40.2	25.4	18.4	1369	69.2	674.2	454.9
1-14	683269.5	5729454.3	1079.8	ICHmk2	1	197	0.00	25.58	67	20.9	25.1	27.8	27.6	16.5	13.9	2133	45.5	401.9	205.0
1-16	682940.1	5729045.2	1176.4	ICHmk2	3	21	4.17	20.69	25	9.6	9.5	11.3	11.3	11.4	10.5	796	8.1	30.9	32.8
1-17	683332.9	5728890.1	1152.9	ICHmk2	3	–	0.00	0.00	16	4.7	4.6	5.3	5.2	5.4	5.3	509	1.1	2.7	3.1
1-18	684164.8	5728998.6	1056.9	ICHmk2	1	205	0.00	8.70	46	23.4	29.5	29.5	29.0	16.8	12.3	1464	32.5	303.6	160.1
1-19	683651.8	5727989.9	1215.8	ICHmk2	2	24	0.00	10.00	50	10.0	10.3	13.4	13.0	14.1	12.8	1592	24.8	102.3	100.9
1-20	684615.7	5727641.7	1222.3	ICHmk2	2	25	0.00	3.85	78	11.8	12.8	15.0	15.0	12.4	11.3	2483	29.8	134.8	130.1
2-1	686329.1	5730124.0	1353.0	ICHmk2	2	19	0.00	22.22	51	10.6	10.9	14.2	11.8	13.0	11.2	1623	21.5	87.6	87.5
2-2	686266.1	5729236.6	1172.6	ICHmk2	2	38	3.28	15.38	61	13.8	16.5	18.0	17.4	10.9	9.8	1942	18.1	100.3	69.3
2-3	688150.6	5729357.1	1399.3	ICHmk2	4	109	36.36	36.36	21	27.9	26.6	31.4	31.1	29.2	27.7	668	44.7	538.0	228.6
2-4	685543.7	5728479.1	1050.9	ICHmk2	2	153	13.95	16.28	43	17.4	18.3	22.7	21.4	16.4	14.1	1369	28.9	213.3	128.6
2-5	687301.0	5728744.0	1238.1	ICHmk2	2	30	16.13	16.13	31	11.0	11.1	14.7	14.7	15.0	13.8	987	17.4	76.4	68.4
2-6	686119.6	5728338.2	1044.6	ICHmk2	3	33	6.78	11.02	118	9.7	10.3	15.5	15.1	9.1	8.3	3756	24.7	109.7	78.5
3-1	686613.7	5726885.6	1066.9	ICHmk2	2	44	18.46	18.46	66	14.7	15.5	18.2	15.9	14.2	12.8	2101	33.2	198.2	140.3
3-2	686911.1	5726539.3	1094.7	ICHmk2	2	34	0.00	6.25	46	12.1	12.4	17.5	17.0	13.3	12.0	1464	20.3	98.1	85.0
3-3	687098.4	5726645.1	1053.3	ICHmk2	2	28	4.17	8.00	49	14.4	15.4	18.2	15.7	13.9	12.6	1560	23.6	135.2	103.7
3-4	688924.0	5726548.6	1009.6	ICHmk2	7	32	33.33	37.86	103	14.1	18.1	21.1	21.1	10.3	8.4	3279	27.1	171.5	94.8
3-6	686417.7	5725585.5	1150.3	ICHmk2	2	130	0.00	8.89	43	21.6	22.7	27.8	25.3	22.5	18.9	1369	54.6	461.6	258.5
3-7	686629.5	5725380.2	1192.8	ICHmk2	8	–	0.00	7.84	49	25.5	27.8	33.4	24.5	27.2	23.8	1560	90.6	851.5	510.9
3-8	688686.2	5725124.8	1127.9	ICHmk2	8	–	0.00	1.72	57	25.2	31.0	34.9	32.5	22.0	19.5	1814	69.2	695.4	366.8
3-9	689267.7	5725001.9	1134.7	ICHmk2	2	–	56.52	56.52	68	11.1	11.2	14.0	11.4	11.5	10.7	2165	22.4	107.0	71.1
3-11	689334.8	5722874.8	1057.9	ICHmk2	4	84	7.14	17.86	23	36.1	35.4	40.1	40.1	34.0	23.8	732	66.5	819.6	507.8
4-1	691106.3	5727811.3	1516.8	ESSFwc2	8	101	51.61	64.52	18	25.9	26.7	29.4	29.0	26.9	23.6	573	32.6	346.4	162.5
4-2	691761.0	5727466.2	1410.7	ICHmk2	1	102	0.00	46.27	40	22.7	25.1	27.8	27.0	19.0	15.5	1273	36.0	344.0	173.8
4-4	690678.9	5726204.2	1130.3	ICHmk2	3	14	0.00	0.00	69	3.9	3.7	4.9	4.8	4.2	3.9	2196	3.0	4.9	6.3
4-5	692176.4	5725442.1	1090.9	ICHmk2	4	91	16.67	33.33	27	34.3	34.2	39.3	39.1	31.2	26.1	859	65.9	883.7	410.9
4-7	694176.6	5725827.3	1049.6	ICHmw3	–	90	0.00	7.41	52	24.6	24.8	33.2	30.3	20.2	15.6	1655	53.0	491.5	231.8
4-8	692394.8	5724313.3	1074.3	ICHmk2	3	–	0.00	0.00	38	4.5	4.4	5.2	5.1	5.7	5.6	1210	3.1	5.6	6.9
4-9	694434.7	5724829.0	1196.1	ICHmw3	4	–	53.13	53.13	29	28.8	28.8	34.6	33.6	26.2	24.4	923	49.7	636.9	260.8
4-10	693804.5	5725463.0	1070.4	ICHmw3	2	–	0.00	16.22	63	17.5	17.4	21.8	21.5	15.5	12.8	2005	37.8	269.3	179.0

Note: UTM coordinates are Zone 10 N, NAD83; BEC – Biogeoclimatic Ecosystem Classification; Structure Class – *k*-means cluster; Age – stand age; MPB – mountain pine beetle; HL – Lorey's height; HDC – average height of dominant/codominant crown classes; H100T – average height of 100 tallest trees per hectare; H100D – average height of the largest 100 diameter trees per hectare; QMD – quadratic mean diameter; SD – stem density; BA – basal area; VOL – total volume; TAGB – total aboveground biomass (dry weight).

Table 2b. Species composition by basal area proportion for 35 ground-reference plots

Plot ID	BEC Subzone	Elevation (m.a.s.l.)	Struct. Class	Age (yr)	Total BA (m^2/ha)	Species Proportion by Percent Basal Area													First Leading Species	Second Leading Species	Third Leading Species
						Alnus (%)	*Populus* (%)	*Betula* (%)	*Acer* (%)	*Salix* (%)	Cw (%)	Fd (%)	Bl (%)	Hw (%)	Pl (%)	Se (%)	Sw (%)	Sx (%)			
1-2	SBSmm	1109.8	1	120	33.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	29.4	0.0	37.7	32.9	0.0	0.0	Pine	Spruce	Balsam
1-3	SBSmm	1080.2	7	120	44.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.3	0.0	7.0	80.7	0.0	0.0	Spruce	Balsam	Pine
1-5	SBSmm	1087.7	1	160	45.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	23.3	0.0	31.9	44.8	0.0	0.0	Spruce	Pine	Balsam
1-10	SBSmm	1072.1	7	–	54.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	54.7	0.0	0.0	45.3	0.0	0.0	Balsam	Spruce	
1-12	ICHmk2	1105.5	3	–	16.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	31.1	0.0	65.9	0.0	2.5	0.0	Pine	Balsam	Spruce
1-13	ICHmk2	1109.8	5	218	69.2	0.0	0.0	0.0	0.0	0.0	4.7	41.3	4.1	3.6	0.0	0.0	46.3	0.0	Spruce	Fir	Cedar
1-14	ICHmk2	1079.8	1	197	45.5	0.0	0.0	0.0	0.0	0.0	0.1	0.0	42.7	0.0	0.0	0.0	57.2	0.0	Spruce	Balsam	
1-16	ICHmk2	1176.4	3	21	8.1	0.1	0.0	0.0	0.0	0.0	0.6	53.7	14.4	0.0	12.1	19.1	0.0	0.0	Fir	Spruce	Balsam
1-17	ICHmk2	1152.9	3	–	1.1	0.0	0.0	0.0	0.0	0.0	35.7	49.0	0.0	0.0	15.3	0.0	0.0	0.0	Fir	Cedar	Pine
1-18	ICHmk2	1056.9	1	205	32.5	0.5	0.0	0.0	0.0	0.1	7.0	0.0	10.1	0.0	0.0	82.3	0.0	0.0	Spruce	Balsam	Cedar
1-19	ICHmk2	1215.8	2	24	24.8	0.0	0.0	0.0	0.0	0.0	16.1	2.7	30.3	0.0	0.0	50.9	0.0	0.0	Spruce	Balsam	Cedar
1-20	ICHmk2	1222.3	2	25	29.8	0.0	1.8	0.0	0.0	0.7	1.7	70.6	0.8	0.0	2.9	21.5	0.0	0.0	Fir	Spruce	Pine
2-1	ICHmk2	1353.0	2	19	21.5	2.8	7.1	0.0	0.0	20.5	0.0	0.0	29.9	0.0	0.0	6.8	0.0	32.9	Spruce	Balsam	Willow
2-2	ICHmk2	1172.6	2	38	18.1	0.0	54.7	0.0	1.7	19.1	3.0	6.5	0.0	0.0	13.5	0.0	1.5	0.0	*Populus spp.*	*Salix spp.*	Pine
2-3	ICHmk2	1399.3	4	109	44.7	0.0	0.0	0.0	0.0	0.0	0.0	11.4	0.4	0.0	30.8	57.4	0.0	0.0	Spruce	Pine	Fir
2-4	ICHmk2	1050.9	2	153	28.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	70.4	0.0	10.3	19.3	0.0	0.0	Balsam	Spruce	Pine
2-5	ICHmk2	1238.1	2	30	17.4	0.0	15.8	0.0	0.0	0.6	0.0	0.0	6.3	0.0	41.4	35.9	0.0	0.0	Pine	Spruce	*Populus spp.*
2-6	ICHmk2	1044.6	3	33	24.7	0.0	0.0	0.0	0.0	0.2	0.0	0.2	64.6	0.0	28.4	6.6	0.0	0.0	Balsam	Pine	Spruce
3-1	ICHmk2	1066.9	2	44	33.2	0.0	1.8	0.0	0.0	4.1	0.0	44.9	0.0	0.0	42.0	0.0	7.1	0.0	Fir	Pine	Spruce
3-2	ICHmk2	1094.7	2	34	20.3	0.0	11.7	0.0	0.0	0.4	0.0	33.4	4.4	0.0	0.5	49.5	0.0	0.0	Spruce	Fir	*Populus spp.*
3-3	ICHmk2	1053.3	2	28	23.6	0.0	2.2	5.4	0.0	7.8	0.0	49.3	0.8	0.0	14.6	19.9	0.0	0.0	Fir	Spruce	Pine
3-4	ICHmk2	1009.6	7	32	27.1	0.0	0.0	0.0	0.0	0.0	0.0	16.9	0.0	0.0	81.4	1.7	0.0	0.0	Pine	Fir	
3-6	ICHmk2	1150.3	2	130	54.6	0.0	0.0	0.0	0.0	0.0	34.4	0.0	57.0	0.0	0.0	8.6	0.0	0.0	Balsam	Cedar	Spruce
3-7	ICHmk2	1192.8	8	–	90.6	0.0	0.0	0.0	0.0	0.0	33.0	53.9	0.9	0.0	0.0	12.2	0.0	0.0	Fir	Cedar	Spruce
3-8	ICHmk2	1127.9	8	–	69.2	0.0	13.4	0.0	0.0	0.0	28.1	40.7	0.0	0.0	0.0	17.8	0.0	0.0	Fir	Cedar	Spruce
3-9	ICHmk2	1134.7	2	–	22.4	0.7	3.7	0.0	0.0	3.9	0.0	0.3	2.1	0.0	88.2	1.1	0.0	0.0	Pine		
3-11	ICHmk2	1057.9	4	84	66.5	0.4	0.0	0.0	0.0	0.0	0.0	65.0	8.3	0.0	0.0	26.2	0.0	0.0	Fir	Spruce	Balsam
4-1	ESSFwc2	1516.8	8	101	32.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.3	0.0	41.1	0.0	49.6	0.0	Spruce	Pine	Balsam
4-2	ICHmk2	1410.7	1	102	36.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	47.6	0.0	0.0	0.0	52.4	0.0	Spruce	Balsam	
4-4	ICHmk2	1130.3	3	14	3.0	0.3	0.0	0.0	0.0	0.7	0.0	0.8	0.0	0.0	82.2	16.0	0.0	0.0	Pine	Spruce	
4-5	ICHmk2	1090.9	4	91	65.9	0.0	0.0	0.0	0.0	0.0	0.5	39.8	1.7	0.0	21.3	0.0	36.7	0.0	Fir	Spruce	Pine
4-7	ICHmw3	1049.6	–	90	53.0	0.0	0.0	0.0	0.0	0.0	74.8	0.0	6.4	0.0	0.0	18.8	0.0	0.0	Cedar	Spruce	Balsam
4-8	ICHmk2	1074.3	3	–	3.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	Pine		
P4-9	ICHmw3	1196.1	4	–	49.7	0.0	0.0	0.0	0.0	0.0	1.8	44.8	1.0	0.0	52.4	0.0	0.0	0.0	Pine	Fir	
P4-10	ICHmw3	1070.4	2	–	37.8	0.0	0.0	0.0	0.0	0.0	0.0	13.5	35.6	30.8	0.0	20.1	0.0	0.0	Balsam	Hemlock	Spruce

Note: BEC – Biogeoclimatic Classification System; Structure Class – *k*-means cluster; Age – stand age; Total BA – total basal area per hectare; Cw – western redcedar; Fd – Douglas-fir; Bl – subalpine fir (balsam); Hw – western hemlock; Pl – lodgepole pine; Se – Engelmann spruce; Sw – white spruce; Sx – hybrid spruce; leading species determined by first, second, and third largest basal area proportions.

Table 3a. LiDAR canopy height statistics (i.e., quantiles and *L*-moments) for each of the 35 ground-reference plots.

Plot ID	Sample Quantiles of LiDAR Canopy Height																					*L*-Moments			
	Lh_{0} (m)	$Lh_{0.05}$ (m)	$Lh_{0.1}$ (m)	$Lh_{0.15}$ (m)	$Lh_{0.2}$ (m)	$Lh_{0.25}$ (m)	$Lh_{0.3}$ (m)	$Lh_{0.35}$ (m)	$Lh_{0.4}$ (m)	$Lh_{0.45}$ (m)	$Lh_{0.5}$ (m)	$Lh_{0.55}$ (m)	$Lh_{0.6}$ (m)	$Lh_{0.65}$ (m)	$Lh_{0.7}$ (m)	$Lh_{0.75}$ (m)	$Lh_{0.8}$ (m)	$Lh_{0.85}$ (m)	$Lh_{0.9}$ (m)	$Lh_{0.95}$ (m)	$Lh_{1.0}$ (m)	Lh_{Mean} (m)	Lh_{COV}	Lh_{Skew}	Lh_{Kurt}
1-2	2.17	2.60	3.27	3.82	4.14	5.02	6.61	8.74	10.07	10.65	11.93	13.32	14.64	14.96	15.51	16.28	17.55	18.47	19.67	24.02	27.66	11.78	0.324	0.066	0.025
1-3	2.05	3.43	4.03	4.83	5.38	6.67	8.70	11.55	13.55	16.38	18.13	19.11	19.97	21.85	22.55	23.64	25.66	26.98	27.83	29.93	31.81	16.46	0.315	-0.023	-0.033
1-5	2.04	3.52	4.87	6.05	7.35	7.80	8.38	8.87	10.41	11.53	12.67	13.69	14.93	16.48	17.42	18.49	20.21	22.07	24.33	27.26	30.63	13.80	0.300	0.103	0.045
1-10	2.35	3.49	4.19	5.31	6.55	6.99	7.80	8.88	10.27	11.22	12.65	14.48	15.97	17.94	20.37	22.22	24.82	26.21	28.93	31.30	35.09	14.70	0.343	0.153	0.015
1-12	2.02	2.21	2.44	2.61	2.82	2.89	2.94	3.09	3.11	3.25	3.37	3.42	3.53	3.62	3.79	3.94	4.30	4.65	5.18	6.00	13.55	3.70	0.186	0.364	0.323
1-13	2.61	5.22	6.44	7.52	9.26	11.52	13.70	15.59	16.90	18.71	21.40	22.59	24.06	25.32	27.09	28.08	29.13	30.50	32.57	35.83	40.59	20.19	0.275	-0.019	0.024
1-14	2.02	3.52	4.26	5.15	6.32	7.41	8.18	10.10	11.12	12.26	13.41	14.11	16.61	17.60	18.26	19.29	20.44	21.08	22.97	27.14	33.54	13.62	0.304	0.082	0.040
1-16	2.02	2.13	2.27	2.37	2.51	2.67	2.80	2.96	3.10	3.24	3.50	3.74	4.00	4.15	4.52	4.86	5.32	5.85	6.50	7.64	9.74	4.00	0.233	0.274	0.115
1-17	2.02	2.02	2.04	2.11	2.18	2.22	2.22	2.30	2.33	2.37	2.48	2.69	2.96	3.19	3.40	4.38	5.19	5.54	5.82	7.40	8.40	3.40	0.268	0.468	0.165
1-18	2.01	2.26	2.92	3.47	4.98	5.66	6.16	6.96	7.91	9.11	10.03	11.16	12.88	13.99	15.67	16.97	18.86	20.09	23.39	26.69	33.16	11.70	0.368	0.182	0.052
1-19	2.07	2.73	3.07	3.33	3.59	3.93	4.16	4.34	4.49	4.68	4.94	5.04	5.29	5.57	5.80	6.35	6.76	7.32	7.83	9.26	13.81	5.30	0.204	0.191	0.149
1-20	2.02	3.46	3.92	4.33	4.77	5.24	5.52	5.81	6.13	6.43	6.65	6.95	7.24	7.54	7.95	8.27	8.70	9.13	9.72	10.45	13.10	6.75	0.181	0.041	0.091
2-1	2.01	2.36	2.82	3.16	3.62	3.94	4.36	4.59	4.92	5.33	5.62	5.99	6.28	6.65	6.97	7.29	7.68	8.14	8.85	9.86	14.69	5.79	0.227	0.095	0.092
2-2	2.06	2.79	3.79	5.48	6.58	7.16	7.42	7.84	8.04	8.25	8.53	8.84	9.99	10.80	11.43	12.32	13.11	14.31	14.85	15.54	20.58	9.37	0.238	0.059	0.106
2-3	2.24	7.37	9.26	14.06	15.83	16.96	18.00	19.06	20.08	20.88	21.48	22.23	22.79	23.36	24.18	24.67	25.73	26.45	27.65	29.18	31.03	20.32	0.172	-0.172	0.136
2-4	2.10	2.75	3.21	3.59	4.16	4.41	4.66	4.96	5.45	6.11	6.96	8.44	9.41	10.06	11.06	12.11	12.81	13.61	15.03	17.21	20.18	8.31	0.311	0.167	0.018
2-5	2.11	2.78	3.29	3.63	4.27	4.90	5.45	5.85	6.47	6.86	7.18	7.49	8.19	8.79	9.33	10.04	10.99	11.50	12.48	13.76	15.50	7.56	0.254	0.087	0.051
2-6	2.07	2.42	2.65	3.17	3.46	3.81	4.03	4.22	4.57	4.92	5.30	5.56	5.78	6.05	6.73	7.39	7.76	8.60	9.79	11.55	13.39	5.78	0.252	0.203	0.116
3-1	2.17	5.33	6.73	7.53	8.37	9.29	9.72	10.01	10.20	10.41	10.72	11.00	11.20	11.47	11.81	12.10	12.41	13.02	13.41	14.34	15.95	10.37	0.140	-0.127	0.171
3-2	2.06	2.97	3.38	4.13	4.67	5.29	5.68	6.19	6.79	7.48	7.97	8.34	8.62	8.94	9.74	10.24	10.71	11.21	11.74	12.60	13.69	7.75	0.227	-0.003	0.042
3-3	2.14	5.85	6.58	7.21	7.64	8.09	8.29	8.53	8.75	9.01	9.32	9.70	10.24	10.52	10.83	11.10	11.50	12.12	12.54	13.45	17.06	9.50	0.142	0.007	0.132
3-4	2.05	2.33	2.67	3.14	3.45	4.01	4.40	5.21	5.63	7.48	8.50	10.17	11.53	12.30	12.84	13.60	14.52	15.31	16.08	18.67	20.91	9.23	0.332	0.095	-0.024
3-6	2.20	3.21	4.23	5.63	7.03	8.30	9.33	10.28	10.76	11.39	12.16	12.78	13.44	14.08	15.02	15.98	16.94	18.17	19.41	21.41	26.98	12.18	0.261	0.028	0.100
3-7	2.47	5.03	6.78	8.42	10.61	13.43	15.79	17.74	19.12	20.58	21.96	23.19	23.64	24.46	25.25	25.85	26.49	27.52	28.15	28.98	32.08	19.58	0.226	-0.178	0.032
3-8	2.67	4.65	5.45	6.03	6.82	7.99	9.16	10.70	12.70	15.51	20.70	24.12	25.52	26.56	27.47	28.35	29.35	30.22	31.10	31.87	34.05	18.75	0.310	-0.040	-0.106
3-9	2.01	2.49	2.98	3.38	3.66	4.13	4.52	5.07	5.35	5.67	5.91	6.16	6.44	6.63	6.85	7.14	7.50	7.80	8.30	8.71	9.99	5.74	0.192	-0.031	0.058
3-11	2.19	3.88	12.39	18.86	20.82	23.07	23.85	25.33	26.73	27.73	28.37	29.08	30.16	31.17	31.94	32.44	33.82	36.03	36.76	39.32	41.10	26.28	0.195	-0.203	0.167
4-1	2.03	2.60	3.31	4.40	5.69	7.09	9.29	11.22	13.61	15.50	16.66	17.77	18.60	19.79	21.06	22.04	22.70	23.66	24.45	25.94	32.11	15.07	0.303	-0.059	-0.015
4-2	2.07	3.12	4.64	8.30	9.55	10.59	12.06	12.84	14.09	15.16	15.96	16.85	17.59	18.36	19.41	20.03	20.69	21.71	23.21	24.99	29.99	15.17	0.241	-0.071	0.082
4-4	2.06	2.06	2.07	2.08	2.10	2.13	2.17	2.21	2.24	2.28	2.30	2.33	2.40	2.47	2.53	2.53	2.53	2.61	2.73	2.77	2.81	2.35	0.065	0.179	0.046
4-5	2.04	5.57	11.61	18.54	21.57	22.72	24.08	25.79	26.75	27.49	28.48	29.35	30.16	30.92	31.51	32.42	33.04	34.02	34.95	36.15	38.00	26.05	0.178	-0.258	0.154
4-7	2.09	3.04	4.12	5.17	5.90	6.61	6.99	7.49	8.24	9.06	9.71	10.43	11.25	12.13	12.65	14.28	15.39	17.39	19.64	21.74	26.15	10.77	0.293	0.166	0.093
4-8	2.02	2.03	2.11	2.24	2.28	2.31	2.31	2.37	2.40	2.60	2.68	2.70	2.83	2.85	2.90	2.94	2.96	2.97	3.07	3.17	3.18	2.62	0.078	-0.038	-0.025
4-9	2.12	7.14	10.39	12.47	13.05	14.53	15.44	16.49	17.46	18.25	19.35	20.06	20.52	21.66	22.40	22.96	23.85	24.78	26.12	28.19	30.67	18.46	0.187	-0.074	0.108
4-10	2.27	5.28	6.45	7.34	8.57	9.11	9.65	10.10	10.55	11.02	11.78	12.37	12.92	13.40	14.02	14.83	15.27	16.25	17.22	18.31	23.61	11.79	0.196	0.023	0.112

Note: LiDAR canopy height statistics were estimated using all LiDAR points with heights >2 m; all *L*-moments except the mean are unitless.

Table 3b. LiDAR canopy density statistics (CC_F) for each of the 35 ground-reference plots.

Plot ID	CC_0 (%)	$CC_{0.05}$ (%)	$CC_{0.1}$ (%)	$CC_{0.15}$ (%)	$CC_{0.2}$ (%)	$CC_{0.25}$ (%)	$CC_{0.3}$ (%)	$CC_{0.35}$ (%)	$CC_{0.4}$ (%)	$CC_{0.45}$ (%)	$CC_{0.5}$ (%)	$CC_{0.55}$ (%)	$CC_{0.6}$ (%)	$CC_{0.65}$ (%)	$CC_{0.7}$ (%)	$CC_{0.75}$ (%)	$CC_{0.8}$ (%)	$CC_{0.85}$ (%)	$CC_{0.9}$ (%)	$CC_{0.95}$ (%)	$CC_{1.0}$ (%)	Total # Pts	# Pts >2 m
1-2	45.06	41.26	35.39	32.76	31.81	30.20	28.96	27.60	23.61	22.15	20.44	18.74	13.96	11.14	8.81	6.39	4.46	3.66	3.00	2.85	2.06	222	100
1-3	54.44	52.17	46.67	42.54	40.01	38.35	37.30	35.82	34.23	32.41	30.84	28.07	25.43	22.08	19.97	16.46	13.62	11.01	8.02	5.07	2.55	314	171
1-5	70.80	67.98	64.80	60.99	57.78	50.63	44.62	41.66	38.22	33.94	29.50	27.01	21.92	18.08	15.71	13.70	10.80	8.86	6.21	5.12	3.24	185	131
1-10	74.41	71.38	66.46	60.76	51.67	47.66	42.28	38.14	34.62	31.68	28.13	25.29	24.03	20.26	18.09	15.97	12.61	10.31	8.56	5.40	3.57	250	186
1-12	28.99	27.78	26.30	24.65	23.23	18.67	15.97	13.07	9.97	8.21	6.79	5.81	5.23	4.17	3.44	3.06	2.57	1.94	1.62	1.45	1.17	257	75
1-13	66.43	64.75	62.92	58.74	55.47	52.22	49.88	46.88	44.05	39.20	36.31	34.66	31.35	27.42	22.47	19.38	14.01	9.74	6.98	4.81	3.21	308	204
1-14	60.97	59.70	54.87	50.59	47.84	42.85	40.09	37.18	34.88	30.81	26.65	24.62	22.69	19.19	15.44	11.71	9.14	6.55	4.87	3.54	2.74	223	136
1-16	44.87	39.95	34.97	30.10	25.96	23.44	20.28	17.40	14.66	12.87	11.26	9.99	8.58	7.45	5.88	4.97	4.17	3.63	3.26	2.61	2.08	436	195
1-17	4.44	2.96	2.22	1.98	1.73	1.23	1.23	1.23	1.23	0.99	0.99	0.99	0.99	0.74	0.49	0.25	0.25	0.25	0.25	0.25	0.25	405	18
1-18	45.24	39.49	35.82	33.00	29.38	26.65	24.01	20.54	19.41	17.13	14.75	13.64	11.34	11.00	8.58	6.31	5.63	4.50	4.45	2.71	2.03	304	137
1-19	64.33	63.28	61.23	57.86	53.79	49.20	44.87	38.05	33.18	26.49	21.81	18.24	16.16	13.59	11.43	8.59	6.59	5.60	4.94	4.02	3.12	699	450
1-20	80.90	80.20	79.13	77.97	75.47	70.75	67.46	63.40	59.09	53.04	47.33	40.66	35.06	29.86	25.02	20.02	16.21	11.95	9.18	6.65	3.98	810	656
2-1	85.04	80.61	76.92	71.87	68.51	63.47	59.07	53.93	48.58	43.62	39.51	33.15	29.29	23.19	18.88	14.46	11.69	9.81	7.03	5.41	4.14	337	287
2-2	75.63	72.40	69.50	67.11	66.46	64.49	61.90	59.19	52.70	43.73	34.40	31.57	29.40	26.42	22.37	19.49	16.16	14.19	11.85	7.05	3.65	247	186
2-3	48.76	47.88	47.88	47.51	46.70	45.47	43.35	42.85	41.86	41.31	39.91	37.27	33.19	30.59	26.94	21.78	16.28	11.78	7.43	4.37	2.30	306	149
2-4	63.82	60.70	54.56	48.70	41.11	36.46	33.60	31.75	29.45	27.06	24.81	21.51	18.75	16.22	13.30	10.26	7.26	6.41	4.58	3.71	3.13	382	244
2-5	58.34	56.54	53.03	48.50	45.91	43.29	40.03	36.55	33.17	28.00	24.13	21.96	19.05	16.87	13.56	12.24	9.01	6.87	5.20	4.44	2.73	461	269
2-6	45.91	43.09	40.80	37.39	34.44	29.25	26.02	22.68	18.30	15.31	13.51	11.88	9.44	7.69	6.63	6.08	4.87	3.79	3.22	2.93	2.24	625	287
3-1	75.91	75.17	74.23	73.94	73.61	72.80	71.26	69.98	67.44	64.04	61.38	58.95	56.04	48.76	38.86	29.63	21.89	14.75	10.36	5.95	3.61	374	284
3-2	55.69	53.88	51.79	49.25	47.45	44.68	42.36	38.94	35.78	33.37	30.76	27.74	23.80	19.44	17.36	14.64	11.38	8.52	5.97	4.27	2.68	525	292
3-3	69.29	68.93	68.57	68.57	67.84	67.12	66.46	65.02	62.29	58.90	54.59	48.19	39.88	33.36	28.41	23.13	16.88	12.32	8.47	5.40	3.42	553	383
3-4	47.17	41.93	37.45	33.01	30.19	28.24	26.98	25.32	23.35	22.20	20.84	19.44	17.58	13.43	10.79	8.90	6.71	4.59	3.87	2.92	2.25	462	218
3-6	71.55	68.37	66.11	62.53	59.65	57.80	54.66	51.87	48.97	43.10	38.46	32.68	27.81	23.78	18.83	15.80	12.72	9.62	6.93	4.66	3.46	347	248
3-7	83.88	82.61	80.21	77.33	74.50	70.32	68.13	66.77	64.49	62.00	59.77	56.98	53.19	49.29	45.21	41.51	34.40	27.19	18.33	11.96	4.09	494	414
3-8	94.20	92.63	88.40	78.02	70.88	65.48	59.88	57.20	54.50	52.30	51.39	49.87	48.50	46.52	44.50	42.06	36.10	29.25	21.11	12.69	4.60	369	348
3-9	76.03	74.14	70.93	68.23	65.72	60.78	58.69	54.97	51.84	49.85	45.58	41.71	36.65	31.57	25.14	19.67	16.30	12.67	9.67	6.76	3.70	486	370
3-11	64.22	61.26	60.10	58.95	58.41	57.60	56.64	55.90	55.10	53.88	51.14	48.94	43.49	39.16	33.72	25.79	18.35	12.85	9.28	4.45	2.85	184	118
4-1	51.39	46.53	43.73	41.21	39.00	37.65	36.00	34.66	32.85	31.78	30.35	28.73	26.65	23.60	20.11	17.70	15.03	11.56	8.18	4.48	2.51	906	465
4-2	57.32	54.54	52.13	51.03	50.14	49.17	47.85	44.47	41.24	39.12	35.80	32.55	29.44	25.24	20.73	17.44	12.54	8.84	6.12	4.97	2.81	903	518
4-4	1.56	1.56	1.36	1.17	1.17	0.97	0.97	0.97	0.78	0.58	0.58	0.58	0.58	0.58	0.19	0.19	0.19	0.19	0.19	0.19	0.19	514	8
4-5	75.83	73.42	72.17	70.76	69.69	68.61	67.73	66.85	66.05	65.09	62.79	60.95	57.47	52.24	47.98	40.33	33.73	25.14	16.55	9.31	3.69	404	306
4-7	74.49	71.22	68.89	64.67	59.17	52.43	45.67	41.57	35.51	31.13	26.99	21.21	19.55	16.95	14.04	12.31	10.76	9.22	7.07	5.58	3.61	430	320
4-8	5.16	4.73	4.73	4.73	4.52	4.09	3.44	3.01	3.01	3.01	2.80	2.80	2.37	2.15	2.15	1.72	1.29	0.65	0.43	0.43	0.22	465	24
4-9	59.90	59.29	58.25	57.71	56.77	56.40	54.96	52.36	50.63	45.94	42.76	39.43	34.58	30.63	24.82	20.36	14.46	9.86	6.83	4.18	2.87	383	230
4-10	76.11	75.43	74.79	73.75	72.34	69.78	66.92	63.82	61.29	55.29	48.53	42.16	37.32	32.24	25.97	21.20	16.95	12.46	9.20	5.86	3.69	394	300

Note: CC_F is the percent canopy density estimated at relative canopy height F, where $F = 0$ was fixed at 2 m above the ground surface, and $F = 1$ was equal to $Lh_{0.95}$ or the 95th percentile of LiDAR canopy height.

2.6.3. Predictive modelling and spatial extrapolation of inventory attributes

We use ordinary least-squares (OLS) regression to predict cell-level variations in selected inventory variables (i.e., top height, QMD, BA, VOL, and TAGB) across BCGS map sheet 92P.069 (and the surveyed portion of the TFL) based on findings presented by Næsset et al. (2005). We chose to fit the linearized form of a nonlinear (multiplicative) regression model based on Næsset (2002, 2004) and Lim and Treitz (2004), and through exploratory analyses using a large, well-documented inventory and LiDAR dataset derived from 205 simulated forest stands (Frazer et al., 2005; Frazer, 2007):

Multiplicative form: $$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \times \ldots \times x_n^{\beta_n} \varepsilon$$

Linearized form: $$\ln y = \ln \beta_0 + \ln \beta_1 x_1 + \ldots + \ln \beta_n x_n + \ln \varepsilon$$

where y is a field-measured inventory attribute (i.e., H_L, H_{100T}, H_{100D}, QMD, BA, VOL, or TAGB); x_1 to x_n are independent or predictor variables (biometrics) extracted from the LiDAR dataset; β_0 is the y-intercept when all xs are equal to zero; β_1 to β_n are partial regression coefficients estimated by OLS regression procedures; ε is the error term, and ln denotes the natural logarithm.

Our procedure for fitting each of the seven separate regression models was iterative and involved a number of steps. First, we examined simple correlations (Pearson product-moment correlations) found between log-transformed inventory variables and log-transformed LiDAR biometrics to identify potential candidate predictors. Second, we investigated simple correlations among LiDAR biometrics to avoid the inclusion of redundant variables into any one model specification. Third, we kept the model as simple as possible (i.e., a minimum number of predictors) to prevent model overfitting. Fourth, we eliminated obvious leverage points (i.e., remote points on the regression line) and influential observations (i.e., outliers far off the regression line) prior to fitting the final models (Montgomery et al., 2006). Last, we selected models that maximized the coefficient of determination (R^2) and minimized the root-mean-square error (RMSE), mean absolute error (MAE), and error bias of the predicted values in their original arithmetic units of measure:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{1/2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)$$

where n is the total number of observations or samples, and $\hat{y}_i$ and y_i are respectively the predicted and observed (ground-reference) values associated with observation i in arithmetic units. Back-transformation of the fitted models to arithmetic units was accomplished following the procedures outlined by Sprugel (1983). We used these back-transformed predicted values to populate each of the 20 m grid cells stored in the image array.

2.7. Extraction of forest inventory attributes

2.7.1. Comparison of VRI and remotely sensed datasets

Canfor Ltd. and the BC Ministry of Forests and Range, Forest Analysis and Inventory Branch, initiated a Vegetation Resources Inventory (VRI), Phase I, for TFL 18 as a baseline for comparison with our remotely sensed inventory estimates. Conventional 1:20,000 scale colour photography (Zeiss RMK Top 15 camera) was acquired for the entire TFL on August 2, 2007, under clear skies (approximately 1 yr. after the remotely sensed data were collected). VRI photos were digitally scanned at 14 microns (28 cm spatial resolution) prior to softcopy photo-interpretation by Forest Dimensions Inc., Saanichton, BC. Photo-interpretation and quality-assurance methods conformed to the latest VRI standards[4] and procedures set by the BC Ministry of Forests and Range. Stand heights were based on estimates of the average height of dominant, co-dominant, and high-intermediate crown classes, and guided by several ancillary information sources, such as ground calls, air calls (helicopter), and softcopy height measurements. All VRI

[4] BC MOFR, Photo Interpretation Procedures Version 2.5 (April 2009); BC MOFR, Vegetation Resources Inventory Photo Interpretation Standards and Quality Assurance Procedures Version 3.2 (April 2009); http://www.for.gov.bc.ca/hts/vri/standards/index.html

attributes were estimated to fall within a certain range of values (e.g. stand height is 30 to 40 m), rather than being expressed as precise, continuous measures.

There were five inventory variables common to the remotely sensed and VRI datasets (i.e., stand height, basal area, crown cover, volume, and leading species). To compare these two separate datasets, we compared the 20 m LiDAR-derived raster surfaces (GeoTIFF format) and the VRI polygonal coverage (ESRI Shapefile), and then computed stand-level statistics (minimum, maximum, mean, and standard deviation) for each inventory variable (i.e., H_L, H_{100T}, H_{100D}, QMD, BA, VOL, TAGB, and canopy cover). We used mean values of LiDAR-derived attributes as stand-level estimates of height, basal area, and crown cover, and regress them against the equivalent VRI attributes. To compare stand-level leading species, we overlaid the VRI polygon files onto the treelist and computed species proportions for each of the 1057 forest cover polygons.

3. Results & Discussion

3.1. LiDAR-based inventory attributes

3.1.1. Ground-reference sampling

Regression-based methods for predicting forest inventory variables using LiDAR-derived canopy height and density metrics have been well described and demonstrated in the research literature (Næsset et al., 2005). Less well developed, however, are reliable techniques for selecting an unbiased sample of ground-reference plots for model building. In this study, we introduce a stratified random sampling approach based on discrete landscape units (strata), each defined by their own distinct vertical canopy structure. Rather than using the original 42 LiDAR biometrics, which were all highly intercorrelated (e.g., absolute Pearson's r averaged 0.745 ± 0.224 s.d. for all unique, non-diagonal pairwise correlations), we instead use the first five non-trivial, synthetic components (95.2% of the total variance explained) derived from PCA to represent multi-dimensional feature space (Figure 13). We employ cluster analysis along with several global stopping rules to identify eight relatively homogeneous canopy classes (Figure 14 and 15), and then randomly sample within each class to derive a list of potential field plots.

Simple 2-D or 3-D scatterplots of the first few non-trivial principal components summarize the magnitude and direction of similarity/dissimilarity in vertical canopy structure among individual sample units (Figure 16). These simple plots also define the extent or breadth of variability in vertical canopy structure that one could expect to find on the TFL. Thus, random sampling within discrete, non-overlapping *k*-means clusters (ideally at one or two standard deviations from the cluster centroid) should, in theory, provide an efficient sampling approach to eliminate the potential bias associated with small or limited sample sizes. Once ground-reference sites have been selected, one way to find out how well they might represent the full range of structural variability is to plot them in dimensionally reduced (PCA) feature space along with a large, random selection of sample units taken from across the TFL (Figure 17). Our ground-reference plots appear to cover the full range of structural variability within the first three dimensions (87% of total variance) very well despite their limited number (n = 35). Nevertheless, several of our field plots also fall well outside the confidence ellipse fit to the distribution of randomly selected sample units. We expect that this is due to the fact that our field sampling was spatially restricted to a single 1:20,000 BCGS mapsheet (92P.069) and largely confined to the ICH zone (i.e., 30 of the 35 plots were located in ICHmk2 or ICHmw3 subzones) while the larger dataset used to define the scatter plots uses the entire project dataset. The ICH zone is noted for its high productivity, species diversity, and structural complexity (Ketcheson et al., 1991), occupies only a small portion of the TFL (16% of the land base), and is therefore likely to be ecological outlier when compared to other more common forest types (ESSF and SBS).

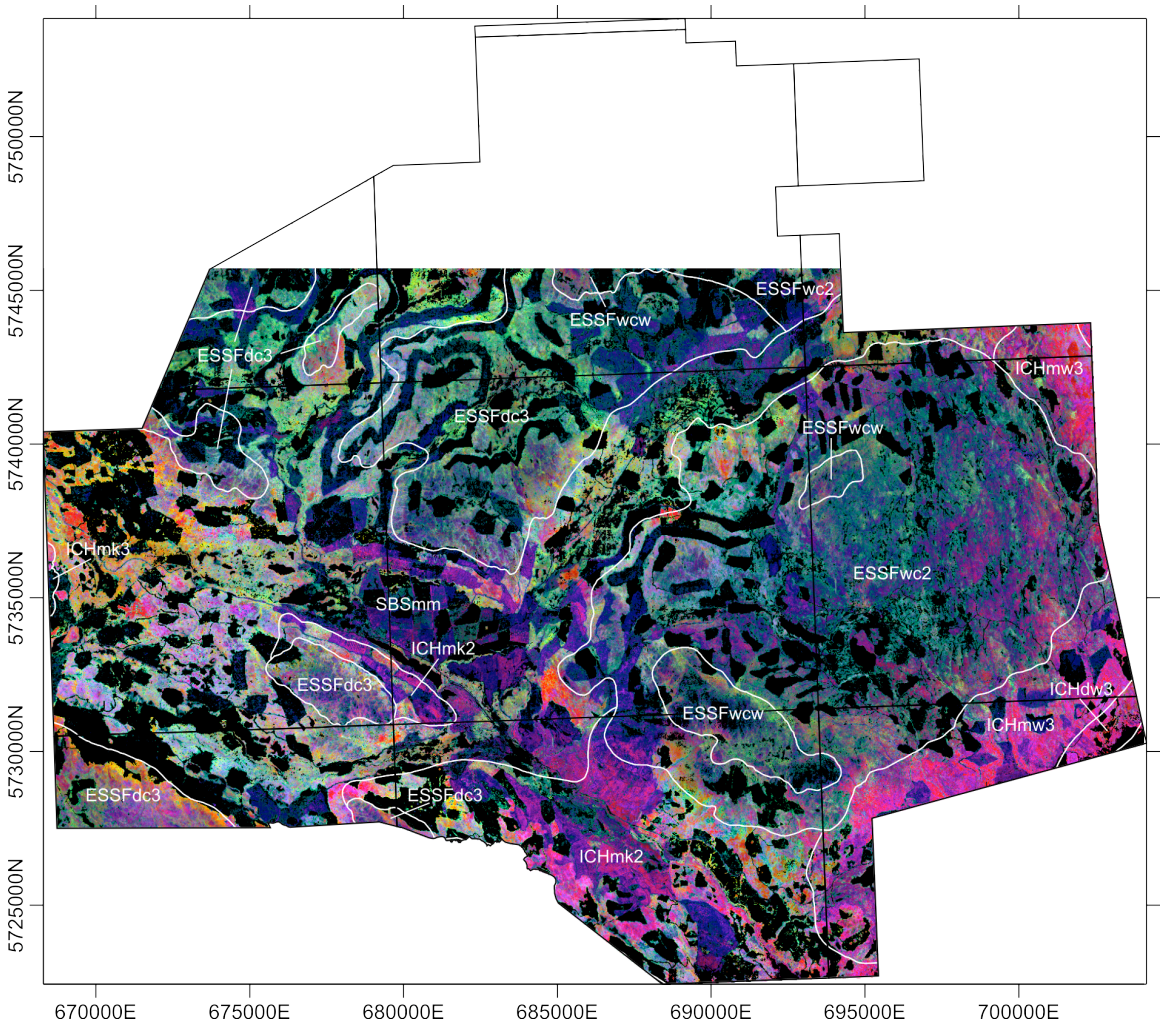


Figure 13. Vertical canopy structure defined by the first 3 principal components of a PCA (86.9% of the total variance explained). Principal components, PC1 (62.4%), PC2 (15.7%), and PC3 (8.9%), occupy the red (R), green (G), and blue (B) colour bands, respectively, of a three band RGB colour display. Black areas denote locations lacking vertical canopy structure (i.e., harvested and non-forested areas). Solid white lines represent biogeoclimatic subzone boundaries. Colour differences indicate that vertical canopy structure varies continuously across the TFL in response to changes in species composition, stand age, disturbance history, and site productivity. The data will find further application in the areas of predictive ecosystem (PEM), fire, habitat, and wood-quality modelling, and as data layers (bands) for object- or region-based stand delineations.

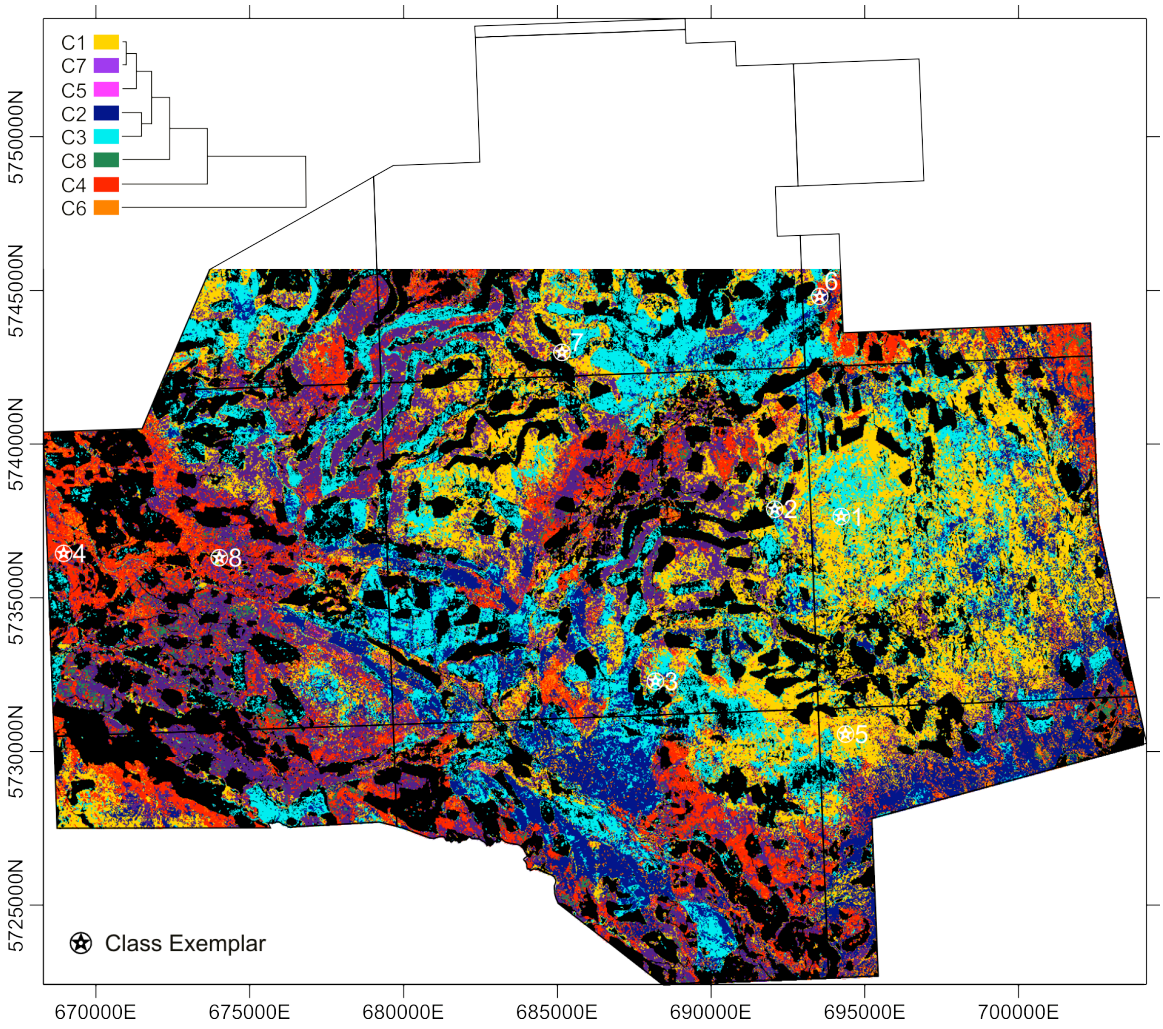


Figure 14. Eight discrete landscape units (vertical canopy structure classes) derived by *k*-means partitioning of the first 5 non-trivial principal components (95.2% of the total variance explained) extracted from the 42 original LiDAR biometrics. We used a *k*-means cluster model based on 30K randomly sampled points to train a maximum likelihood (ML) classifier, which in turn was applied to spatially extend the classification across the TFL. White symbols indicate the location of class exemplars – sample units found closest to their cluster centroid. We used Ward's minimum variance method to summarize the hierarchical connectivity among *k*-means clusters (the dendrogram is scaled to the total sum of squared errors). An unweighted majority filter (3 × 3 kernel) was applied to eliminate fine-scale speckle from the raw ML classification results.

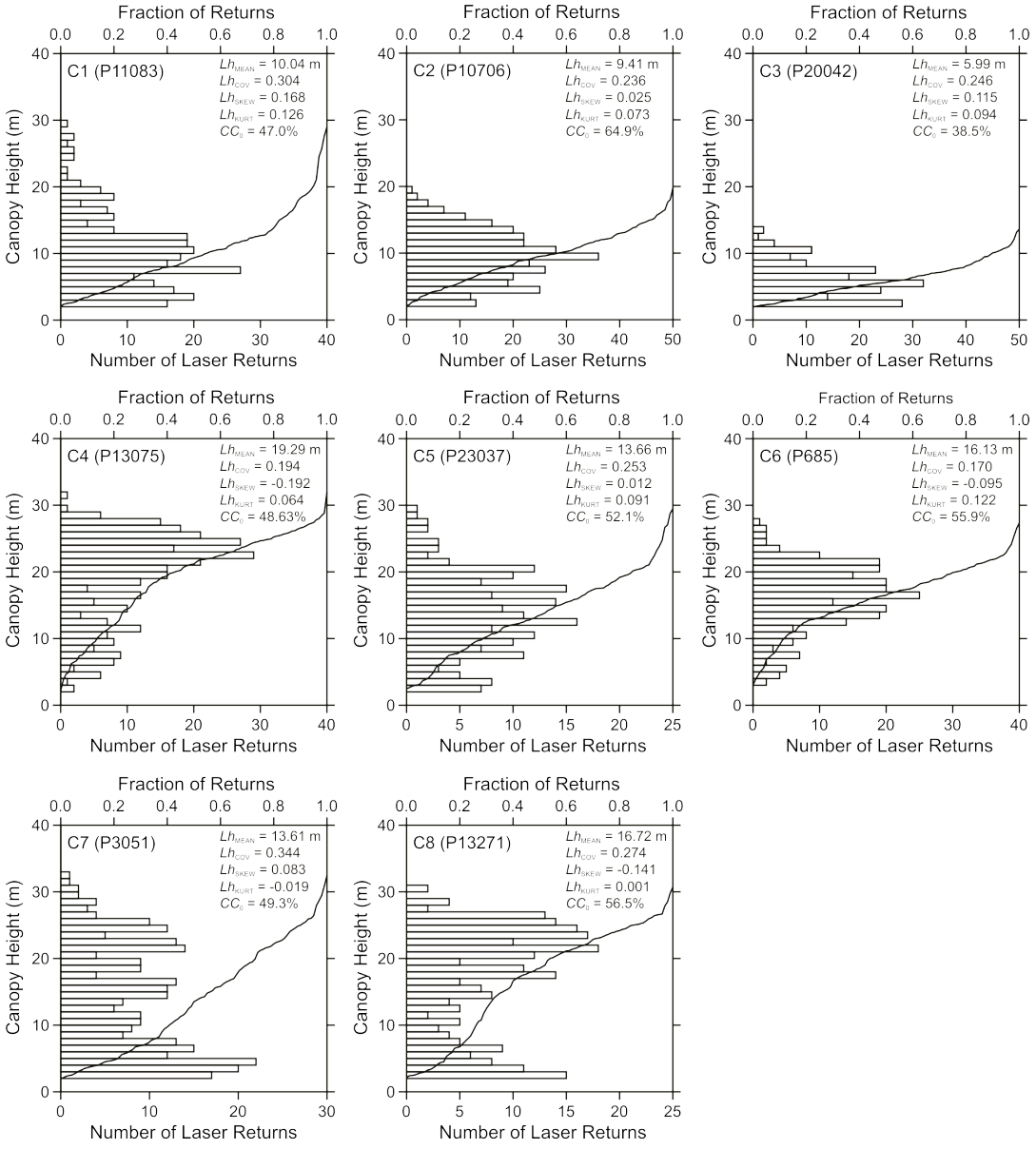


Figure 15. LIDAR canopy height profiles (also known as apparent foliage height profiles) and selected distributional statistics for 8 class exemplars identified by *k*-means partitioning. Exemplars represent the average or 'typical' vertical canopy structure found within a particular cluster or group; they are the sample units found closest to the group centroid. Canopy height profiles are derived from height-frequencies of LiDAR points > 2 m in height. A large frequency of laser returns indicates that many reflective surfaces occur at that particular height interval within the canopy. Canopy height profiles can be described as top-loaded (C4), bottom-loaded (C1 and C3), bimodal (C8), uniform, or normally distributed (C6). The superimposed quantile curve shows the fraction of LiDAR canopy returns found below a given height threshold. The overall shape of this curve can be thought of as a unique structural identifier or "fingerprint".

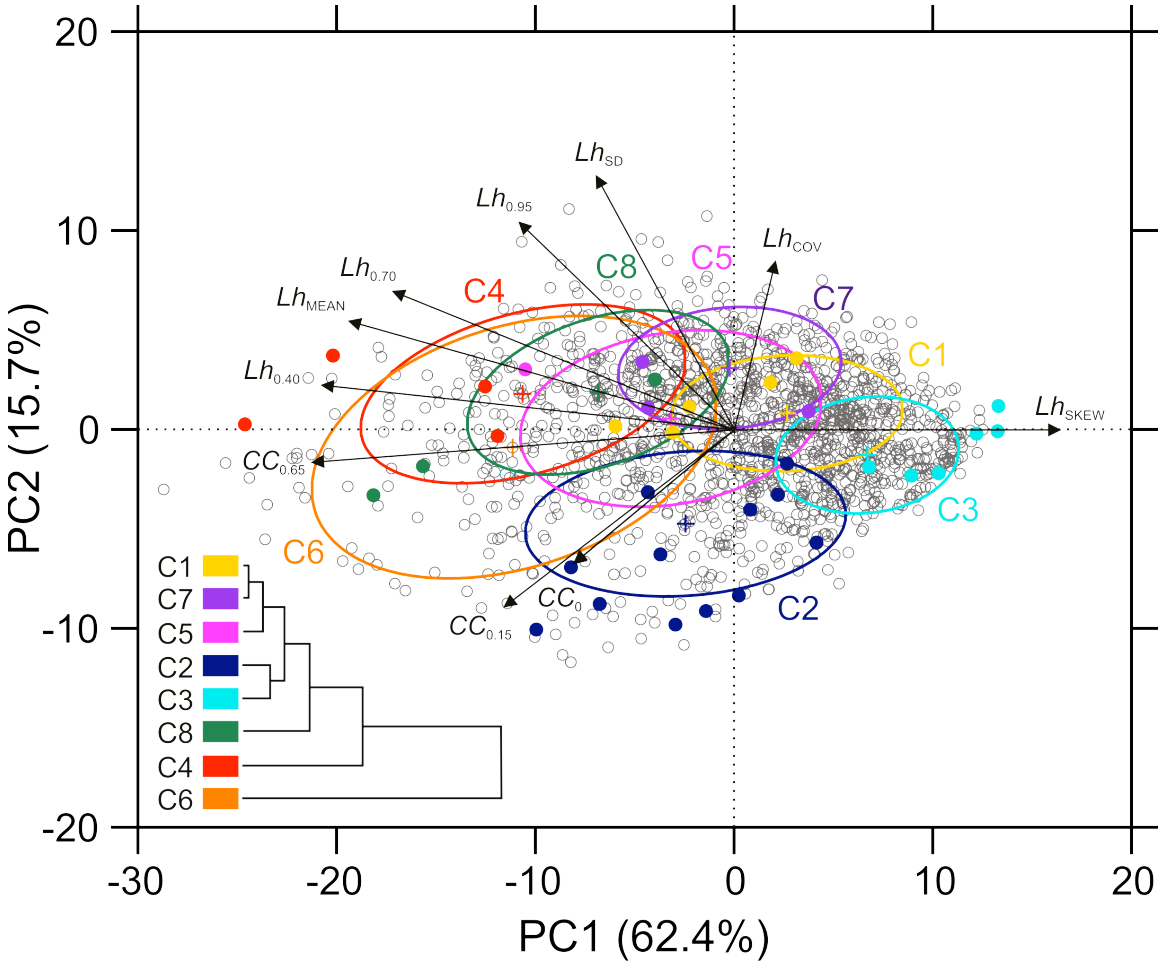


Figure 16. Joint plot of the first two synthetic components (78.1% of the total variance) extracted by PCA from a correlation matrix of LiDAR biometrics. Grey open circles denote sample units (20 m grid cells) randomly selected from across the TFL, while closed, coloured circles show the distribution of our ground-reference sites. Coloured ellipses are 2-D confidence limits (one standard deviation) belonging to each of the eight *k*-means clusters (C1–C8). Plus signs (+) identify the cluster centroid (location of class exemplars). The direction and length of the labeled arrows summarize the strength and direction of simple (Pearson's) correlations between the original LiDAR biometrics and the first two principal components. For example, Lh_{Mean} is negatively correlated ($r = -0.88$) with PC1, and positively correlated ($r = 0.47$) with PC2. The wide-ranging scatter of ground-reference sites (at least in the first two synthetic dimensions) indicates that this small sample ($n = 35$) covers a broad range of structural conditions within the TFL and may therefore be relatively unbiased. The dendrogram (scaled to the total sum of squared errors) summarizes the hierarchical connectivity between clusters. For example, C1 and C7 are found close together in feature space (i.e., they have similar canopy structures), and are therefore the first clusters to combine at $k = 7$. C6 is the last cluster to join at $k = 1$, due to its structural uniqueness.

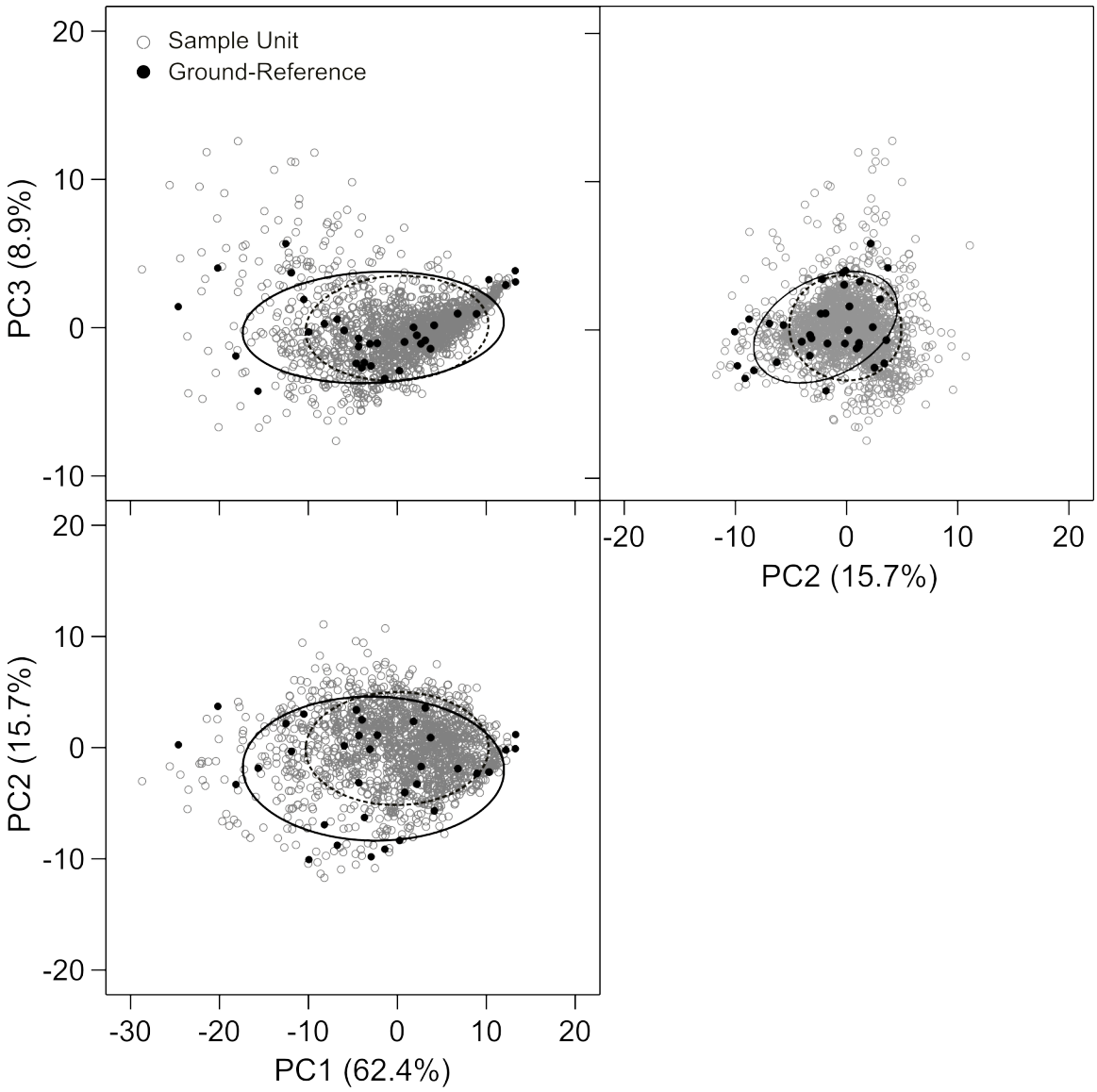


Figure 17. The distribution of ground-reference sites (n = 35) and randomly selected sample units in the first three synthetic dimensions (87% of the total variance explained) of reduced d-dimensional space. Substantial overlap between confidence ellipses (p = 0.683) constructed for each of these two distributions (broken line for sample units and solid line for ground-reference plots) suggests that our ground-reference plots cover a broad range of feature space, show little obvious bias, and may therefore be well selected.

3.1.2. Regression model building and performance

We relied on two classes of LiDAR biometrics to model and predict forest inventory variables for BCGS map sheet 92P.069 and across the remainder of the TFL. The first group or class of biometrics (i.e., LiDAR canopy height quantiles and moments) was derived exclusively from LiDAR points >2 m in height. These variables describe the vertical distribution of canopy heights and are highly correlated with the size distribution of individual trees (Gobakken and Næsset, 2005). The second class of biometrics was related to the density of LiDAR points found at increasing depths (top to bottom) within the forest canopy and is tightly coupled to vertical changes in canopy density and cover. We found that field measures of tree size (i.e., stand height and diameter) were well predicted by the first class of LiDAR biometrics (i.e., canopy height statistics), while basal area (BA), gross volume (VOL), and total aboveground biomass (TAGB) also required the addition of canopy density metrics (CC_F) to improve model fit.

Coefficients of determination (R^2) ranged between 0.86 and 0.95 and averaged 0.91 for all back-transformed regression models (Table 4 and Figure 18). The statistical relationship between LiDAR and field-measured inventory variables was strongest ($0.92 \leq R^2 \leq 0.95$) for estimates of stand height (i.e., H_L, H_{100T}, and H_{100D}) and VOL, and weakest ($0.86 \leq R^2 \leq 0.89$) for attributes more directly dependent on tree diameter, such as quadratic mean diameter (QMD), BA, and TAGB. The Root Mean Square Error (*RMSE*), when expressed as a percentage of the mean, was lowest ($10.7\% \leq RMSE \leq 15\%$) for tree-size attributes (i.e, H_L, H_{100T}, H_{100D}, and QMD) and highest ($19.9\% \leq RMSE \leq 21.8\%$) for inventory attributes requiring height (*Lh*) and density (CC_F) estimates as explanatory variables (i.e., BA, VOL, and TAGB). Prediction bias was always positive, but very close to zero for each of the regression models. Comparison of these regression results with fit statistics reported by others is complicated by inherent differences in the scale of analysis (field plot size) and magnitude and range of the ground-reference data. Nevertheless, our R^2, *RMSE*, and bias estimates appear similar in magnitude to those presented by others (Means et al., 2000; Næsset et al., 2004; Reutebuch et al., 2005).

Table 4. Ordinary least-squares (OLS) regression models and associated fit statistics describing the statisti
between ground-reference measurements (response) and LiDAR canopy height and density metrics (predic
sheet 092P.069 (TFL 18, Clearwater, BC).

Inventory Attribute	**Linear Regression Equation**	**Units**	***BCF***	***R^2***
Lorey's Mean Height	$\ln H_L = 0.7779 + 0.8593 \ln Lh_{Mean}$	m	1.0095	0.94
Average Height of 100 Tallest Trees	$\ln H_{100T} = 1.0552 + 0.7751 \ln Lh_{0.70}$	m	1.0094	0.95
Average Height of 100 Largest DBH Trees	$\ln H_{100D} = 0.9882 + 0.7800 \ln Lh_{0.70}$	m	1.0113	0.92
Quadratic Mean Diameter	$\ln QMD = 1.3074 + 0.6674 \ln Lh_{0.40}$	cm	1.0185	0.88
Basal Area	$\ln BA = 0.3348 + 0.7459 \ln Lh_{0.55} + 0.3302 \ln CC_0$	m^2/ha	1.0300	0.86
Gross Volume	$\ln VOL = -0.0763 + 1.4534 \ln Lh_{0.70} + 0.4191 \ln CC_0$	m^3/ha	1.0381	0.94
Total Aboveground Biomass (Dry Weight)	$\ln TAGB = 0.0914 + 1.0496 \ln Lh_{0.55} + 0.5719 \ln CC_0$	Mg/ha	1.0531	0.89

Notes for Table 4:

(1) Bias correction factor, $BCF = \exp\left(SEE^2/2\right)$ used to back-transform data to arithmetic units, where *SEE* is the standard error of the estimate computed by linear regression (Sprugel, 1981); e.g., $\hat{H}_L = \exp\left(0.7779 + 0.8593 \ln Lh_{Mean}\right) \times BCF$.

(2) Root-mean-square error, $RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2\right]^{1/2}$.

(3) Mean absolute error, $MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|$.

(4) Bias, $Bias = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)$.

(5) R^2, *RMSE*, *MAE*, and *Bias* statistics were all computed using back-transformed estimates of predicted $\hat{y}_i$ and the original (untransformed) field measurements of y_i; n is the number of samples, and p is the number of fitted coefficients in the model, including the intercept.

Total number of ground-reference plots was 35.

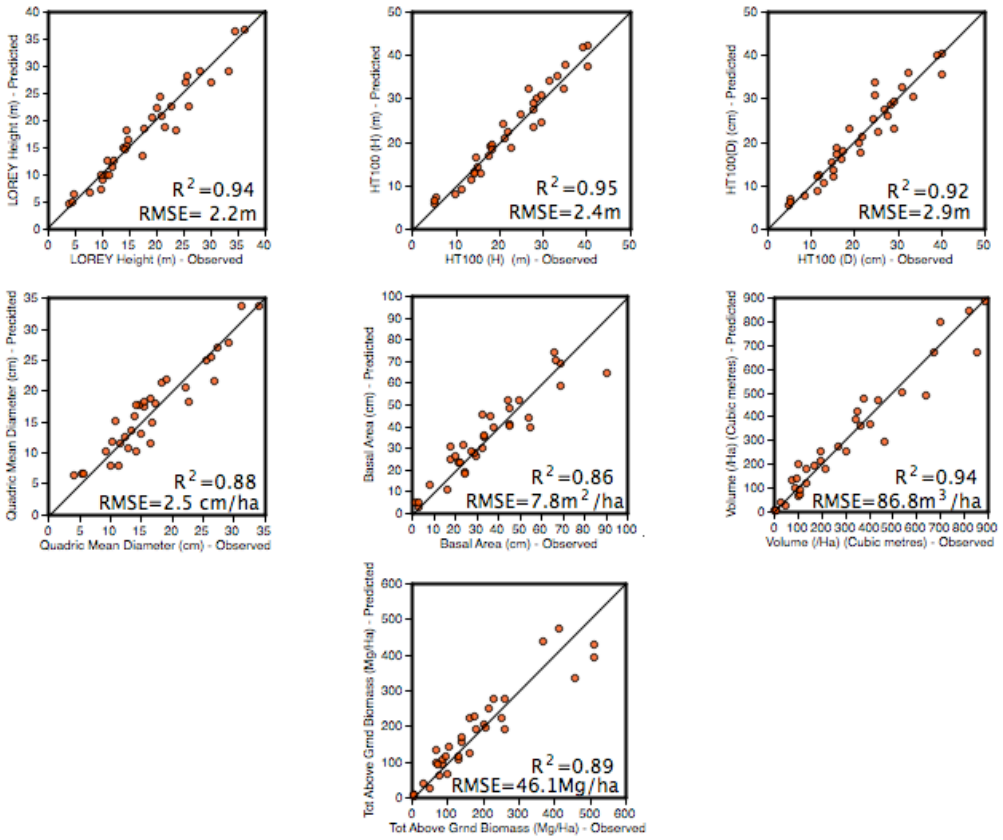


Figure 18. Scatterplots of observed (ground) versus predicted (LiDAR) estimates of Lorey's mean height (H_L), average height of the 100 tallest (H_{100T}) and largest diameter (H_{100D}) trees per hectare, quadratic mean diameter (QMD), basal area (BA), total volume, and total aboveground biomass (dry weight). Predicted values have been back transformed to arithmetic units following the methods of Sprugel (1983). The solid diagonal is the 1:1 line.

We expect that a number of factors have negatively impacted the statistical models and results summarized above: (1) ground-reference data were collected two years after the LiDAR survey, and no attempt was made to account for structural changes related to tree growth, mortality, or damage (TFL 18 is currently experiencing a MPB epidemic); (2) GPS error resulted in a horizontal offset of unknown distance between the LiDAR and ground-reference datasets (we were able to assess GPS precision, but not accuracy); (3) LiDAR canopy height and density metrics were distorted by the inclusion and/or exclusion of partial crowns located along the plot boundaries (edge effect), as well as by the presence of leaning, broken, or MPB defoliated trees; (4) small plots, such as the ones used in this study (i.e.,10 m fixed radius), are more susceptible to the deleterious effects of GPS positional error and boundary issues, and also capture substantially less spatial variability compared to larger sample plots (Zenner, 2005; Gobakken and Næsset, 2008); (5) regression models have been incompletely specified (i.e., they omit key explanatory variables), and are also unlikely to meet the critical assumption of stationarity (i.e., the efficacy of a predictor variable will not change across the study area; Fotheringham et al., 2002), due to strong local to regional scale differences in species composition, disturbance regime, site productivity, and other site-related characteristics.

We suggest that the following recommendations may potentially improve model fit and behaviour for future LiDAR surveys: (1) ground-reference samples and remotely sensed data should be collected in same year and preferably during the same part of the growing season to eliminate structural discrepancies induced by growth, mortality, disturbance, or phenology; (2) larger ground-reference plots (15 m to 25 m fixed radius) should be installed to dampen the impact of edge effects and GPS positional error, and as a way to capture more of the natural stand variability; (3) the GPS survey needs to be planned as an integral part of the field campaign using mapping or survey grade GPS equipment and procedures to minimize spatial uncertainty; (4) adopt an efficient stratified random sampling approach (like the one implemented in this study) to capture as much regional structural variability as possible; (5) maximize the number of ground-samples collected and always include sufficient replication (Gobakken and Næsset, 2008); (6) improve model specification by including other kinds of plant-relevant explanatory variables extracted from the LiDAR-generated DEM (e.g., Swanson et al., 1988; Franklin, 1995;

Lookingbill and Urban, 2005; Hudak et al., 2008); (7) explore other modelling methods that do not depend on the stationarity assumption (e.g., Kimsey et al., 2008) and allow for complex interactions among predictors (e.g., De'ath and Fabricius, 2000; Hudak et al., 2008).

3.1.3. Map sheet and TFL inventory statistics

We constructed spatially contiguous maps of predicted inventory variables (i.e., H_L, H_{100T}, H_{100D}, QMD, BA, VOL, TAGB, and crown cover) for BCGS mapsheet 92P.069 (Figures 19–25) and for the entire surveyed portion of the TFL (Figures 26–32) using the fitted regression models and back-transformation procedures described above. Because these regression models were developed specifically for map sheet 92P.069 (and almost exclusively for the ICHmk2 subzone), we expect to find greater prediction error and/or bias in areas outside the map sheet and ICHmk2 subzone. Subalpine fir (balsam) crowns, for instance, have a very distinctive architecture (i.e., extremely narrow, long, and dense), and may therefore relate differently to LiDAR canopy height and density metrics than spruce, fir, or pine leading stands (Nelson, 1997). Hence, we anticipate that the largest prediction error and/or bias will be associated with mature, balsam-leading ESSF stands.

We computed basic descriptive statistics for each of the eight LiDAR-derived inventory variables using a binary mask, from a GIS (ESRI) SHAPE file supplied by Canfor Ltd, to identify productive and non-productive areas of the land base. These statistics describe the productive portion of the land base only, and were computed separately for map sheet 92P0.69 and for the surveyed portion of the TFL (Table 5). For example, average Lorey's stand height (H_L) is 13.04 m ± 9.65 m (one standard deviation) for all 212,873 20 m cells found in the productive portion of map sheet 92P.069; H_L is 0 m (i.e., stands had been harvested) in at least 25% of the 212,873 cells, and ≤ 20.4 m in 75% of the cells; maximum H_L is 46.04 m. The 95% confidence limits on the mean are extremely narrow due to the large sample size (n).

Table 5: Cell-level (20 m) inventory statistics computed for BCGS map sheet 92P.069 and for the surveye
Statistics exclude all areas of the land base classified by Canfor as non-productive (e.g., wetlands, outcrop

Region	Attribute	Units	20 m Cell Statistics								
			Min.	P25th	Mean	Median	P75th	Max.	s.d.	LCL	U
92P.069	Lorey	m	0.00	0.00	13.04	13.64	20.40	46.04	9.65	13.00	13.
	H100h	m	0.00	0.00	16.75	17.43	26.88	50.33	12.27	16.69	16.
	QMD	cm	0.00	0.00	12.44	13.18	18.55	39.82	9.02	12.40	12.
	BA	m^2/ha	0.00	0.00	25.26	26.21	38.41	88.78	19.24	25.17	25.
	Volume	m^3/ha	0.00	0.00	201.42	150.81	316.37	1210.05	201.89	200.57	202.
	TAGB	Mg/ha	0.00	0.00	118.92	105.01	178.74	613.04	107.69	118.46	119.
	Cover	%	0.00	18.80	43.00	47.49	64.38	100.00	27.14	42.89	150.
TFL18	Lorey	m	0.00	5.73	13.11	14.03	20.13	46.04	9.28	13.07	13.
	H100h	m	0.00	7.41	16.97	18.04	26.57	50.33	11.95	16.91	17.
	QMD	cm	0.00	7.21	12.25	13.00	17.97	39.82	8.54	12.21	12.
	BA	m^2/ha	0.00	9.08	24.71	25.64	37.73	93.89	18.17	24.61	24.
	Volume	m^3/ha	0.00	20.90	196.83	154.00	317.91	1304.76	187.72	195.86	197.
	TAGB	Mg/ha	0.00	21.27	113.35	101.02	174.59	672.62	98.89	112.84	113.
	Cover	%	0.00	23.50	40.92	45.13	57.56	100.00	23.83	40.79	41.

Note: P25th – 25th percentile; P75th – 75th percentile; s.d. – standard deviation
LCL – lower 95% confidence limit for mean; UCL – upper 95% confidence limit for mean
Lorey – Lorey's mean stand height
H100h – mean height of the tallest 100 trees per hectare (pro-rated to tallest 4 trees per plot)
QMD – quadratic mean diameter; BA – basal area; Volume – total or gross volume
TAGB – total aboveground biomass (dry weight); Cover – vertically projected canopy cover

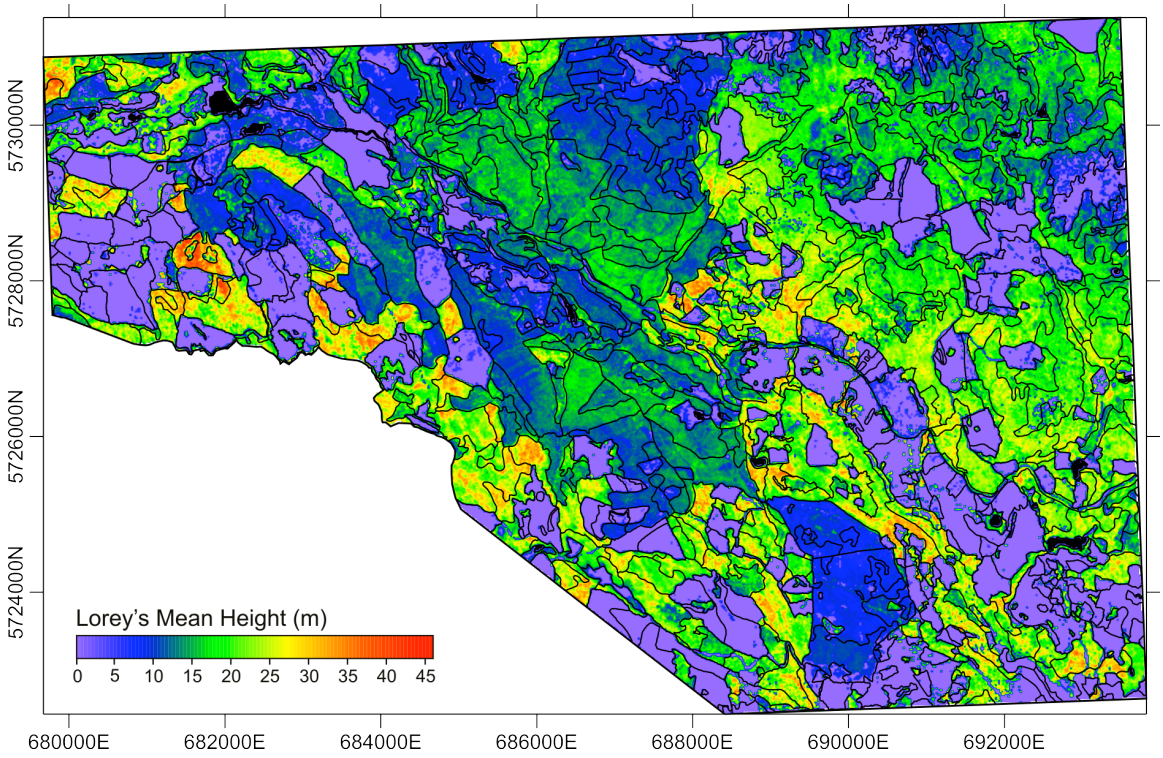


Figure 19. LiDAR predicted Lorey's mean height (H_L). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 m, 46.03 m, 13.04 m, and 9.65 m, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

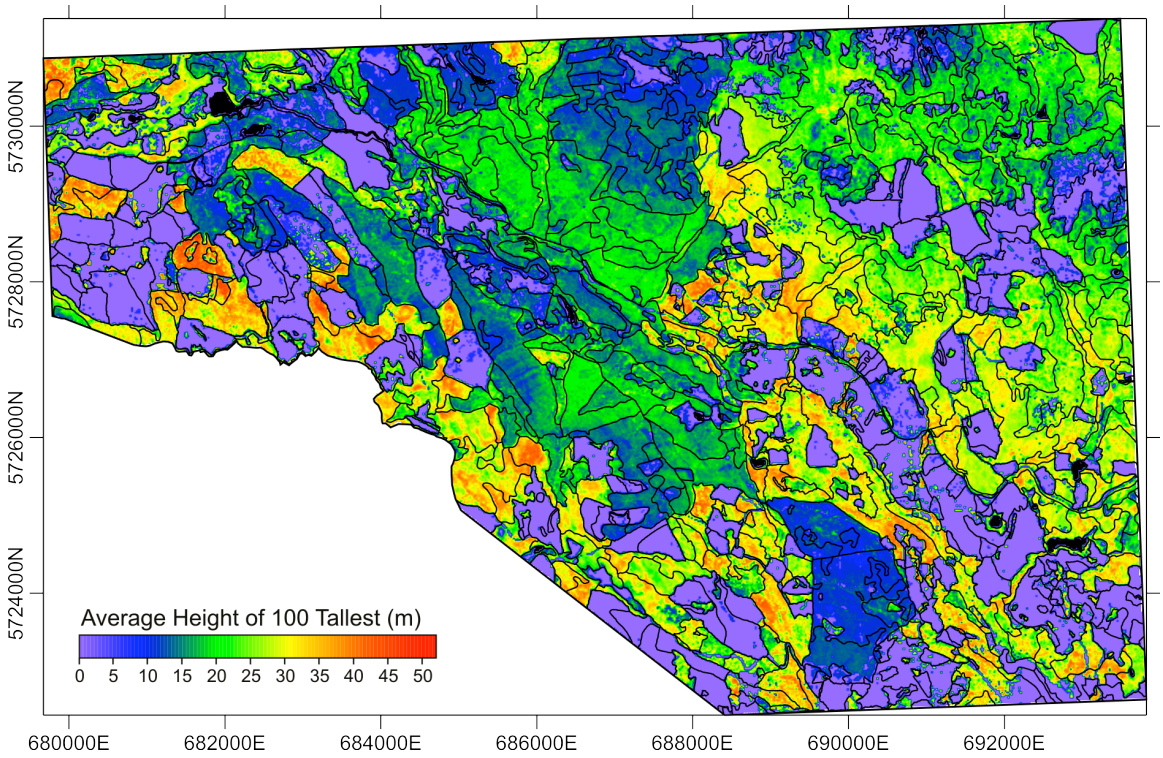


Figure 20. LiDAR predicted average height of the tallest 100 trees per hectare (H_{100H}). Minimum, maximum, mean, and standard deviation statistics for the map sheet (212,873 20 m grid cells) are 0 m, 50.33 m, 16.75 m, and 12.27 m, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

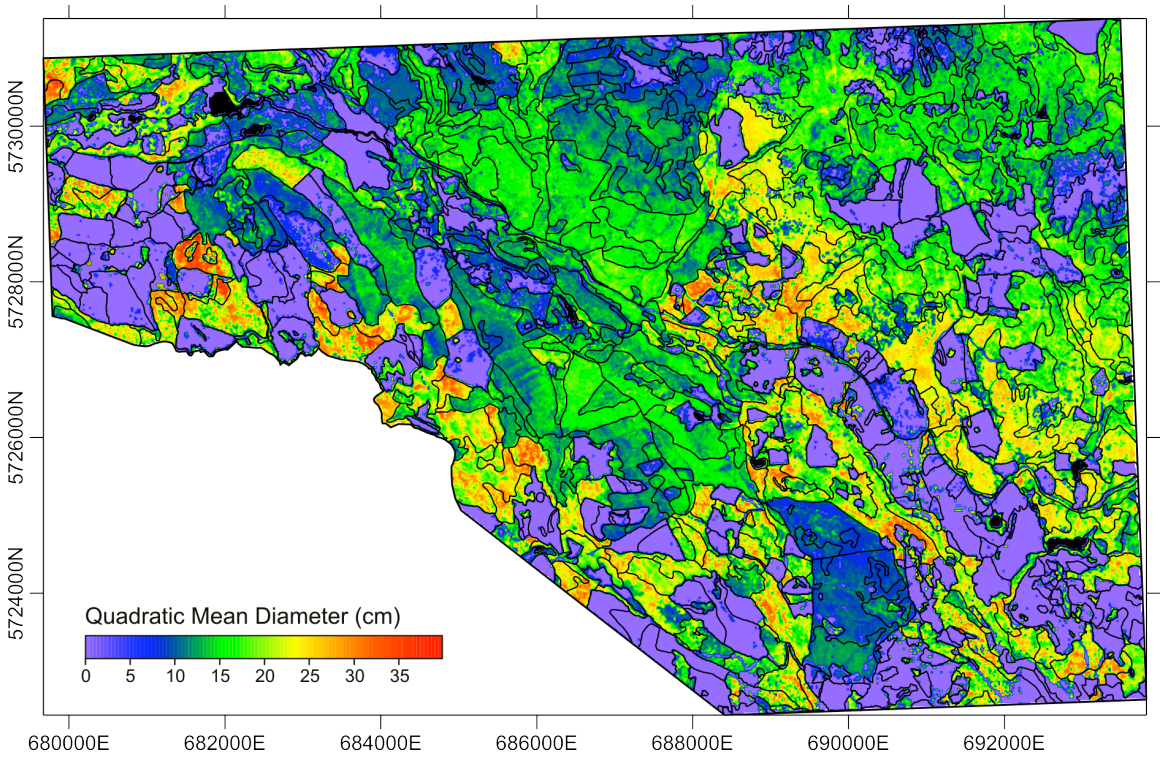


Figure 21. LiDAR predicted quadratic mean diameter (QMD). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 cm, 39.82 cm, 12.44 cm, and 9.02 cm, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

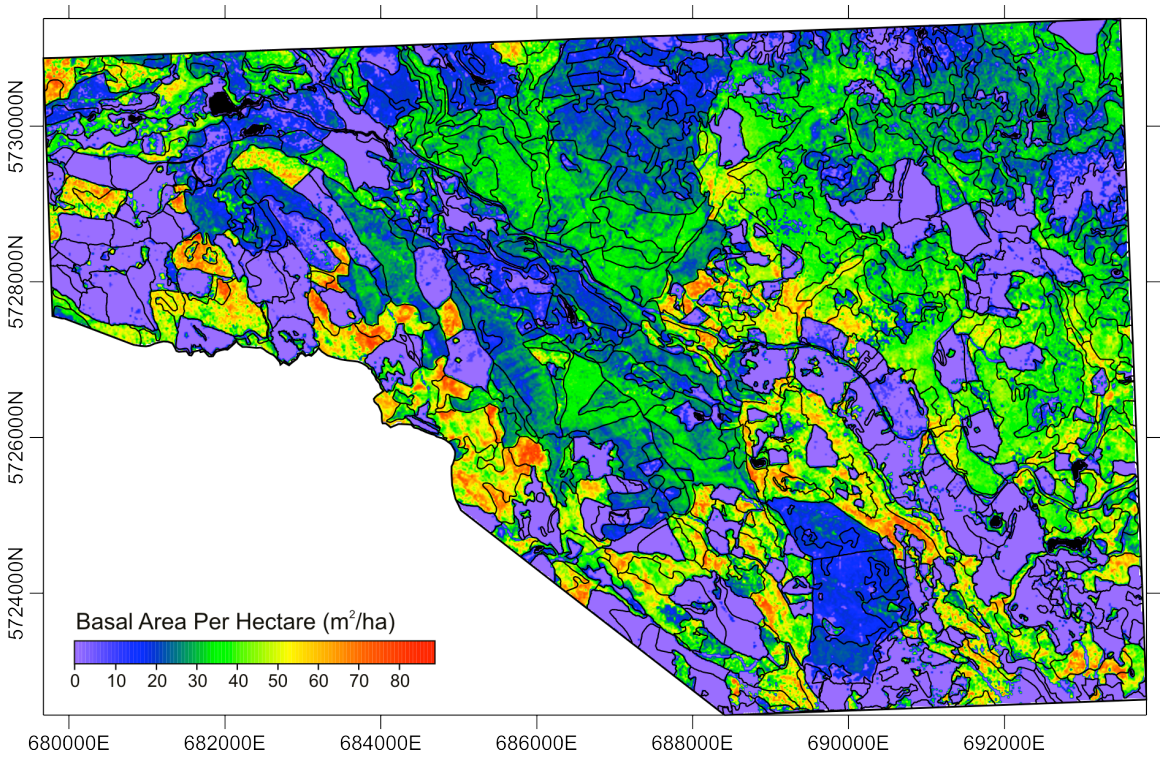


Figure 22. LiDAR predicted basal area per hectare (BA). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 m^2/ha, 88.78 m^2/ha, 25.26 m^2/ha, and 19.24 m^2/ha, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

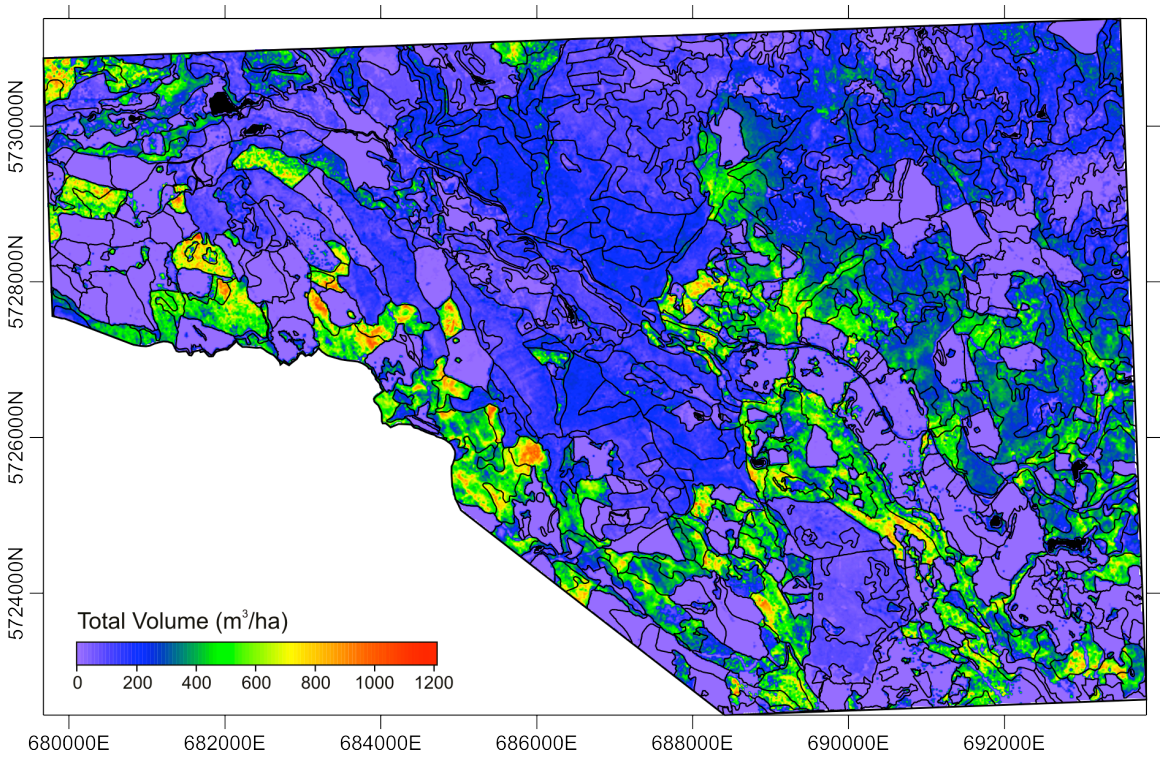


Figure 23. LiDAR predicted total volume per hectare (VOL). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 m^3/ha, 1210.05 m^3/ha, 201.42 m^3/ha, and 201.89 m^3/ha, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

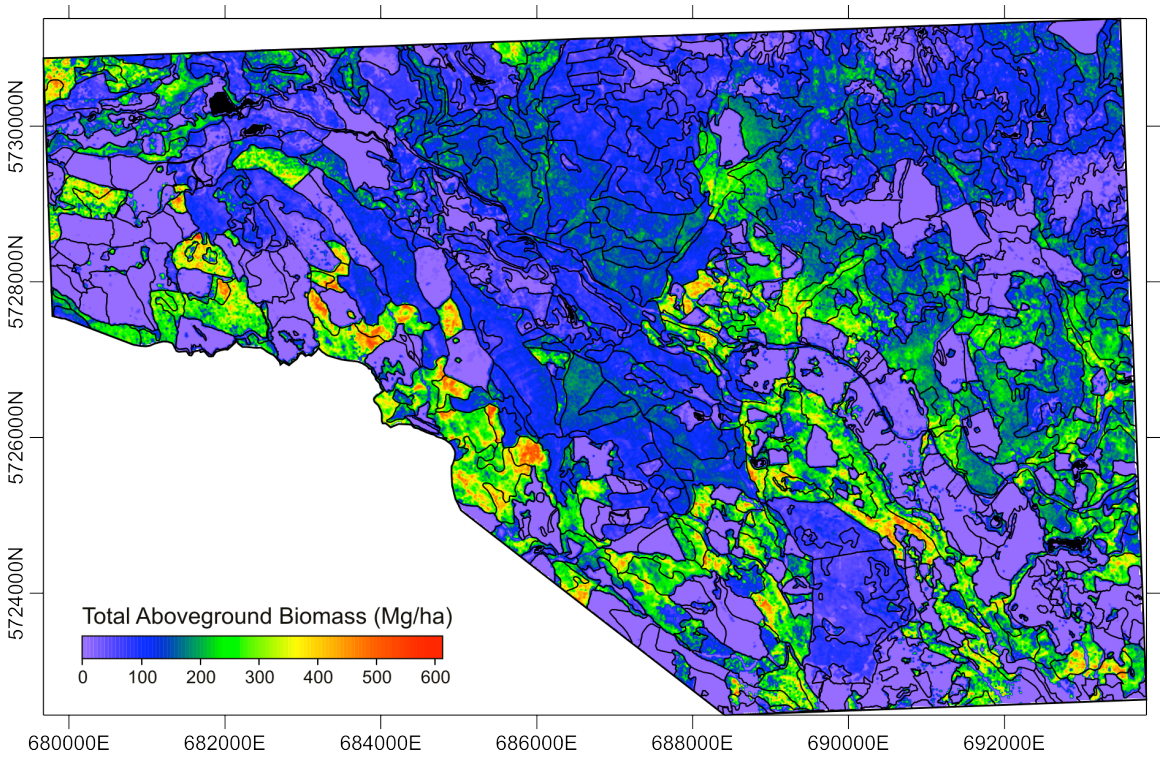


Figure 24. LiDAR predicted total aboveground biomass (dry weight) per hectare (TAGB). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 Mg/ha, 613.04 Mg/ha, 118.92 Mg/ha, and 107.69 Mg/ha, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

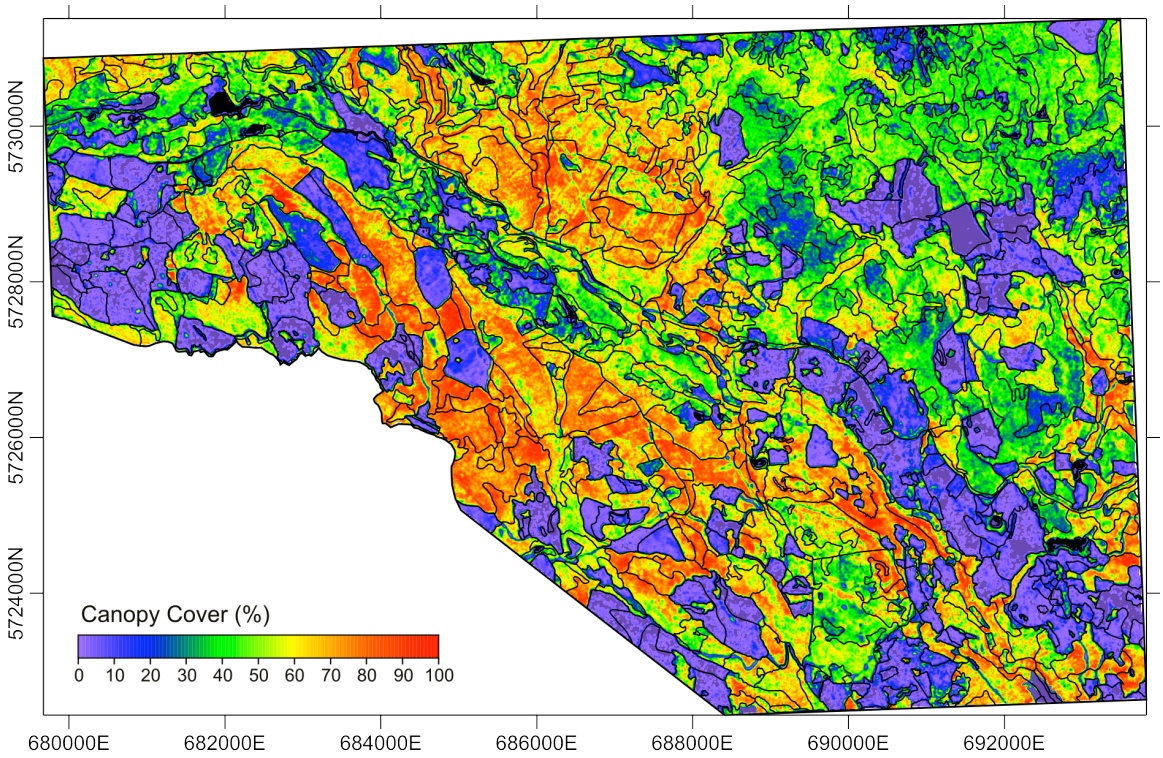


Figure 25. LiDAR predicted crown or canopy cover (reported in the VRI as crown closure). Minimum, maximum, mean, and standard deviation statistics for the map sheet are 0 %, 100 %, 43 %, and 27.14 %, respectively. Forest cover polygons were independently produced by VRI photo-interpreters using 1:20,000 colour photography collected in 2007.

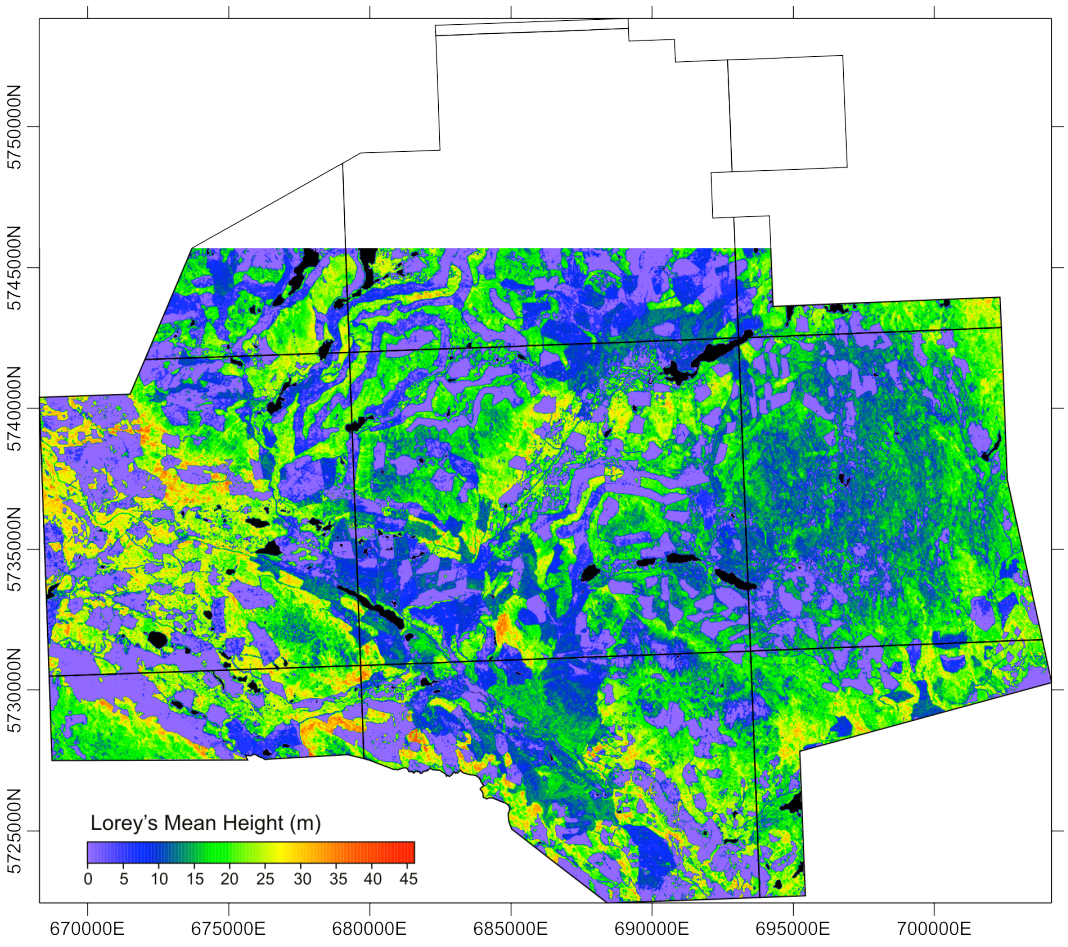


Figure 26. LiDAR predicted Lorey's mean height (H_L). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 m, 46.04 m, 13.11 m, and 9.28 m, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

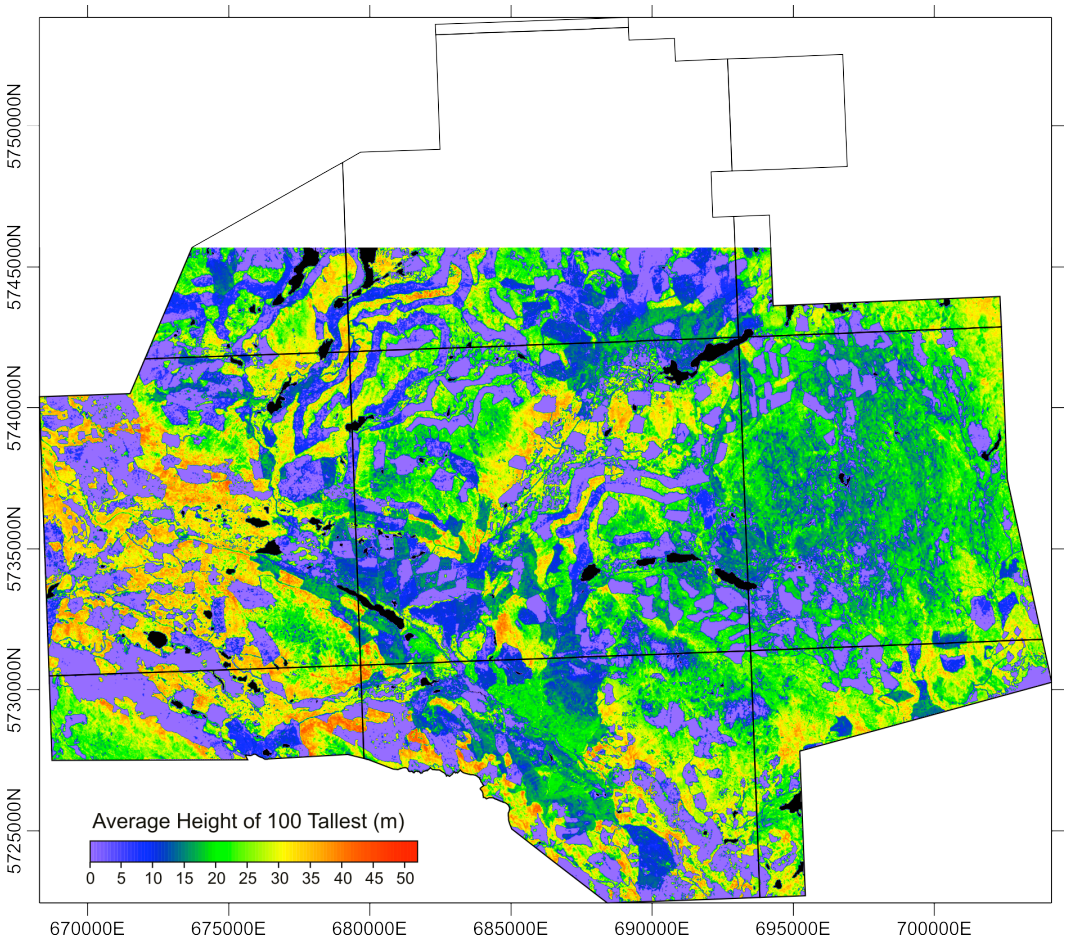


Figure 27. LiDAR predicted average height of the tallest 100 trees per hectare (H_{100T}). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 m, 50.33 m, 16.97 m, and 11.95 m, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

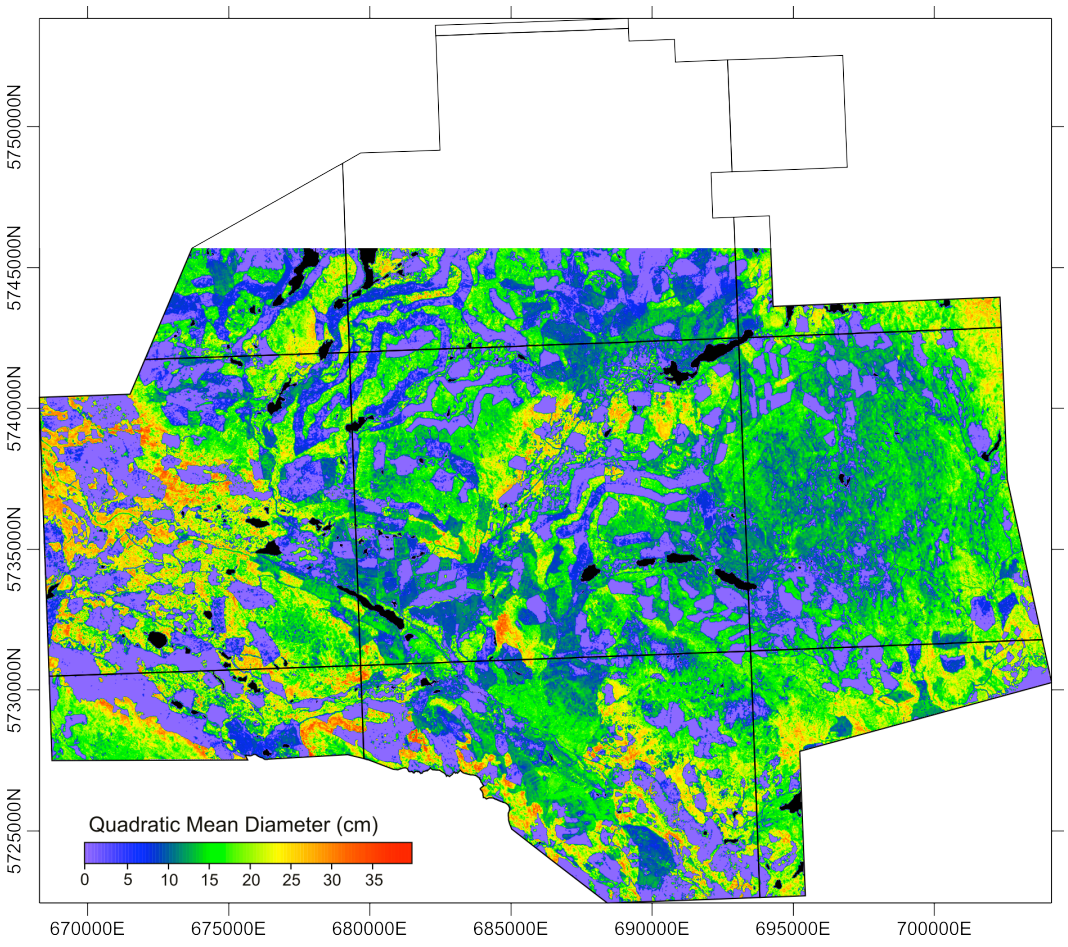


Figure 28. LiDAR predicted quadratic mean diameter (QMD). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 cm, 39.82 cm, 12.25 cm, and 8.54 cm, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

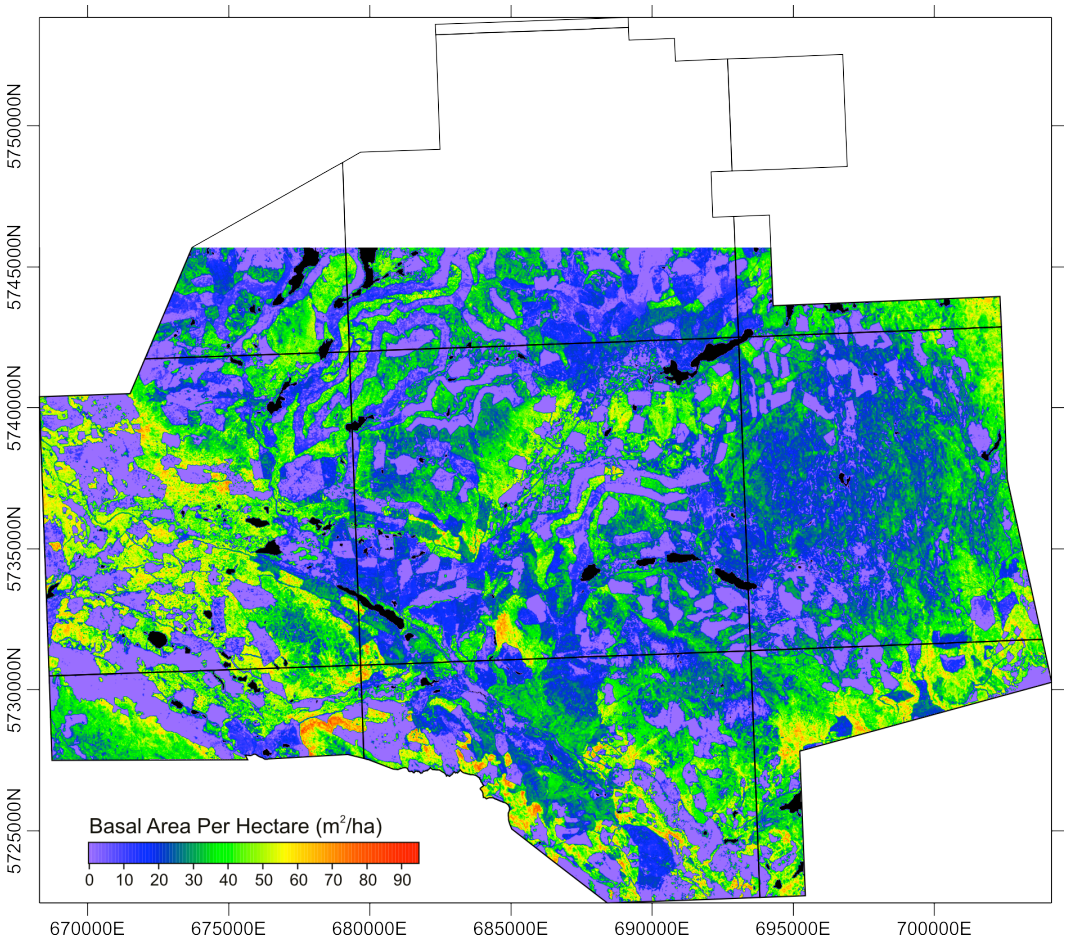


Figure 29. LiDAR predicted basal area per hectare (BA). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 m^2/ha, 93.89 m^2/ha, 24.71 m^2/ha, and 18.17 m^2/ha, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

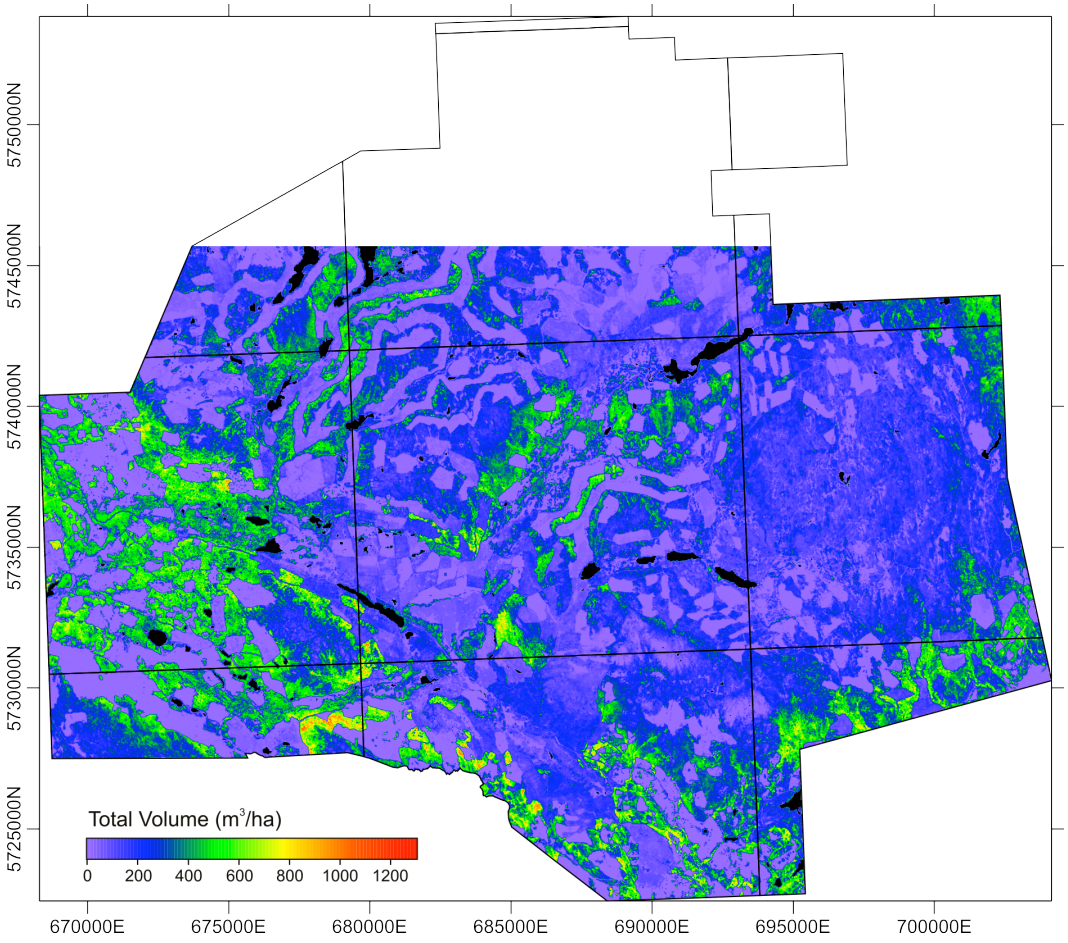


Figure 30. LiDAR predicted total volume per hectare (VOL). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 m^3/ha, 1304.76 m^3/ha, 196.83 m^3/ha, and 187.72 m^3/ha, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

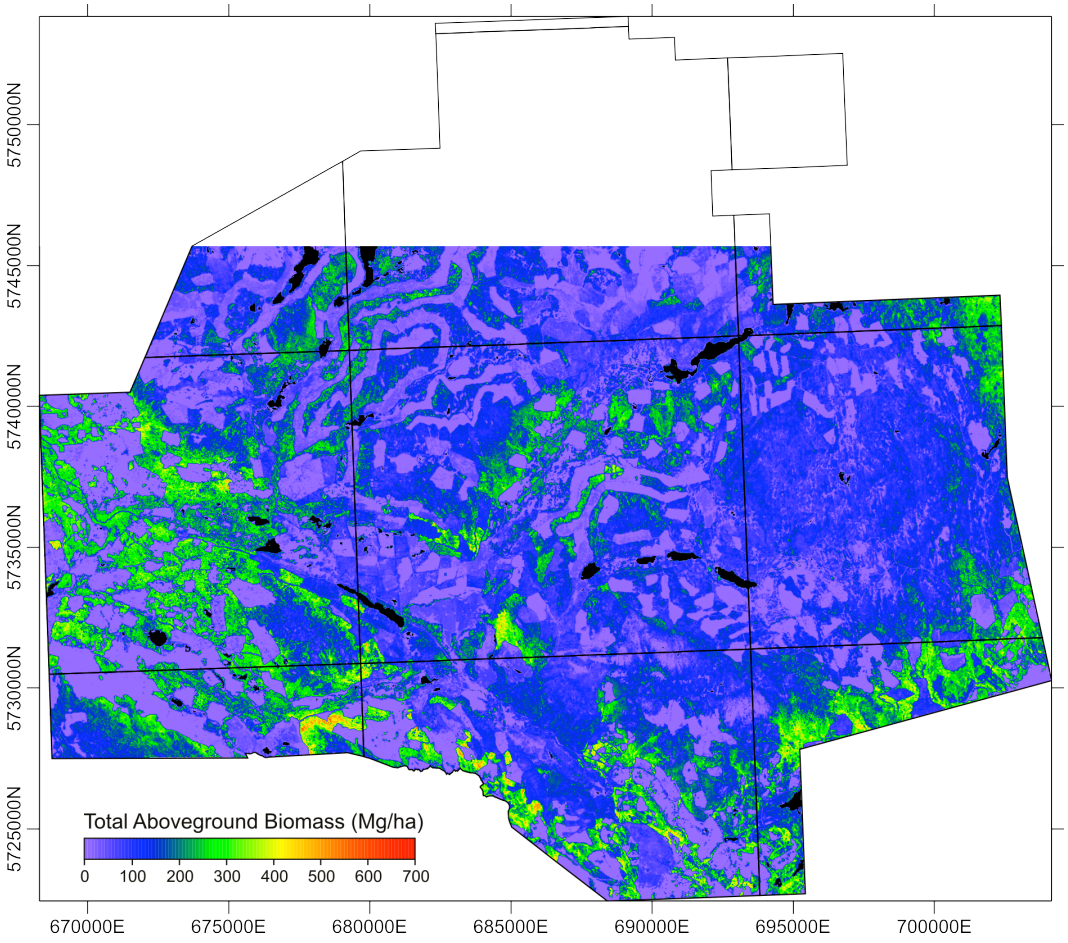


Figure 31. LiDAR predicted total aboveground biomass (dry weight) per hectare (TAGB). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 Mg/ha, 672.62 Mg/ha, 113.35 Mg/ha, and 98.89 Mg/ha, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

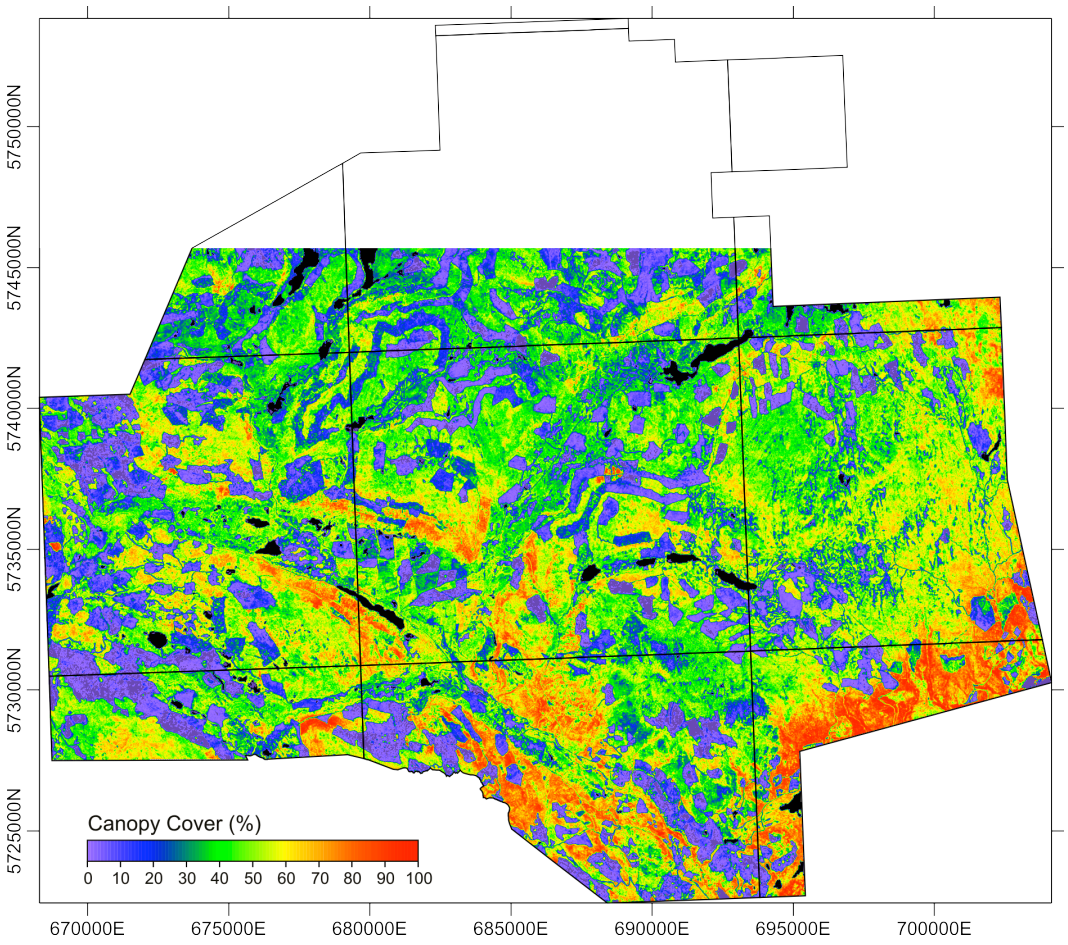


Figure 32. LiDAR predicted crown or canopy cover (reported in the VRI as crown closure). Minimum, maximum, mean, and standard deviation statistics for the surveyed portion of the TFL are 0 %, 100 %, 40.92 %, and 23.83 %, respectively. TFL and BCGS 1:20,000 map sheet boundaries are shown in black. Map sheet 92P.069 is located at the centre bottom of the TFL.

Average gross volume (VOL) was 201.42 m^3/ha ± 201.89 m^3/ha s.d. for all 212,873 cells; VOL was ≤ 316.37 m^3/ha in 75% of the cells, and never > 1210.05 m^3/ha. Crown cover (referred to as crown closure in the VRI) averaged 43% ± 27.14% s.d. across all 212,873 cells, and was ≤ 18.8% and ≤ 64.38% in 25% and 75% of the cells, respectively. Summary statistics differed for the TFL due to regional differences in logging intensity and the inclusion of other forest types (SBS and ESSF; Figure 13).

3.2.1. Spectral-based species classification

It is difficult to validate the classification based on extraction of individual spectra. There were few single species stands to sample for both calibration and validation. To overcome this we employed an evaluation on the individual library spectra using a *leave-one-out cross-validation* approach. This approach iteratively extracts a single spectrum from the calibration dataset, evaluates it against the remaining dataset, thus allowing us to evaluate the performance of the classifier. The results of this evaluation, displayed in Figure 33, suggest that the overall ability of the classifier to predict the observed species was 76%. A kappa adjustment (Cohen, 1960) was performed on the accuracies. This adjustment addresses the overestimation of accuracies due to chance agreements. It is defined as:

$$k = \frac{\mathrm{Pr}_o - \mathrm{Pr}_e}{1 - \mathrm{Pr}_e}$$

Where k is the Cohen's Kappa Coefficient; Pr_o is the observed agreement; Pr_e is the hypothetical probability of chance agreement. If the k value approaches 1 then the predicted and observed are the same. In our case k=0.65, which is classified as there is a substantial agreement between the two groups.

When addressing the individual species however the areas of confusion become apparent. There is considerable confusion between Douglas-fir

Predicted

Observed

	Fir	Spruce	Balsam	Cedar	Pine	Deciduous	Red Attack	Total
Fir	14	2	1	3	2			22
Spruce	2	5	3					10
Balsam	2	1	14		4			21
Cedar		2		18				20
Pine	1				18	1		20
Deciduous					1	6		7
Red Attack							5	5
Total	19	10	18	21	25	7	5	105

Figure 33. Cross tabulation between the species of the observed (library spectra) and the SAM predicted species using a leave one out cross tabulation. The overall accuracy is 0.76. The Cohen's Kappa Coefficient for the overall classification is 0.65.

(termed *fir* in Figure 33) and Englemann/white spruce, and to a lesser degree Balsam-fir. This confusion is expected given the similarities of the species, and low number of pure stands from which to sample. The pine distinguishes from the other conifer species as does the western redcedar. Deciduous vegetation, including predominantly poplar and willow were grouped, and separate from the other spectra. We also identify beetle infested individuals (pine) at the red-attack stage, which had very distinct and unique spectra. The classification accurately predicts the presence of most of the observed species. Spruce was the species that had the lowest success rate, while other species yielded success rates of 70% or greater. Again there was confusion amongst the three (fir, spruce, and Balsam) conifers. While the SAM classifier was used here, our approach affords us the flexibility to employ other methods to partition the feature space.

The species classified spectra were also compared to the species identified in the VRI field plots. The comparison the leading, secondary and tertiary species. The comparison was restricted to those individuals identified as belonging to the dominant/co-dominant canopy. The comparison of the two data sources is summarized in Figure 34. The results suggest that, for the most part, the hyperspectrally – derived species conform to the expected for the surveyed plots. There is discrepancy in the observed proportions, but for the most part there is a generally agreement in species between the two.

The discrepancies can be attributed to a number of factors. The first is that mentioned above where the overall classification accuracy between primarily the conifer classes is at a 76% level. This level of accuracy will translate into a certain confusion when extrapolated into a large area. Another area of discrepancy may be in the classification of the dominant / co-dominant canopy. Some of the trees identified through the use of the LiDAR-derived treetops may not correspond to

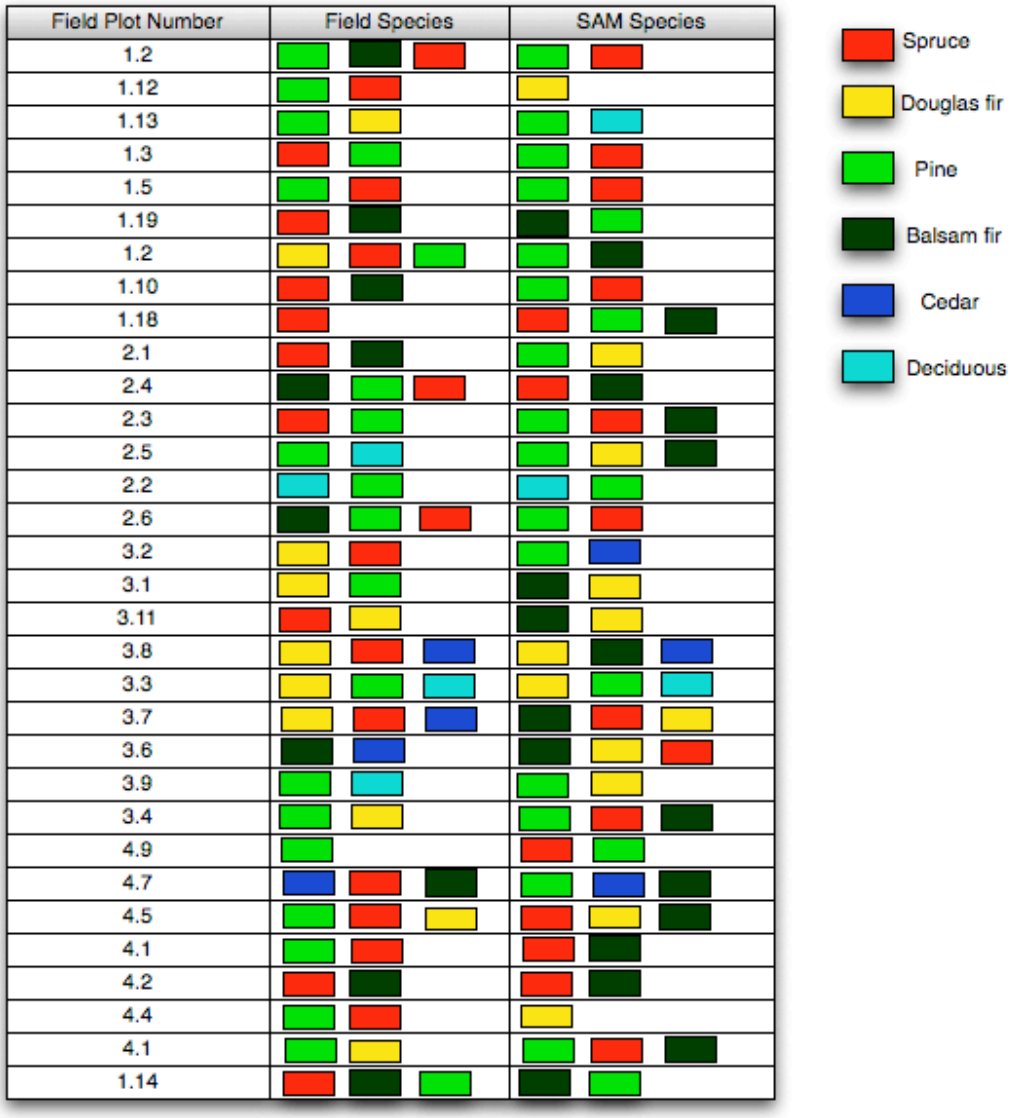


Figure 34. Comparison of the leading species as mapped in the field plots and those as determined by the SAM classifier. Leading species was by number of occurrences. Only the dominant / codominant canopy was considered for the field plots and all trees over 7 metres in height for the SAM classification.

the dominant / co-dominant canopy, and vice versa. The variation in species composition between the understory vegetation and that of the overstory may have some influence on the ultimate classification accuracy. [5]

A further expansion of the species mapping, and one where the power of this approach is highlighted is the development of a *leading species* layer. The evaluation was based on 20 metre grid cells. In this case gridded layer is created based on the example of the output is presented in Figure 35.

3.3. Comparison of VRI and remotely sensed inventories

3.3.1. Structure-based inventory results

Comparison of the VRI inventory with our remotely sensed data was made difficult by inherent differences in the scale of measurement, analytical methods, variable definition and data precision, and the timing of data collection. VRI inventory results, for example, were reported as stand-level estimates, while remotely sensed inventory data were estimated at individual-tree (species only) and 20 m grid-cell scales. The estimation of VRI attributes was largely subjective (dependent on the skill of the photo-interpreter and other less-subjective ancillary data sources) and reported at a much lower level of precision than the purely quantitative, high-precision estimates derived from the remote sensing approach. Last, the VRI photos were collected one year after the remotely sensed data, and therefore subject to forest cover changes related to logging, MPB outbreak, and other disturbances.

Our initial method of comparison was entirely qualitative. We overlaid the forest cover polygons generated by VRI photo-interpreters on top of the LiDAR imagery to see how well VRI stand boundaries compared to spatial patterns of

[5] A factor in all of the classification accuracies, and specific to the data collected for this project, was the influence of the atmosphere, specifically the smoke from fires in the region. The smoke became progressively heavier throughout the 4 days of the data collection. The ultimate results of the varying amounts of smoke/haze is that ground calibration at a single point in the study area will not result in sufficient data to correct for the changing optical conditions.

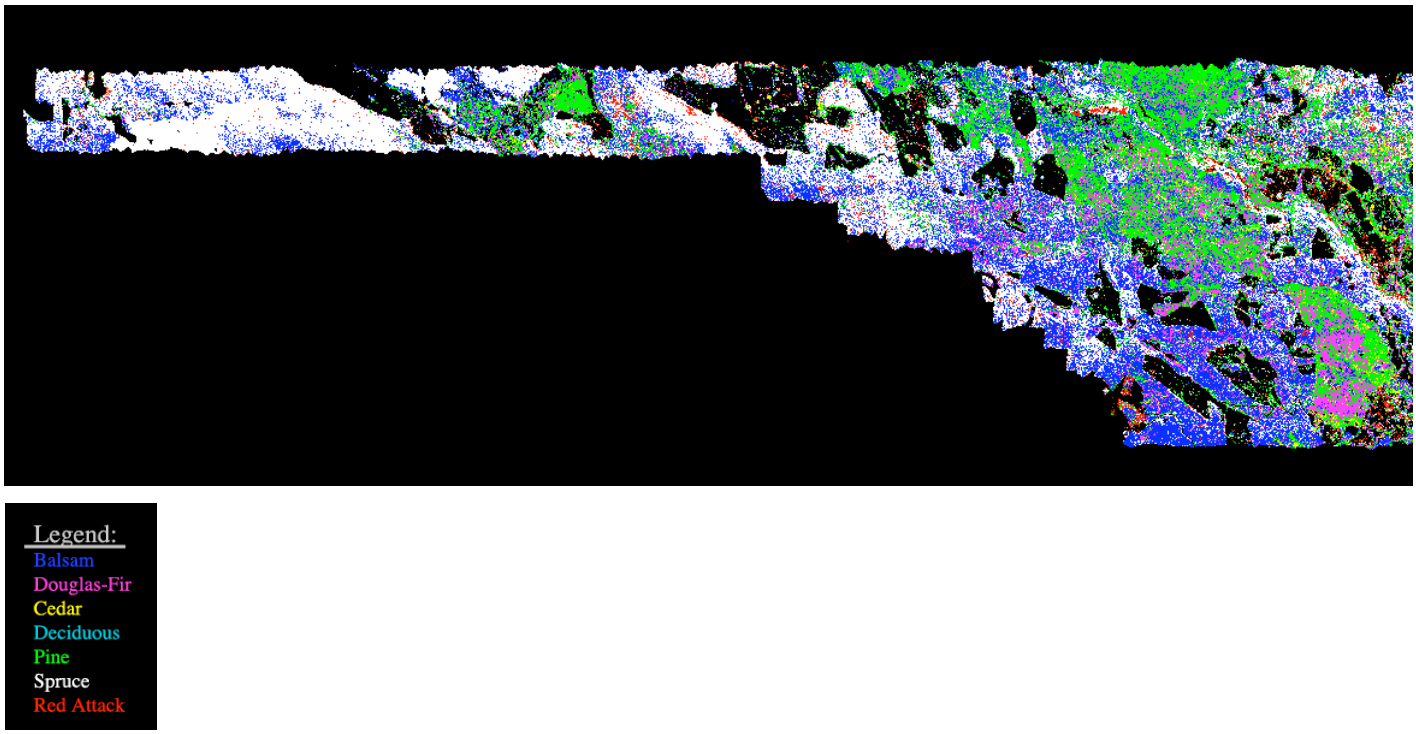


Figure 35. Example of the species classification applied to the first day of flying and summarized to a 20 me leading species.
\

LiDAR height and cover (Figures 20–25). The correspondence between these two different datasets was generally very good (i.e., placement of stand boundaries seemed logical and appropriate). However, it was immediately apparent that all forest cover polygons exhibited some measure of within-stand heterogeneity, and that some polygons were markedly less homogeneous in height and cover than others.

To compare the VRI and remotely sensed inventory data quantitatively, we computed stand-level estimates of LiDAR height, basal area, volume[6], and crown cover for each VRI forest cover polygon (n = 1057) using the average of all 20 m grid-cells contained within the polygon boundaries. We then regressed these upscaled (averaged), stand-level LiDAR estimates against equivalent VRI attributes (Figure 36). VRI and LiDAR estimates of stand basal area were the most highly correlated (R^2 = 0.68), followed by crown cover (R^2 = 0.67), height (R^2 = 0.60), and volume (R^2 = 0.58). We suspect that two main factors may have negatively impacted the magnitude of these correlations. First, stand-level LiDAR inventory estimates were influenced by the inclusion of mixed pixels (i.e., cells that straddled the boundary between two or more polygons, and did not accurately reflect the characteristics of anyone polygon). This would be a particular issue along high-contrast edges (e.g., border between clearcut and mature stand), or where polygons have a very high perimeter-to-area ratio (i.e., lots of mixed pixels). Second, no attempt was made to remove stands from the analysis that had been subject to major disturbance (i.e., logging and MPB) between August 20, 2006 (after the LiDAR survey) and August 1, 2007 (before the VRI photos).

Finally, we compared VRI and LiDAR inventory results at the map sheet scale by computing polygon-area-weighted mean values for each of the four attributes. We found again that the VRI and LiDAR methods produced similar results: mean height was 14.02 m (VRI) vs. 12.67 m (LiDAR); mean basal area

[6] We applied a non-linear equation (BA = 1.3549*VOL^0.5887; r = 0.96) derived from our 35 ground-reference plots to impute stand-level volume estimates for the VRI. This equation is unlikely to be scale-invariant, so we expect that this procedure is not implemented without some element of error.

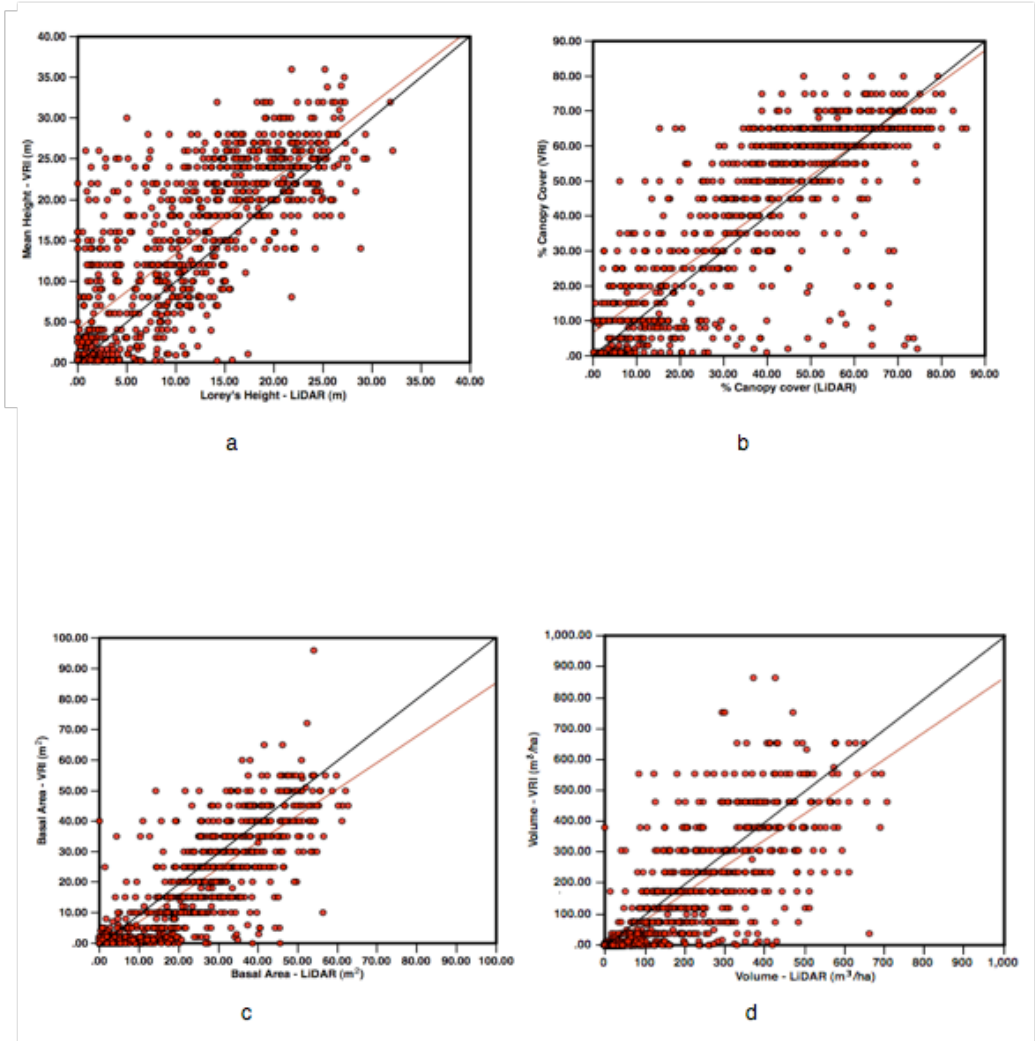


Figure 36. Comparison of the distribution of attributes by polygon for common VRI and LiDAR-derived attributes

.

was 21.43 m^2/ha (VRI) and 24.43 m^2/ha (LiDAR); mean volume was 188.20 m^3/ha (VRI) and 193.27 m^3/ha (LiDAR), and mean crown cover was 45.9% (VRI) and 42.1% (LiDAR). Unweighted mean values calculated directly from the LiDAR grid-cell data (Table 5) were slightly higher than the LiDAR polygon-area-weighted means reported above: 13.04 m (height), 25.26 m^2/ha (basal area), 201.43 m^3/ha (volume), and 43% (crown cover). These differences, although seemingly small, highlight the need to account for the continuously varying nature of inventory data at spatial scales finer than forest cover polygons.

In addition to the LiDAR -VRI comparison, we also compared the species classification from the VRI and the hyperspectral data. For this we summarized the leading species from the single tree-based classification for each of the VRI polygons. The leading species in this case was based on the numeric count for each polygon. The comparison between the two sources is presented in Figure 37. Consistent with the previous accuracy assessment and comparison with the field plots there is confusion between the first leading species. When comparing all of the leading species for a number of polygons however (Figure 38) we see that in most cases the species are represented, but in differing proportions. A complete listing of the VRI-hyperspectral species comparison is included in the DVD accompanying the report. One issue that does come up with the comparison of the two classifications is that the hyperspectral-based product classifies the leading species of low density stand such as those found in wetlands. The VRI classification for these typically integrated some interpretation of the landform into its final assessment. At this point we have not had to opportunity to integrate this aspect into the classification and evaluation, however this is an area of research that is currently underway.

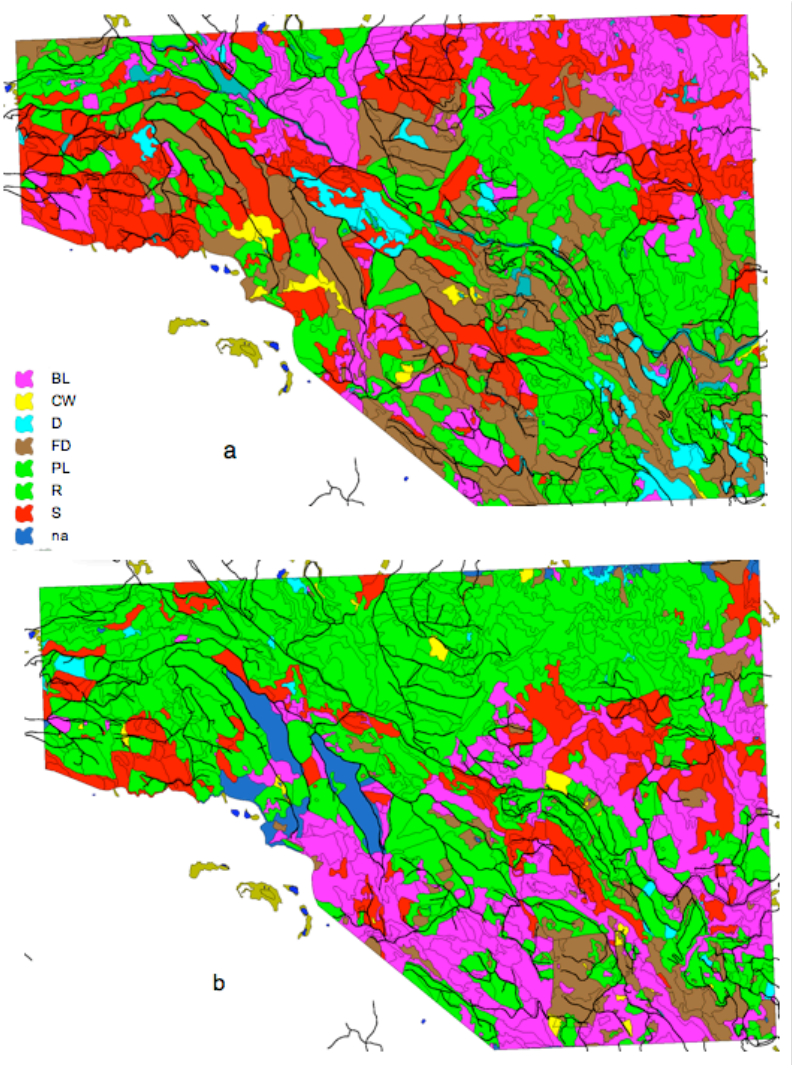


Figure 37. A Comparison of the leading specie per VRI polygon for the test map sheet. The top map sheet (a) is the VRI species designation while the lower map (b) is the one generated from the hyperspectral data.

ID	Species 1 (VRI)	Perc 1	Species 2 (VRI)	Perc 2	Species 3 (VRI)	Perc 3	Species 1 (SAM)	Perc 1	Species 2 (SAM)	Perc 2	Species 3 (SAM)	Perc 3
4	FD	40	PL	30	CW	10	PL	62	S	20	FD	11
5	FD	50	PL	45	SX	5	FD	34	PL	29	S	26
6	FD	40	PL	40	AT	20	PL	32	FD	27	CW	22
9	FD	50	PL	40	AT	10	PL	31	S	27	FD	26
10	PL	48	LW	22	AT	10	PL	54	FD	21	CW	13
11	PL	55	FD	30	AT	10	PL	61	S	16	FD	12
12	EP	42	AT	16	FD	16	PL	36	S	28	FD	15
14	FD	40	PL	40	SX	20	S	49	FD	18	PL	17
16	SX	40	BL	40	PL	20	S	46	FD	21	PL	16
17	FD	80	PL	10	CW	10	S	48	FD	22	PL	14
18	SX	60	BL	30	CW	10	S	57	PL	19	R	12
21	SX	50	BL	25	PL	15	PL	36	S	30	FD	26
22	SX	80	BL	20	0	0	FD	39	S	29	PL	18
23	AT	60	PL	20	FD	20	PL	56	FD	15	S	13
25	FD	51	LW	23	BL	10	PL	35	FD	32	S	23
26	SX	50	BL	50	0	0	S	44	FD	16	BL	16
28	FD	40	PL	30	SX	10	PL	57	S	25	FD	11
29	PL	60	SX	30	FD	10	PL	46	FD	28	S	10
30	PL	70	SX	20	FD	10	PL	26	S	25	R	16
32	PL	75	FD	25	0	0	PL	36	S	26	R	20
33	FD	60	CW	20	PL	20	PL	33	FD	31	S	16
34	FD	85	PL	10	SX	5	S	52	FD	18	PL	14
35	PL	50	FD	40	SX	10	PL	29	S	26	FD	24
36	BL	100	0	0	0	0	PL	35	S	35	R	13
37	SX	60	BL	30	PL	10	R	40	S	28	PL	24
38	PL	50	FD	35	BL	10	S	44	BL	17	FD	15
39	PL	100	0	0	0	0	PL	45	S	33	R	10
41	PL	45	AT	25	FD	20	BL	36	D	33	FD	16
42	SX	50	PL	30	BL	20	S	40	BL	18	FD	18
43	PL	51	FD	24	BL	17	PL	49	FD	21	CW	14
46	AT	60	PL	20	FD	20	PL	31	FD	29	S	16

Figure 38. Details of the leading species breakdown on a polygon basis for the VRI hyperspectral (SAM) comparison.

4. Conclusions and Future Directions

LiDAR produces extremely accurate 3-D geometric reconstructions of forest canopies. The task for this project was to convert these already accurate measures of canopy height into stand attributes that have greater meaning to foresters. The current project attempted to provide an environment where these data were processed to extract attributes common with the VRI.

LiDAR canopy heights have tremendous merit in their raw form – they allow us to perceive and quantify the continuously varying nature of forest canopy height and cover in all four spatiotemporal dimensions. We need, however, to start addressing how LiDAR/hyperspectral-derived canopy metrics can be used more directly for forest management, rather than trying to alter them to fit the existing paradigm of stand-level management. The project reported on here is one that did not directly address resource management issues directly. However, given the features that were extracted, the implications for management issues are apparent.

4.1. Summary of key findings

The project has demonstrated the utility of integrated hyperspectral and LiDAR data for forest inventory, and forest biology, feature extraction. A number of key conclusions have been attained. These include:

- The collection and integration of airborne data for a large cohesive area has been shown to be feasible;
- The data volumes, while high, are manageable with desktop computer systems;
- There is a consistency between the LiDAR/hyperspectral - derived attributes and those collected through traditional aerial photographic - based surveys;
- The advantage of the LiDAR/hyperspectral - derived surveys over the more traditional ones is that the former retains fine scale spatial variability, while the latter generalizes and retains central tendencies at the stand level;
- The hyperspectral data, which are complementary to LiDAR can be used to derive species, with accuracies at the same level of accuracy as

traditional surveys. The exploitation of hyperspectral data is, however, still nascent, with further research required; and

- The data acquisition costs of the LiDAR are, at present, higher than comparable photographic missions. The cost savings are realized in the processing and feature extraction. There are also benefits for multiple stakeholders to share costs. Information extraction costs for topographic mapping, for example, will not be incrementally higher once the initial bare earth model has been extracted.

The quality and usefulness of the information derived have been demonstrated to be at least on par with the more traditional survey techniques. We have restricted our analysis to the canopy portion of the overall LiDAR returns. The bare earth model, which has not been included in this analysis offers additional uses. Some of these uses are readily apparent, especially those pertaining to the support of forestry operations, such as cut block layout and road design, terrain analysis/slope stability, hydrological modelling etc. Other potential uses can be as a substitute for the more traditional DEM generation for topographic mapping. Modelling other plant-relevant variables related to site energy balance (i.e., radiation, moisture, temperature, etc.) for predictive vegetation modelling (species distribution), site index, wood quality, ecosystem modelling. The DEM provides invaluable ecological context (i.e., elevation, slope, aspect, etc.) to measurements of radiance spectra or vegetation height data. For example, two or more species that are inseparable in the spectral or structural domain may differ by topographic position (habitat).

4.2. Future directions

This project has provided an insight to the use of these remotely sensed data for forestry related surveys. While much of the energy was spent on developing a processing framework for the efficient extraction of attributes, some relevant attributes were successfully derived. Considerable work still needs to be carried out to fully exploit these data.

4.2.1. Hyperspectral Data:

Sensor Noise: The work on hyperspectral data is preliminary. These data have

not previously been applied at the fine resolutions (both spatially and spectrally) to this type of survey. Effort still needs to be expended out that will characterize the noise of the AISA sensor. The success of further analysis of this sensor will depend on our ability to increase the S/N. This will have the effect on classifiers such as the Spectral Angle Mapper used in this project that uses the shape of the spectrum to derive the species identification. In many instances, as found in this project, the signal strength of the reflected energy will vary depending on a number of factors including time of day, solar orientation, and terrain related shadowing. With a decrease in the overall intensity of the spectrum the effects of the noise becomes more apparent and the benefits of using a classifier such as the one that we used here decreases. In these cases the noise increasingly defines the overall spectral curve so that misclassification will occur.

Alternative approaches to classification: While the supervised classifier used on this project was one that is quite effective in most situations, other strategies to partition an n-dimensional feature space may prove to be more effective. The inclusion of LiDAR-derived variables (both related to the stand structure and terrain) through nonparametric modelling techniques is an avenue that has not been attempted here , but has been reported on in the literature. This increase in the dimensionality has been demonstrated with lower quality data sets in the past (see for example Niemann, 1991), but has not been extended to LiDAR-derived sources.

Foliar health: Work on issues related to foliar health is ongoing parallel to this project. One of the areas of interest is in characterizing pigment (including Chlorophyll a and b) levels as well as the modelling of foliar Nitrogen. This type of analysis, in combination of some of the LiDAR - based analysis, will allow us to more adequately address issues related to overall nutritional status and isolate early stages of the effects of insect infestation and disease.

4.2.2. LiDAR

Feature extraction from the LiDAR data continues with work progressing on

the investigation of a variety of aspects of forest biology and forest information extraction.

Growth environment: The use of the very highly detailed elevation made possible from these data can aid us in deriving an understanding of the distribution of antecedent landscape moisture. We are presently integrating these data with potential evapotranspiration models to better characterize growth potential. The output to this research would be extremely beneficial in ecosystem modelling (PEM).

Stand Delineation The current models of stand delineation using aerial photographic interpretation do not necessarily key in on any single attribute. The variability of the LiDAR-derived attributes within VRI polygons, as demonstrated by the work reported on here, indicates that the definition of individual stands is somewhat fluid. The use of LiDAR-derived attributes with their continuous variation would help to standardize the boundary delineation.

Wood quality: A project addressing the quality of the wood (in terms of elasticity, etc) through a linkage with the acoustical properties of the standing timber and LiDAR - and hyperspectral - derived properties is being initiated.

Fire fuel assessment: The vertical and horizontal structure of a stand is critical in understanding the potential fire behaviour of a stand. We are using the biometrics developed for this project to relate to the vertical structure of the stand. In addition, methods to automatically separate the dominant-codominant structure from the understory fuels are being explored. Once this is achieved then we can address the nature of the understory fuel characteristics.

4.2.3. Sensor Development

A new suite of sensors is being assembled (MAP 2) which is intended to be helicopter-based. This platform integrates a very high frequency LiDAR (260 KHz) with a VNIR spectrometer. In addition a capacity of a thermal sensor is also being included. The LiDAR system will have the capacity to produce a dataset with up to 12 points per sq m. digitizing 4 points per return. The dataset while having a very high data size will allow us to better characterize the vertical and

horizontal canopy structure. In addition, early tests indicate that we can map coarse woody debris. The VNIR spectrometer will have the capacity to collect up to 490 channels for the range between 400 and 1000 nm at a spatial resolution of 50 cm or finer. This will give us the ability to map individual crowns within a stand with multiple pixels. In addition the spectral resolution (potentially of ca 1 nm) will give us the ability to better define the absorption features in VNIR range. This could be useful, for example, to better define pigments such as carotinoids, and Chl a and b.

5. REFERENCES

Andersen, H-E., S.E. Reutebuch, and R.J. McGaughey, 2006. A rigorous assessment of tree height measurements obtained using airborne LiDAR and conventional methods. Canadian Journal of Remote Sensing 32: 255-366.

Avery, T.E. and H.E. Burkhart, 2002. Forest measurements, fifth edition. McGraw-Hill, New York, NY.

BCMOF, 2006. Tree Farm License 18, Canadian Forest Products Ltd., Rationale for Allowable Annual Cut (AAC) Determination, March 9, 2006.

Caliński, R.B. and J. Harabasz,, 1974. A dendrite method for cluster analysis. Communications in Statistics 2: 1–27.

Cohen, J. , 1960. A coefficient of agreement for nominal scales, Educational and Psychological Measurement 76: 613-619.

Coops, N. and D. Culvenor (2000). Utilizing Local Variance of Simulated High Spatial Resolution Imagery To Predict Spatial Pattern of Forest Stands. Remote Sensing of Environment 71(3): 1149.

Curtis, R.O. and D.D. Marshall, 2000. Why quadratic mean diameter? Western Journal of Applied Forestry 15: 137–139.

De'ath, G. and K.E. Fabricius, 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81: 3178–3192.

Fotheringham, A.S., C. Brunsdon and M.Charlton, 2002. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, New York, NY.

Franklin, J., 1995. Predictive vegetation mapping: geographical modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19: 474–499.

Frazer, G.W., 2007. Fine-scale, multidimensional spatial patterns of forest canopy structure derived from remotely sensed and simulated datasets. PhD Dissertation, Department of Geography, University of Victoria. 251 p.

Frazer, G.W., M.A.Wulder, and K.O. Niemann, 2005. Simulation and quantification of the fine-scale spatial pattern and heterogeneity of forest canopy structure: a lacunarity-based method designed for the analysis of continuous canopy heights. Forest Ecology and Management 214: 65–90.

Gaveau, D. L. A. and R. A. Hill (2003). Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Canadian Journal of Remote Sensing 29(5): 650-657.

Gobakken, T. and E. Næsset, 2005. Weibull and percentile models for LiDAR-based estimation of basal area distribution. Scandinavian Journal of Forest Research 20: 490–502.

Gobakken, T. and E. Næsset, 2008. Assessing effects of laser point density, ground sampling intensity, and field plot size on biophysical stand properties derived from airborne laser scanner. Canadian Journal of Forest Research 38: 1095-1109.

Guttman, N.B., 1994. On the sensitivity of sample L moments to sample size. Journal of Climate 7: 1026-1029.

Hartigan, J.A., 1985. Statistical theory of clustering. Journal of Classification 2: 63–76.

Hay,G. and K. O. Niemann (1994). "Visualizing 3-D texture: a three-dimensional approach to model forest texture." Canadian J. of Remote Sensing 20: 89-101.

Hopkinson, C. and L. Chasmer, 2008. Testing models of fractional cover across multiple forest ecozones. Remote Sensing of Environment 113: 275-288.

Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society B 52: 105-124.

Hosking, J.R.M., 1992. Moments or L-moments? An example comparing two measures of distributional shape. The American Statistician 46: 186-189.

Hudak, A.T., N.L. Crookston, J.S. Evans, D.E. Hall, and M.J. Falkowski, 2008. Nearest neighbour imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment 112: 2232–2245.

Hyndman, R.J. and F. Fan, 1996. Sample quantiles in statistical packages. The American Statistician 50: 361-365.

Jenkins, J.C., D.C. Chojnacky, L.S. Heath, and R.A. Birdsey, 2003. National-scale biomass estimators for Unites States tree species. Forest Science 49: 12–35.

Kaufmann, L. and P.J. Rousseeuw, 2005. Finding groups in data: an introduction to cluster analysis. John Wiley and Sons Inc., Hoboken, NJ.

Ketcheson, M.V., T.F. Braumandl, D. Meidinger, G. Utzig, D.A. Demarchi, and B.M. Wikeem, 1991. Interior Cedar–Hemlock Zone. In: Meidinger, D., Pojar, J. (Eds), Ecosystems of British Columbia. BC Ministry of Forests, Research Branch, Special Report Series 6, February 1991, pp. 168–181.

Kimsey, Jr., M.J., J. Moore, and P. McDaniel, 2008. A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. Forest Science 54: 356–366.

Krzanowski, W.J. and Y.T. Lai, 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering. Biometrics 44: 23–34.

Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K.B. Heidebretch, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz, (1993). "The spectral image processing system (SIPS)--interactive visualization and analysis of imaging spectrometer data." Remote Sensing of Environment 44(2-3): 145-163.

Leckie, D.G., S. Tinis, T. Nelson, C. Burnett, F. Gougeon, E. Cloney, and D. Paradine, 2005. Issues in species classification of trees in old growth conifer stands. Canadian Journal of Remote Sensing 31: 175-190.

Lefsky, M.A., A.T. Hudak, W.B. Cohen, and S.A. Acker, 2005. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. Remote Sensing of Environment 95: 517–531.

.

Legendre, P., K.E. Ellingsen, E. Bjørnbom, and P. Casgrain, 2002. Acoustic seabed classification: improved statistical method. Canadian Journal of Aquatic Science 59: 1085–1089.

Legendre, P., and L. Legendre, 1998. Numerical ecology, second English edition. Developments in Environmental Modelling 20, Elsevier Science B.V., Amsterdam.

Lim, K.S. and P.M. Treitz, 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research 19: 558-570.

Lloyd, C.D. and P.M. Atkinson, 2002. Deriving DSMs from LiDAR data with kriging. International Journal of Remote Sensing 23: 2519–2524.

Lookingbill, T.R. and D.L. Urban, 2005. Gradient analysis, the next generation: towards more plant-relevant explanatory variables. Canadian Journal of Forest Research 35: 1744–1753.

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 28: 1016-1031.

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: 1367-1371.

Montgomery, D.C., E.A. Peck, and G.G. Vining, 2006. Introduction to linear regression analysis. John Wiley & Sons, Inc., Hoboken, NJ.

Nelson, R., 1997. Modeling forest canopy heights: the effects of crown shape. Remote Sensing of Environment 60: 327–334.

Niemann, K.O. (1991) Landscape Drainage Modelling to Enhance LANDSAT Classification Accuracies GeoCarto International 6: 13-30.

Niemann, K. O., S. Adams, and G. Hay (1999). Automated tree crown identification using digital orthophotomosaics. International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry, Victoria, B.C., Natural Resources Canada, Canadian Forest Service.

Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne LiDAR system. Remote Sensing of Environment 56: 1-7.

Næsset, E., 2002. Predicting forest stand characteristics with airborne laser scanner using a two-stage procedure and field data. Remote Sensing of Environment 80: 88-99.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research 19: 164–179.

Næsset, E., O.M. Bollandsås, and T. Gobakken, 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment 94: 541-553.

Næsset, E., T. Gobakken, J. Holmgren, H. Hyyppä, J. Hyyppä, M. Maltamo, M. Nilsson, H. Olsson, Å. Persson, and U. Söderman, 2004. Laser scanning of forest resources: the Nordic experience. Scandinavian Journal of Forest Research 19: 482–499.

Parker, G.G., 1995. Structure and microclimate of forest canopies. In: M.D. Lowman and N.N. Nadkarni (Eds.), Forest Canopies. Academic Press, San Diego, CA, pp. 73–106.

Peres-Neto, P.R., D.A. Jackson, and K.M. Somers, 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis 49: 974–997.

Roberts, D.A., S.L. Ustin, S. Ogunjemiyo, J. Greenberg, S.Z. Dobrowski, J. Chen, and T.M. Hinckley, 2004. Spectral and structural measures of Northwest forest vegetation at leaf to landscape scales. Ecosystems 7: 545-562.

Reutebuch, S.E., H.-E. Andersen, and R.J. McGaughey, 2003. Light detection and ranging (LiDAR): an emerging tool for multiple resource forest inventory. Journal of Forestry 103: 286–292.

Rönnholm, P., J. Hyyppä, H. Hyyppä, H. Haggrén, X. Yu, and H. Kaartinen, 2004. Calibration of laser-derived tree height estimates by means of photogrammetric techniques. Scandinavian Journal of Forest Research 19: 524-528.

Sharma, M., R. L. Amateis, and H.E. Burkhart, 2002. Top height definition and its effect on site index determination in thinned and unthinned loblolly pine plantations. Forest Ecology and Management 168: 163–175.

Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64: 209–210.

Swanson, F.J., T.K. Kratz, N. Caine, and R.G. Woodmansee, 1988. Landform effects on ecosystem patterns and processes. BioScience 38: 92–98.

Tsai, F. and W. Philpot, 1998 Derivative analysis of hyperspectral data, Remote Sensing of Environment, vol. 66, pp. 41-51.

Vogel, R.M. and N.M . Fennessey, 1993. L-moment diagrams should replace product moment diagrams. Water Resources Research 29: 1745-1752.

Vosselman, G., 2000. Slope based filtering of laser altimetry data. International Archives of Photogrammetry and Remote Sensing, Vol. 33 (Part B3/2), Amsterdam, the Netherlands, p. 935–942.

Wulder, M., K. O. Niemann, and D.G. Goodenough (2000). Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. Remote Sensing of Environment 73: 103-114.

Zenner, E.K., 2005. Investigating scale-dependent stand heterogeneity with structure-area-curves. Forest Ecology and Management 209: 87–100.