Development of a forest inventory using 2018 Single Photon LiDAR and assessing decadal forest change

Nicholas Coops

Grant McCartney

Martin Queinnec

March 2020

University of British Columbia

Table of Contents

1. Introduction

1.1. Focus Site

The Province of Ontario launched a province-wide ALS mapping project using the ALS SPL Leica SPL100. The first acquisitions carried out during summer 2018 covered the Romeo Malette Forest (RMF) (630,000 ha), the Petawawa Research Forest (10,000 ha) and part of Hearst Forest (1.23 million ha). Additional acquisitions are occurring and scheduled in 2020 and beyond, to cover the remaining managed forested areas of the Province. This research focuses on the data acquired at Romeo Malette Forest (RMF) displayed in Figure 1.

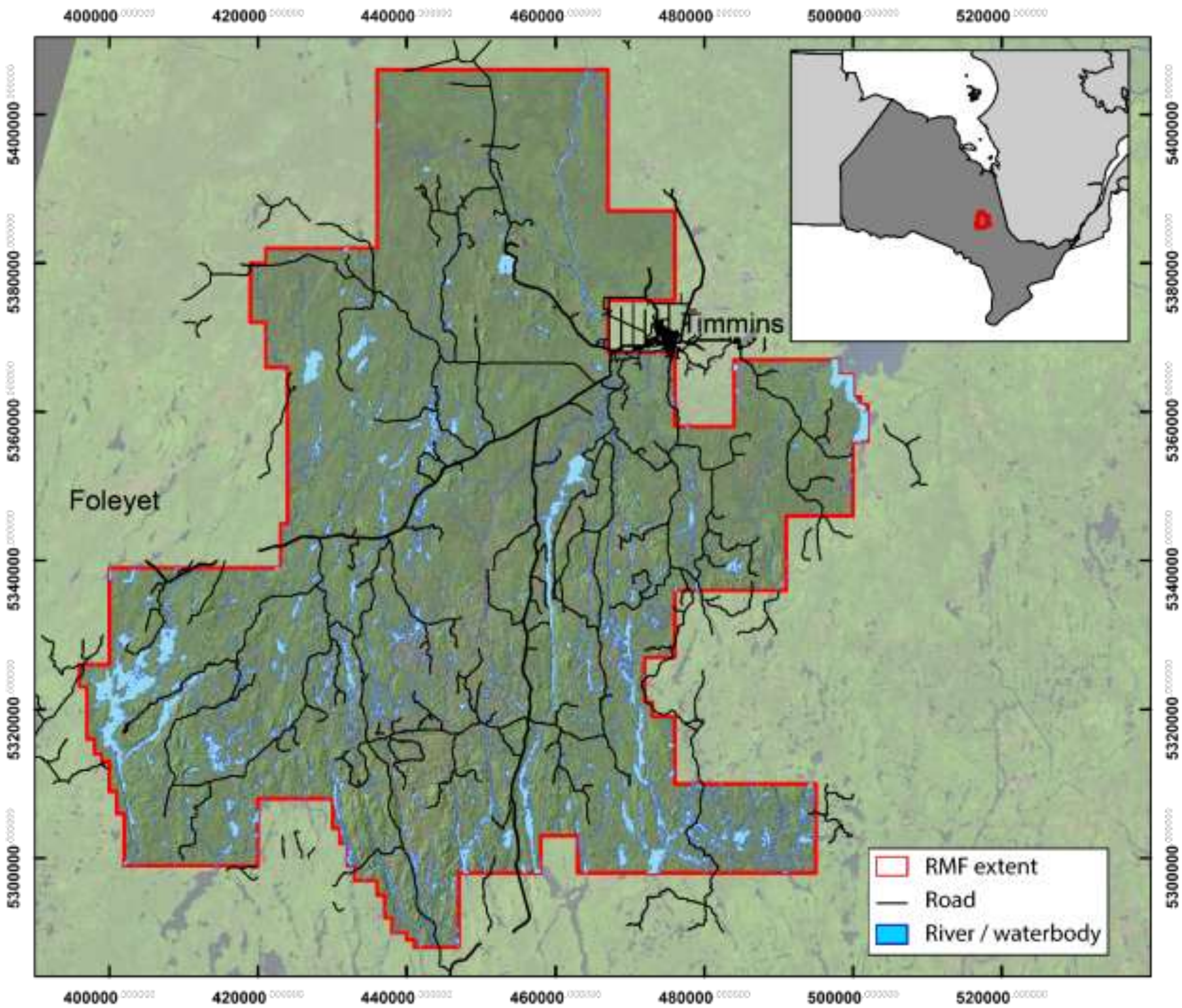

Figure 1. Location and extent of the Romeo Malette Forest. Landsat 8 OLI true color imagery (path 20 ; row 26 and path 20; row 27) acquired on June 25th 2018 is used as background. Coordinate system: NAD83 / UTM Zone 17N.

The RMF is a managed boreal forest of approximately 630,000 ha located in Northern Ontario. It lies within the Boreal Shield ecozone (Ecological Stratification Working Group, 1996) characterized by a continental climate with long cold winters and short warm summers. The terrain consists in flat to moderately rolling topography (305 – 380 m a.s.l) and poorly drained soils resulting in extensive wetlands, watercourses and lakes. The forest is mostly composed of black spruce (*Picea mariana*), jack pine (*Pinus banksiana*), poplar (*Populus* spp.), white birch (*Betula papyrifera*), white spruce (*Picea glauca*), cedar (*Thuja spp.*), larch (*Larix spp.*) and balsam fir (*Abies balsamea*).

1.2 ALS acquisition

The first Enhanced Forest Inventory (EFI) in RMF was carried out using lidar data collected during summer seasons of 2004 and 2005 with an upgraded version of the near-infrared (λ = 1064 nm) Leica ALS40 instrument at a nominal flight altitude of 2740 m above ground. The acquisition was performed with a field of view of 20°, a scan rate of 30 Hz and a maximum pulse repetition frequency of 32 300 Hz resulting in an average point density of 0.46 points m^{-2} (Woods et al., 2011).

A second lidar acquisition over RMF was carried out during June and July 2018 with the Leica SPL100 single-photon sensitive instrument, thereafter referred to as single-photon lidar (SPL). The SPL operates in the green region of the electromagnetic spectrum (λ = 532 nm). Each laser beam is split into a 10 x 10 array of beamlets using a diffractive optical element to enhance the sampling density. The SPL was operated at a nominal altitude of 4000 m above ground, with a field of view of 30° and a pulse repetition frequency of 60 kHz (effective pulse repetition frequency of 60 MHz considering the 10 x 10 array of beamlets). The acquisition parameters resulted in an average point density of 22 points m^{-2}.

2. Structural based sampling of EFI calibration plots

2.1 Structural guided sampling using principal component analysis

When generating forest inventory attributes using an area-based-approach (ABA), the ground plots data used to calibrate ABA regression models need to be representative as much as possible of the full range of forest structure variability within the study area. If this is not the case, regression models might perform poorly in underrepresented forest types (White et al., 2013). LiDAR metrics such as height percentiles, cover or height variability can be used to design a sampling network driven by forest structure.

Principal Component Analysis (PCA) is a method used to summarize the variability of a large number of highly correlated LiDAR structural metrics into a smaller number of uncorrelated variables. The feature space created by the generated principal components can then be stratified into classes that will represent specific types of forest structural conditions. Random sampling can then be performed within each of these classes to ensure a representative characterization of all forest structures occurring across the study area.

2.2 Establishment of field inventory plots at RMF

A network of 182 plots was already established in RMF. The objective of the structural guided sampling was to check if the existing plot network was covering the entire range of structural variability and if not, selecting new plots in underrepresented forest types. First, a set of 20 metrics, listed in Table 1, characterizing vegetation height, cover and vertical distribution were calculated from the SPL on a 20 m x 20 m grid (see Figure 2) to match plots area (r = 11.28 m, area of 400 m^2) as recommended in the ABA approach (White et al., 2013). The software LAStools (Isenburg, 2014) was used to calculate SPL metrics although we could also recommend the lidR package (Roussel & Auty, 2019) implemented in R (R Core Team, 2018). Forest resource inventory polygons obtained from the MNRF were used to constrain the analysis to productive forest types by selecting only cells that intersects with polygons having a POLYTYPE attribute equal to FOR. The rasterPCA function from the RStoolbox package implemented in R was then used to perform a PCA and summarize the 20 metrics into 2 principal components PC1 and PC2 containing

76 % and 11 % of the metrics variance respectively (total variance explained of 87 %). Existing plot-level structural metrics were also calculated from the SPL point cloud clipped to plots location and PCA values determined. A flow diagram of the developed approach is shown in Figure 3.

Table 1. Set of 20 structural metrics calculated with LAStools

Structural metric	Description
avg / qav	Average / Average square height of returns
p05, p10, p20, p30, … , p90, p95, p99	Height percentiles (cutoff height of 1.3 m)
cov_1.3, cov_2, cov_5, cov_10, cov_15	Canopy cover above 1.3 m, 2 m, 5 m, 10 m and 15 m (% first returns above specified height)
std	Standard deviation of returns height

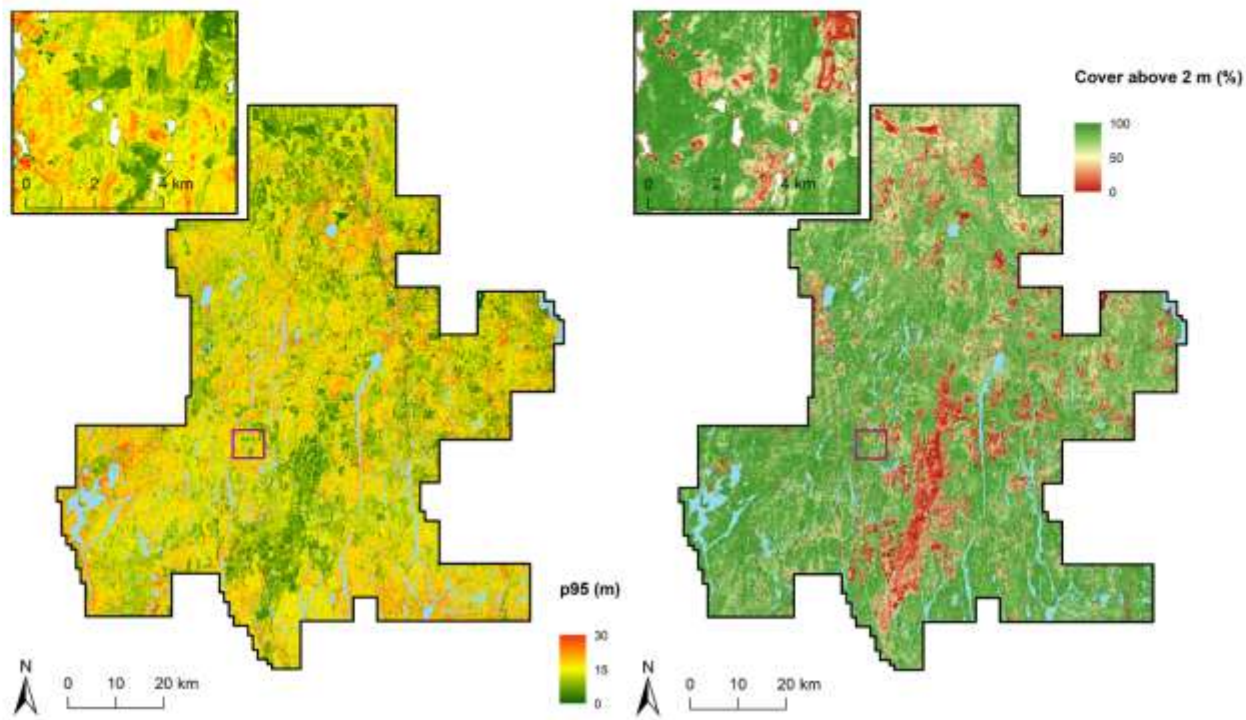

Figure 2. 95th height percentile (left) and canopy cover above 2 m (right) calculated on a 20 m x 20 m grid in Romeo Malette Forest

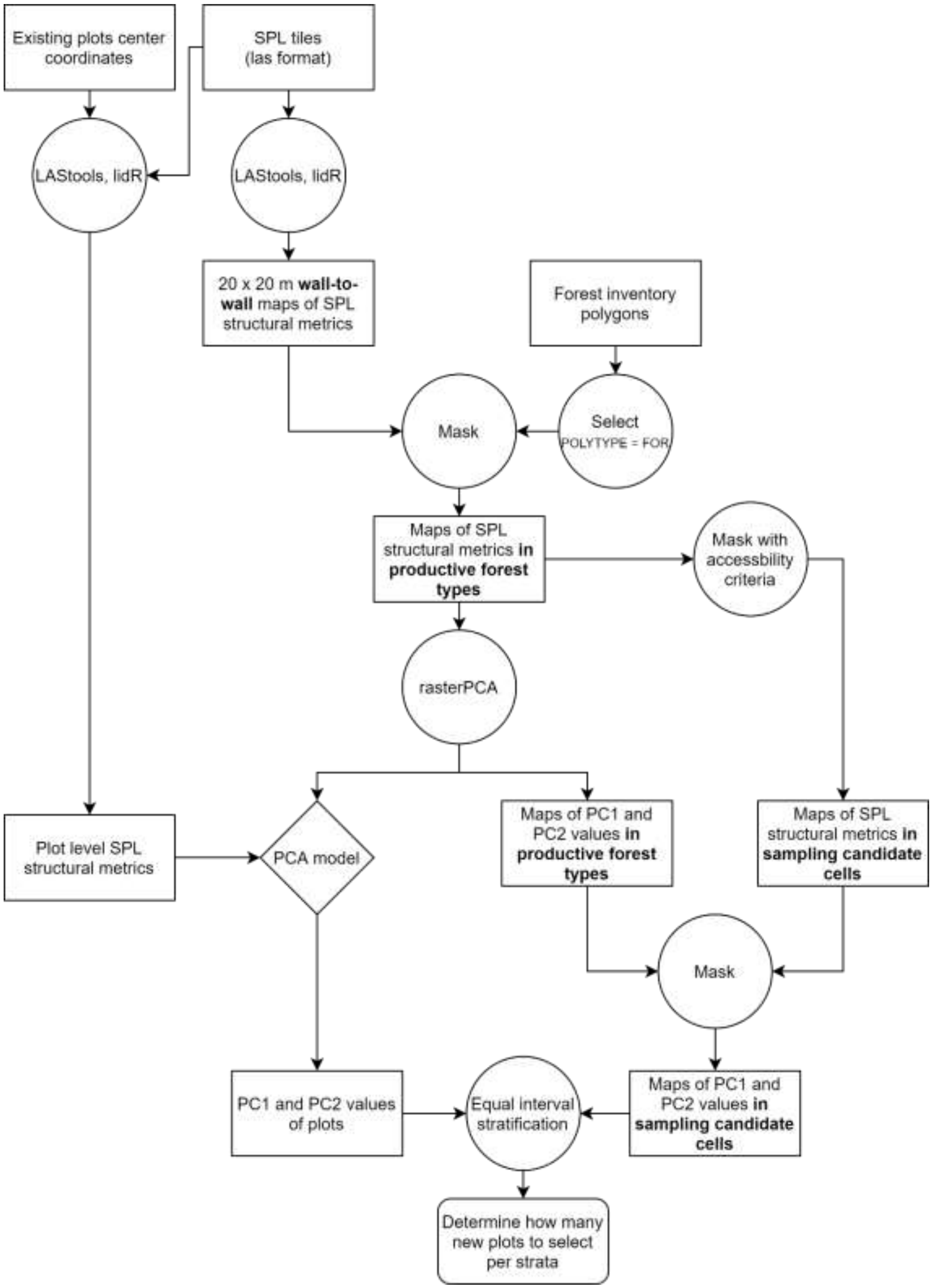

Figure 3. Flowchart presenting the PCA stratification method

As seen in Figure 4, the existing plot network didn't cover the entire range of variability occurring across RMF and new plots needed to be selected. The cells suitable for a new plot establishment, thereafter referred to as candidate cells, were determined based from accessibility criteria such as distance to roads and road type. Specifically, only cells located between 30 m and 200 m of highways, municipal/concession roads, primary roads, branch roads or clay/mineral surface roads accessible year-round were considered.

The feature space formed by PC1 and PC2 values of candidate cells was stratified using 10 and 5 equal intervals for PC1 and PC2 respectively (see summary of stratification scheme in Figure 3). This results in a matrix of 5 x 10 strata covering the candidate cells range of structural metrics. By comparing the distribution of PC1 and PC2 values of candidate cells and existing plots (Figure 1), underrepresented strata could be identified and the number of new plots to establish within each strata could be determined with the overall objective of maximizing the number of existing plots to remeasure and keeping the total number of plots around 250. The 20 Integrated Monitoring Framework (IMF) plots (long-term, multi-purpose monitoring plots) established at RMF were forced to be included in the final plot network.

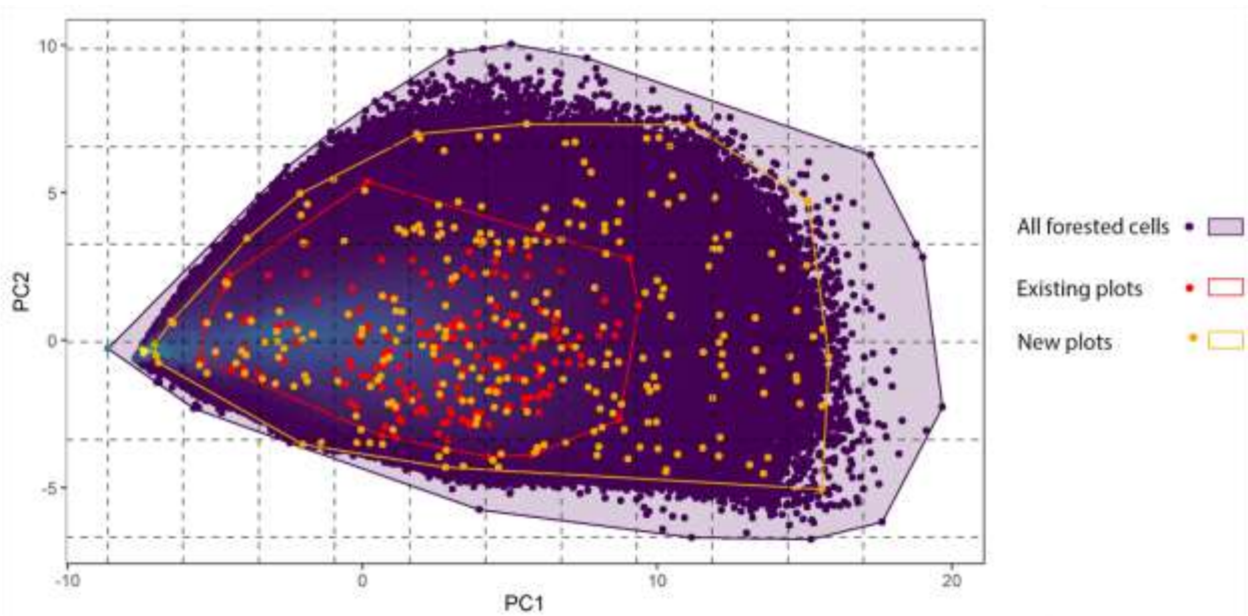

Figure 1. Feature space formed by PC1 and PC2 values of all cells intersecting productive forest types (shaded from purple to yellow from low to high point density), existing plot network (red) and new plot network resulting from the structural based sampling. Solid lines indicate convex hulls of PC1 and PC2 values. Vertical and horizontal dashed lines indicate the equal intervals limits used to separate strata.

In order to minimize the effect of GPS geolocation errors, sampling was performed in priority within cells surrounded by the same strata in a 3 x 3 window. If the required number of sampled cells could not bet met under this condition, cells surrounded by neighbouring strata in the 5 x 10 strata matrix were considered for sampling. Finally, cells with isolated strata were only considered in the case if not enough cells could be sampled under the two aforementioned conditions. A summary of the sampling algorithm is presented in Figure . A total of 168 new plots was selected and 90 existing plots were kept to create a new plot network of 258 plots that cover the entire range of structural variability in the RMF (Figure 1).

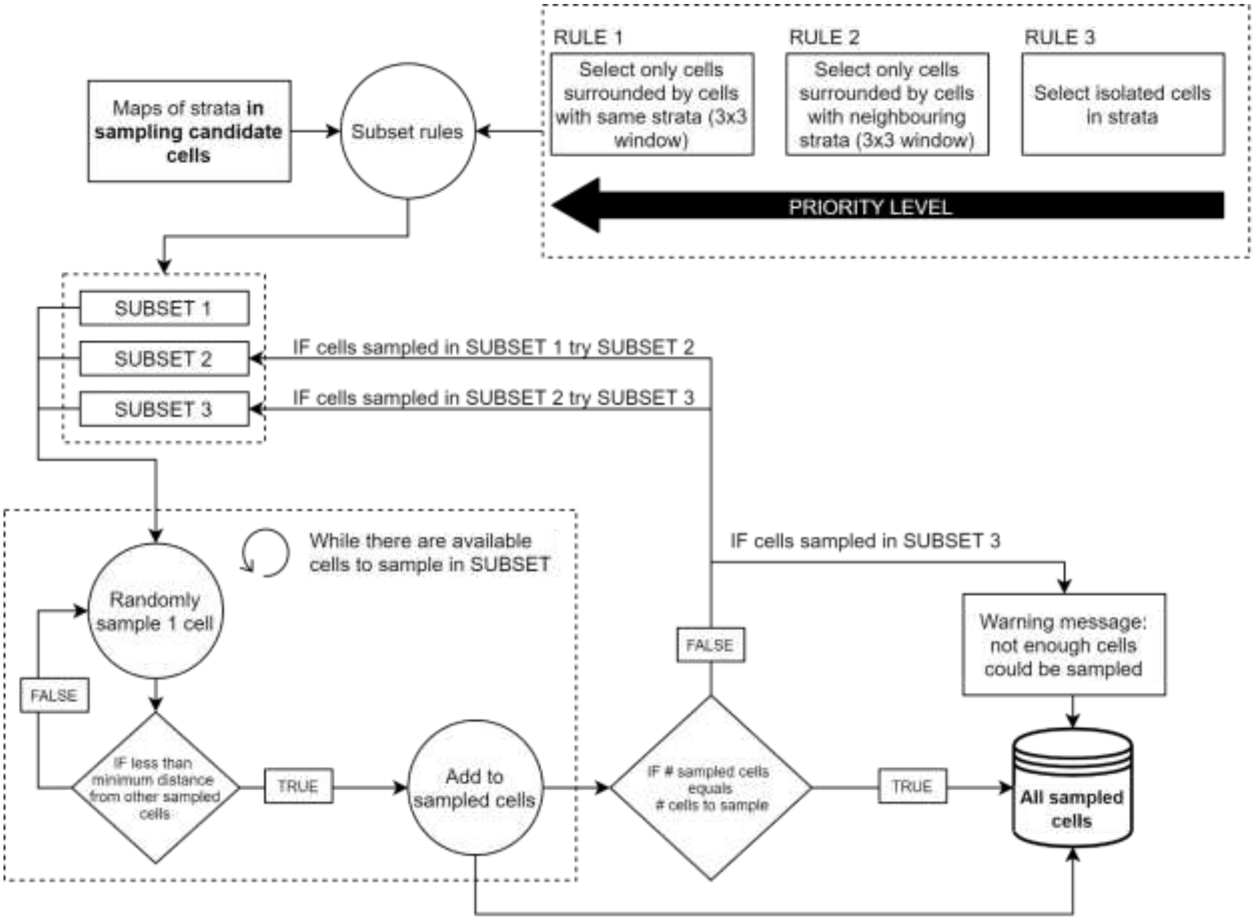

Figure 5. Flowchart summarizing the algorithm used to sample new cells

3. 2019 Field data collection

Through the spring of 2019 MNRF, UBC, RYAM and CWFC developed a new fixed area (400 m^2, with an 11.28 m radius) plot design. Data collected as part of this pilot project serves multiple purposes, including calibration and validation of the new lidar for development of forest resources inventory (FRI), individual crown delineations and species composition assessments, growth and yield modelling, digital soil mapping as well as the opportunity to evaluate the sampling protocols and logistics. The data are also intended to inform selection of additional type C and D plots (Integrated Monitoring Framework (IMF) long-term, multi-purpose monitoring plots).

Once the field plot specification was complete, RYAM led a competitive request for proposal process that was open to all existing Vendor of Record (VOR) contractors that were involved in the T1 eFRI field data collection. Out of this request for proposal process, Sumac Geomatics was selected as the successful vendor to provide field data collection services for the project.

This project required the measurement of 258 fixed area plots that have been pre-selected to represent the range of structural variability that exists in RMF (WP2). Where possible, permanent forest growth plots were selected and re-measured. Other criteria, such as distance to roads (30–200 m), were also considered during plot selection. Each of the 258 plots is classified into one or more of four plot types (A, B, C, D) required to meet the overall project objectives. Table 2 below describes the breakdown of plot by type. A base set of attributes were assessed on all plots (Type A attributes), with added modules requiring additional attributes to be measured on plot types B through D.

The plot data collection began in June 2019 and was completed in early December 2019. Following completion, data entry was done by MNRF growth and yield staff to produce a digital database of the field collected data. A draft plot database was provided to RYAM and UBC in late February 2020 with a final version delivered March 9th, 2020.

Table 2. Number of plots by plot type

Plot Type	Number
A	183
B	50
C	19
D	6
TOTAL	**258**

Figure 6 shows the distribution of measured tree species across the 258 plots.

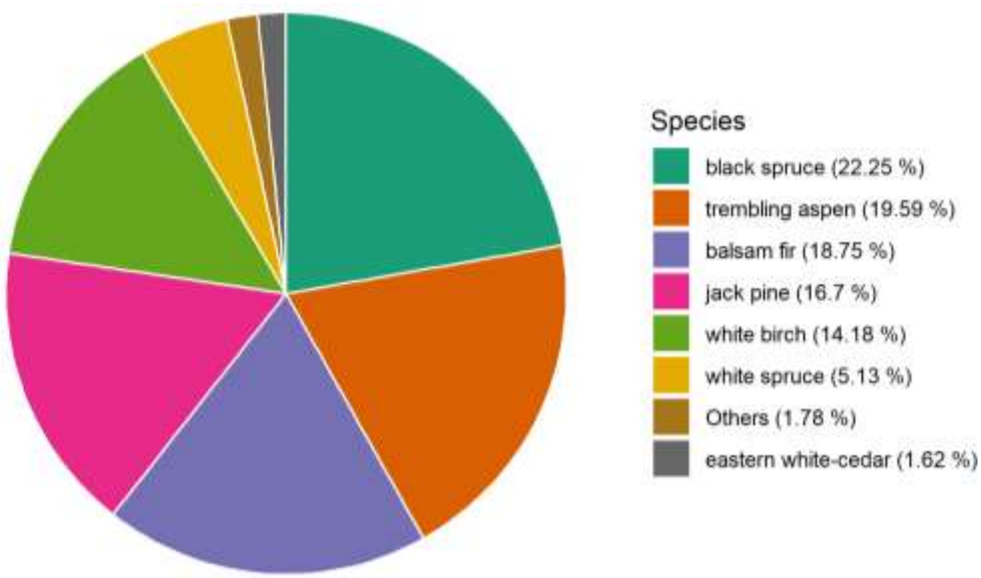

Figure 6. Distribution of measured tree species (number indicated in parenthesis) with the new network of 258 plots

4. Enhanced Forest Inventory using SPL data and field data

4.1 Description of the area-based approach

The estimation of forest attributes in EFI is commonly carried out using an area-based approach (ABA) which consists in generating models that predict plot-level inventory data with spatially coincident ALS point cloud metrics and using these models together with wall-to-wall gridded ALS metrics to create maps of forest inventory attributes. The ABA is presented in details by White et al., (2013). In summary, it consists in the following steps:

1. Collection of fixed-area circular ground plots data
2. Calculate plot-level forest attributes
3. Clip ALS point cloud to ground plots
4. Generate ALS metrics at the ground plot level
5. Generate forest attributes predictive models based on plot-level ALS metrics
6. Generate wall-to-wall maps of gridded ALS metrics. The resolution of the maps is determined so that the area of a pixel corresponds to the plots fixed-area (e.g. 20 x 20 m for plots with fixed-radius of 11.28 m, corresponding to 400 m^2)
7. Use predictive models and wall-to-wall ALS metrics to generate maps of forest attributes

4.2 Measured and modeled forest attributes

The following forest attributes were either directly derived or modeled from field inventory data: Lorey's height (L) , basal area (BA), quadratic mean DBH (QMDBH), stem density (D), whole stem volume (V), merchantable stem volume (VM) and above-ground biomass (AGB). Table 3 provides a description of these forest attributes and their calculation or modeling method.

Table 3. Description of forest attributes calculated or modelled from the field inventory data

Forest attribute	**Description**	**Calculation or modeling method**	**Unit**
Basal area (BA)	Tree cross sectional area (approximated as a circle) at breast height (1.3 m)	$\frac{\pi}{4}\sum_i^n DBH_i^2 \times \frac{1}{A}$, where n is the number of stems and A the plot area in ha	m^2/ha
Lorey's height (L)	Average tree height weighted by basal area	$\frac{1}{n} \times \sum_i^n h_i \times BA_i$, where n is the number of stems and h is the tree height	m
Quadratic mean DBH (QMDBH)	Quadratic mean of DBH	$\sqrt{\frac{\sum_i^n DBH_i^2}{n}}$, where n is the number of stems	cm
Stem density (D)	Number of stems with DBH > 7.1 cm per ha	$\frac{n}{A}$, where n is the number of stems and A the plot area in ha	ha^{-1}
Whole stem volume (V)	Total whole stem volume normalized per hectare	Honer, (1983) and (C. H. Ung, Guo, & Fortin, 2013)	m^3/ha
Merchantable stem volume (VM)	Total merchantable volume normalized per hectare. Stump height is set to 0.2 m and minimum to diameter to 10 cm	Honer, (1983) and C. H. Ung et al., (2013)	m^3/ha
Above-ground biomass (AGB)	Total tree biomass normalized per hectare	Ter-Mikaelian & Korzukhin, (1997) and C.-H. Ung, Bernier, & Guo, (2008)	t C / ha

4.3 ALS metrics

A set of 36 standard lidar metrics, listed in Table 4, were calculated from the SPL point cloud both at the plot and wall-to-wall levels using the lidR package (Roussel & Auty, 2019). More information on these metrics is available at https://github.com/Jean-Romain/lidR/wiki/stdmetrics and in the package documentation.

Table 4. Description of the ALS metrics calculated with the lidR package.

ALS metric description	Abbreviation
Maximum height of all returns	zmax
Mean height of all returns	zmean
Standard deviation of all returns heights	zsd
Skewness of all returns heights	zske
Kurtosis of all returns heights	zkur
Percentage of returns above zmean / 2 meters	pzabovemean / pzabove2
Xth percentile of all returns height distribution	zqx (x = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 65, 70, 75, 80, 85, 90, 95)
Cumulative percentage of returns in the xth layer	zpcumx (x = 1, 2, 3, 4, 5, 6, 7, 8, 9)
Entropy of all returns heights	zentropy

4.4 Modeling framework and accuracy assessment

A random forest regression approach was chosen to build predictive models of the forest attributes based on the SPL metrics described above. Modeling was performed in R using the `randomForest` package (Liaw & Wiener, 2002) and the `caret` package (Kuhn et al., 2015).

The number of tree was set to 500 for each model and the number of variables randomly selected at each tree node (mtry parameter) was determined by training one random forest model for mtry values varying from 2 to 36, the total number of SPL metrics. The mtry value of the model achieving the minimum root mean square error, as defined below, was selected.

Models accuracy was assessed using an iterative k-folds (k = 5) cross-validation approach. At each iteration, a model is trained using k – 1 folds and used to predict response variables on the remaining held-out fold. This process is iterated k times until each fold has been used both as a training a testing set. Folds are created ensuring that they cover approximately the same distribution of observed values. The root mean square error (RMSE), coefficient of determination R^2 and bias were calculated for each held-out fold as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (2)$$

$$bias = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (3)$$

where y_i the observed value of the response variable, $\hat{y}_i$ the predicted value and n is the number of observations id the fold hold-out for validation. Relative RMSE and relative bias were also calculated by dividing by the mean of the observed values. Table 5 summarizes the average and standard deviation of R^2, RMSE and bias calculated on the 5 folds and Figure 7 shows the scatterplot of predicted against observed forest attributes values.

Table 5. Average and standard deviation (in parenthesis) of accuracy measures obtained from the 5-folds cross-validation.

Forest attribute	**R2**	**RMSE**	**Relative RMSE**	**Bias**	**Relative Bias**
Lorey's height	0.88 (0.06)	1.85 (0.55)	9.94 (2.86)	0.02 (0.11)	0.10 (0.58)
Basal area	0.81 (0.03)	7.57 (0.54)	20.29 (1.17)	0.15 (0.98)	0.43 (2.65)
QMDBH	0.76 (0.11)	2.92 (0.65)	14.87 (3.37)	0.03 (0.51)	0.21 (2.62)
Stem density	0.64 (0.06)	409.90 (54.29)	30.61 (2.84)	1.87 (93.51)	0.33 (6.82)
Whole stem volume	0.89 (0.03)	73.09 (12.88)	21.91 (3.79)	2.89 (10.41)	0.85 (3.12)
Merchantable volume	0.90 (0.02)	65.39 (7.66)	23.08 (2.28)	1.59 (13.94)	0.53 (4.81)
Above-ground biomass	0.85 (0.03)	39.88 (3.27)	22.53 (2.02)	0.31 (5.11)	0.23 (2.91)

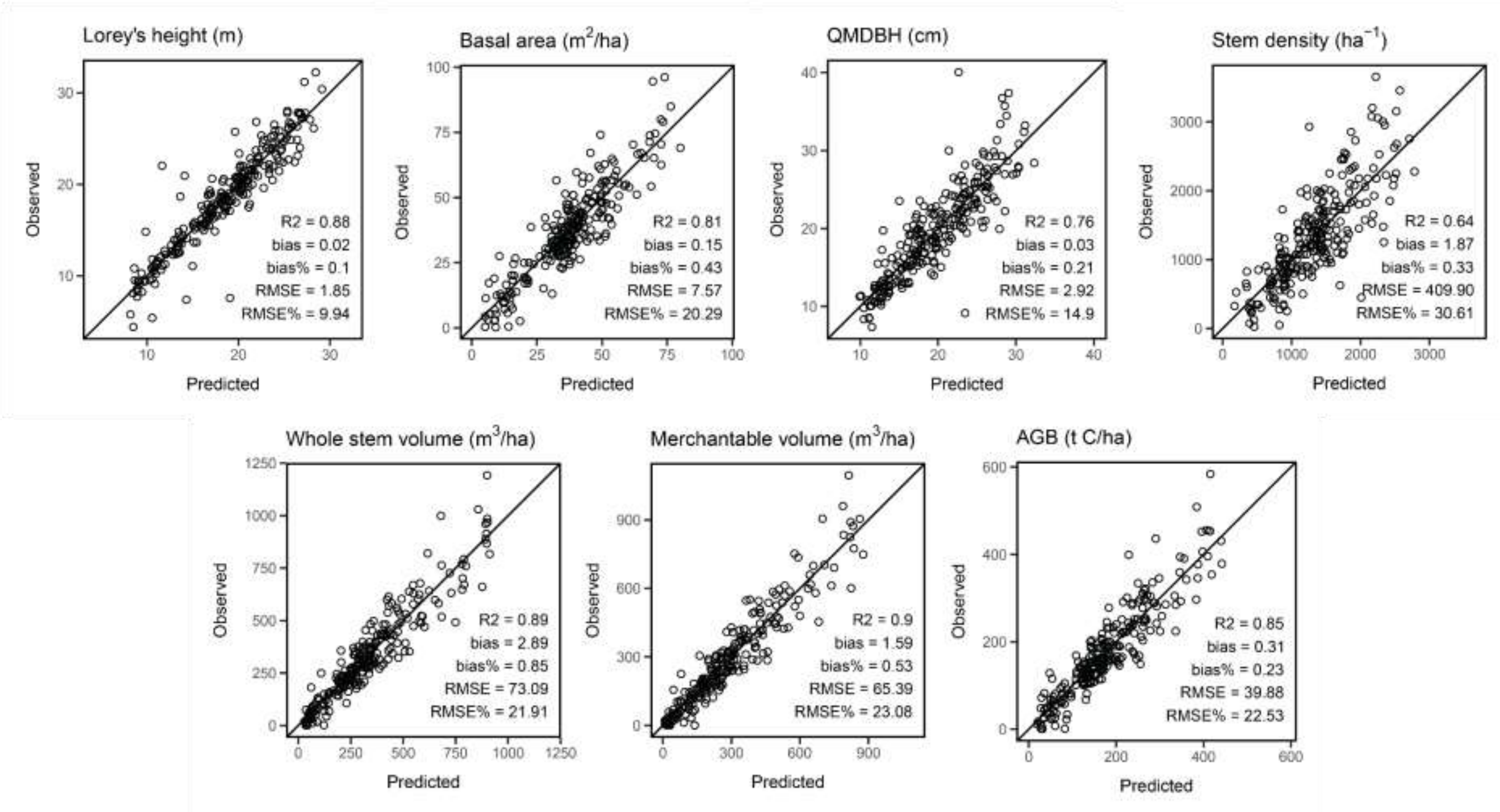

Figure 7. Scatterplots of predicted vs observed values of the forest attributes combined from the 5 independent folds of the k-fold cross-validation.

Finally, the importance of each SPL metrics in the models was assessed using the random forest variable importance measure which is calculated as the percentage decrease of the prediction mean square error when the values of a variable are randomly permuted while others remain unchanged. Figure 8 shows all the SPL metrics that were part of the 5 most important variables in at least one of the random forest models.

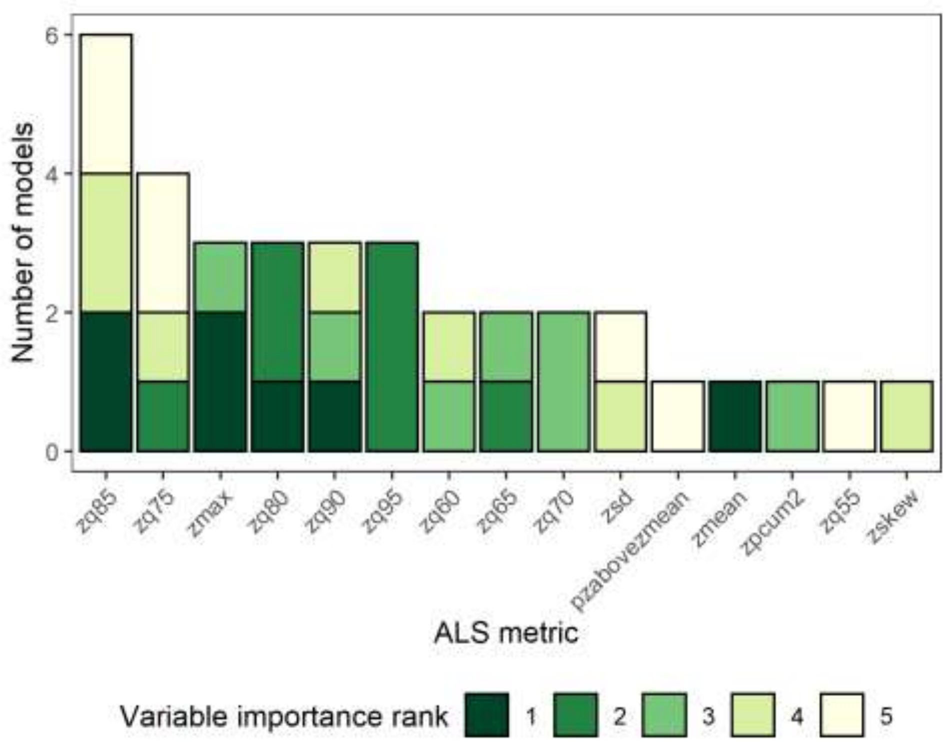

Figure 8. Summary of the SPL metrics importance rank combined from all the models. Metrics that were not at least the 5th most important variable in any of the model are not plotted.

5.0 Next Steps and Integration with Ontario Ministry of Natural Resources and Forestry.

Now that initial ABA models have been developed, we will continue to examine the best approaches and software tools for both model development and applying these ABA models over large management areas. We will work with staff from the Ontario Ministry of Natural Resources and Forestry on cloud-based solutions to extrapolate these developed ABA models over larger management areas focusing on the Romeo Mallete forest management area, and work closely with staff to ensure effective transfer of ABA building tools and best practices.

Ongoing research will focus on the development of individual tree based approaches (ITD) using the SPL data at 50 plots were detailed stem measurements were taken and using these insights to examine the possibility of a hybrid ABA / ITD based inventory alterative and its usefulness to both the and Province and industry.

References

Honer, T. G. (1983). Metric timber tables for the commercial tree species of Central and Eastern Canada. *Can. For. Ser. Inst. Info*, *FMR-X-5*, 139. Retrieved from http://library.wur.nl/WebQuery/clc/188227

Isenburg, M. (2014). LAStools. Retrieved from http://rapidlasso.com/LAStools

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., … Scrucca., L. (2015). caret: Classification and Regression Training, https://CRAN.R-project.org/package=caret. *R Package Version*. Retrieved from http://cran.r-project.org/package=caret

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(December), 18–22. https://doi.org/10.1023/A:1010933404324

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Roussel, J.-R., & Auty, D. (2019). lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. Retrieved from https://cran.r-project.org/package=lidR

Ter-Mikaelian, M. T., & Korzukhin, M. D. (1997). Biomass equations for sixty-five North American tree species. *Forest Ecology and Management*, *97*(1), 1–24. https://doi.org/10.1016/S0378-1127(97)00019-4

Ung, C.-H., Bernier, P., & Guo, X.-J. (2008). Canadian national biomass equations: new parameter estimates that include British Columbia data. *Canadian Journal of Forest Research*, *38*(5), 1123–1132. https://doi.org/10.1139/X07-224

Ung, C. H., Guo, X. J., & Fortin, M. (2013). Canadian national taper models. *Forestry Chronicle*, *89*(2), 211–224. https://doi.org/10.5558/tfc2013-040

White, J. C., Wulder, M. A., Varhola, A., Vastaranta, M., Coops, N. C., Cook, B. D., … Woods, M. (2013). *A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach*. *Forestry Chronicle* (Vol. 89). https://doi.org/10.5558/tfc2013-132

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., & Treitz, P. (2011). Operational implementation of a LiDAR inventory in Boreal Ontario. *The Forestry Chronicle*, *87*(04), 512–528. https://doi.org/10.5558/tfc2011-050