

Article

Not peer-reviewed version

Airborne LiDAR for Basal Area Estimation: Accuracy Assessment and Improvement in Eastern Canada's Mixed Temperate Forests

[David Normandeau](#)*, [Daniel Beaudoin](#), [Martin Riopel](#), [Hakim Ouzennou](#)

Posted Date: 6 February 2026

doi: 10.20944/preprints202602.0535.v1

Keywords: LiDAR; forest inventory; shade-tolerant deciduous; hardwood; uneven-aged forest; stand basal area; open-source machine learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Airborne LiDAR for Basal Area Estimation: Accuracy Assessment and Improvement in Eastern Canada's Mixed Temperate Forests

David Normandeau ^{1,*}, Daniel Beaudoin ¹, Martin Riopel ² and Hakim Ouzennou ²

¹ Department of Wood and Forest Sciences, Université Laval, Québec City, QC G1V 0A6, Canada

² Ministère des Ressources naturelles et des Forêts, Direction des Inventaires Forestiers, 5700, 4e Avenue Ouest, Québec City, QC G1H 6R1, Canada

* Correspondence: david.normandeau.2@ulaval.ca

Abstract

Sustainable forest management requires current, territory-wide data, which is difficult to obtain in vast regions like Quebec, Canada. To complement ground inventories and photo-interpretation, the province developed an ALS-based model that performs well in coniferous stands, but its accuracy in hardwood stands remains untested. This study aims to evaluate the precision of the ALS-based prediction of stand basal area and then test new approaches to increase its performance. Airborne LiDAR data from 2011 to 2020 and 12 506 validation plots from sample plots were used. The ALS model precision was initially compared across the stand types, revealing lower accuracy in shade-tolerant deciduous stands. Three inputs were found to increase prediction accuracy: proportion of each species basal area in the stand, geographical coordinates, and meteorological data associated with location. Parametric and auto machine learning method (AutoML) were employed using those inputs to improve precision, with Auto ML achieving the highest improvement with initial R^2 of 27%, 47% and 54% and after correction R^2 of 31%, 56% and 67% respectively for shade-tolerant deciduous, shade-intolerant deciduous, and coniferous stand. Even with the advancements made, further improvements will be necessary to consider using an ALS-based model for shade-tolerant deciduous species.

Keywords: LiDAR; forest inventory; shade-tolerant deciduous; hardwood; uneven-aged forest; stand basal area; open-source machine learning

1. Introduction

The forested area of Quebec covers nearly 905,800 km², which is more than the four countries with the most forest in Europe combined, excluding Russia (Sweden, Finland, Spain, France) [1,2]. In order to achieve sustainable forest management and effective operational planning in this vast territory, it is essential to have a reliable, complete and sufficiently rapid inventory technique for up-to-date monitoring. Stand basal area (BA) is an important variable for evaluating forest productivity [3]. It represents the area covered by trees at breast height per hectare. It is highly correlated with volume and growth of forest stands [4]. The classical method for determining stand BA uses ground sampling data and applies linear/nonlinear regressions models [5–8] or simultaneous equation methods [9,10] to extrapolate it to the entire stand. These approaches are becoming less common with the potential of new automated survey methods using technologies such as LiDAR (light detection and ranging) and photogrammetry [11].

The use of airborne laser scanning (ALS) for large-scale forest inventory has been used for several years and has shown that it can be the primary tool for evaluating stand dendrometric metrics [12]. In Québec (Canada), the prediction of forest metrics, including BA, is done using the k-nearest neighbors method [13] with data from an airborne LiDAR when available. This approach has

limitations as it relies on ground-based data and photo-interpretation to generate and update its results. Annually, thousands of temporary and permanent samples are collected until the entire province is covered, temporary samples are taken once, while permanent samples are revisited to track changes over the years. It takes about 10 years to update the full inventory [14] which makes it expensive and time-consuming to collect, it also requires significant labor in a labor shortage context [15,16]. From 2022, a tool called the LiDAR Dendrometric Map (LDM) was developed to predict forest metrics, derived from density and height of captured pulses combined with the effective ecoforest map data and a relative density index [17]. This area-based approach (ABA) map is currently available only for bioclimatic domains dominated by conifers in northern Quebec. This approach has shown great potential since, without needing new costly ground sample plot acquisition programs, it's possible to map forest metrics using ALS with a good level of accuracy in the boreal forest of Quebec, where the main species are coniferous or shade-intolerant deciduous [17,18].

The province now wants to extend the map to southern regions where shade-tolerant species are located. Before extending the ALS-based tool to southern forests, an essential step is to validate whether the tool will achieve the same accuracy in predicting basal area as it does in boreal forests. While ALS scanning approaches have demonstrated their efficiency in boreal forest with coniferous and shade-intolerant deciduous species [19–22], this has not been the case for shade-tolerant deciduous forests [23–25]. Since ALS mainly gives information about crown heights and the relation between crown height and mean diameter at breast height (DBH) is weak in these forests, the accuracy tends to be lower [23,24]. To enhance forest metrics estimation from LiDAR, incorporating more canopy characteristics beyond vertical structure is necessary [24,26]. For the LDM, a relative density index (RDI) was generated using inventory data to predict crown closure, thereby strengthening the model. Data from the ecoforest map were also utilized, including dominant species, the proportion of merchantable conifer (FSPL), the standard version of the map, and two ecological land classifications [17]. Given the unique approach used by the LDM for generating forest metrics, it is justified to test its effectiveness in shade-tolerant deciduous stands.

It is known that height-to-diameter relationships are related to growing environment and stand conditions [27,28], and the growth of shade-tolerant species is more influenced by abiotic factors than the growth of intolerant species [29,30]. Targeting improvements, adding data from outside the ecoforest map such as climate data, latitude and longitude to complement independent variables used by the LDM has not been tested. A tool called BIOSIM has been developed in Canada based on weather station data and provides a range of geospatial climate metrics that can be merged with the model [31]. To add independent variables to a model, using a linear or non-linear parametric model is an option that is frequently used in forestry [20,24]. Another way is to use machine learning and train a new model with validation data [32]. In forestry and biology, certain types of machine learning, such as neural networks, Maximum Likelihood and random forests are employed [33,34], but the use of ongoing advancements such as automated machine learning is still emerging [35]. This approach allows one to test different types of machine learning or mix them to obtain the optimal solution for a specific situation [36]. Although the use of this approach facilitates interaction compared to other methods of machine learning, certain criteria such as the proper selection of training and validation data, as well as suitable inputs, are crucial to maximize effectiveness [37]. The large amount of inventory data in Quebec, both from permanent and temporary plots coinciding with LiDAR acquisition, provides an opportunity to test this approach [38–40]. With a robust database and careful selection of inputs, the use of this modern method could lead to increased precision in forest types dominated by shade tolerant deciduous trees, aiming to achieve a level of accuracy similar to that found in boreal forests. Characterizing model error [41] and calculating mean bias, precision and accuracy of a model's predictions are essential steps of the quantitative validation process [42] and are necessary to make a decision on the validity of a model [43] and choose the best model.

The objectives of this study were as follows: (i) assess the accuracy of predictions for stand BA obtained with the LDM across three distinct stand types: shade-tolerant deciduous, shade-intolerant deciduous and conifers. (ii) identify and test independent variables that can increase the precision of

BA predictions, including variables originally used for LDM production as well as additional geospatial and climatic factors. (iii) test model improvements by comparing linear, non-linear, and automated machine learning approaches to enhance the accuracy of LDM predictions.

2. Materials and Methods

2.1. Study Area

The study area (94,800 km²) is located within the bioclimatic domain of the balsam fir-yellow birch forest, which crosses the southern middle of Quebec from west to east (Figure 1)[44]. It extends between the 47th and 49th parallels. The elevation is relatively low, with less than 500 meters in altitude, and the terrain is generally gently undulating. Annual precipitation ranges between 1100 and 1600 mm, with 30% falling as snow. The average temperature ranges from 0 to 2.5°C, and the annual growing season lasts for 160 to 170 days [45,46]. The area includes stands of coniferous, mixed, and deciduous species, primarily dominated by *Abies balsamea*, *Picea mariana*, *Betula papyrifera*, *Betula alleghaniensis*, *Populus tremuloide* and *Acer saccharum*. Within the study area, coniferous and shade-intolerant deciduous stands are managed as even-aged forests in public forests, while shade-tolerant stands are managed as irregular forests [47].

This area represents the southernmost bioclimatic zone covered by the LiDAR Dendrometric Map and serves as the boundary between coniferous and mixed deciduous forests. This makes it the logical choice for conducting the study due to its significant proportion of stands dominated by deciduous species.

2.2. Data

2.2.1. Airborne LiDAR Data and Basal Area prediction (Predicted Basal Area)

The predicted basal area tested in this study is derived from the LiDAR Dendrometric Map (LDM). Developed by the Ministry of Natural Resources and Forest (MNR), the LDM provides predictions of essential forest attributes, including the basal area (BA), which is the focus of this study. The LDM predicts the basal area from the following method:

- The basal area (BA) prediction process is described in Leboeuf et al. (2022) [17], where:
 - BA is derived from a form factor, dominant height, and volume prediction.
 - Predictions are generated using a linear parametric model that incorporates independent variables derived from LiDAR and ecoforest map data.
 - LiDAR data acquisition in the study area took place between 2011 and 2020 during leaf-on periods.
 - Several sensors were used for data collection, with pulse density ranging from 2 to 6 points per square meter.
 - Variable selection was carried out using stepwise linear regression.
 - Basal area predictions are available at a 5 m x 5 m raster scale.
- Known independent variables, including LiDAR-derived, MNR custom-generated, and ecoforest map variables, are presented in Table 2 in Section 2.2.3 (Independent Variables).

2.2.2. Ground Sample Plots (Observed Basal Area)

To test the precision of the predicted basal area, a network of ground sample plots from the MNR was used for the comparison. These plots include permanent plots, which are revisited periodically to track stand evolution, and temporary plots, which are specifically established for 10-year inventory. Each of these circular plots, measuring 400m² (0.04ha) of 11.28 m radius, includes data on tree count, average stand height, and for each tree, the diameter at 1.3 m from the ground (DBH) and tree species [48–51]. The study focuses on the inventory data collected between 2006 and 2020, covering the 4th and 5th decade inventories.

To ensure an accurate comparison of the BA between the ground sample plots and the LDM results, the ground sample plots had to meet four criteria.

1. **LiDAR Coverage:** The ground sample plots must be located in an area covered by LiDAR airborne data.
2. **Measurement Timeframe:** The selected plots must have been measured no more than five years before or after the LiDAR acquisition date.
3. **Absence of Natural Disturbances:** No natural disturbances (e.g., fires, storms, insect outbreaks) must have occurred between the LiDAR acquisition and their field measurement.
4. **Absence of Human Intervention:** The plots must not have undergone any human interventions (e.g., logging, thinning, road construction) between the LiDAR acquisition and their field measurement.

Using these criteria, 12 506 plots were selected to represent three dominant species composition across the study area (shade-tolerant deciduous, shade-intolerant deciduous and conifers) (Figure 1, Table 1). The dominant species composition was established based on the main species indicated in the ecoforest map [52]. Yellow birch, sugar maple, red maple, and ash comprise the class of shade-tolerant deciduous. Aspen and white birch constitute the class of shade-intolerant deciduous. Spruce, larch, pine, fir, and cedar form the class of conifers. Finally, the basal area of each plot (observed value for the analysis) was calculated by summing the basal areas of all trees within it. We will refer to this reference value as the “observed basal area” in the remainder of the paper.

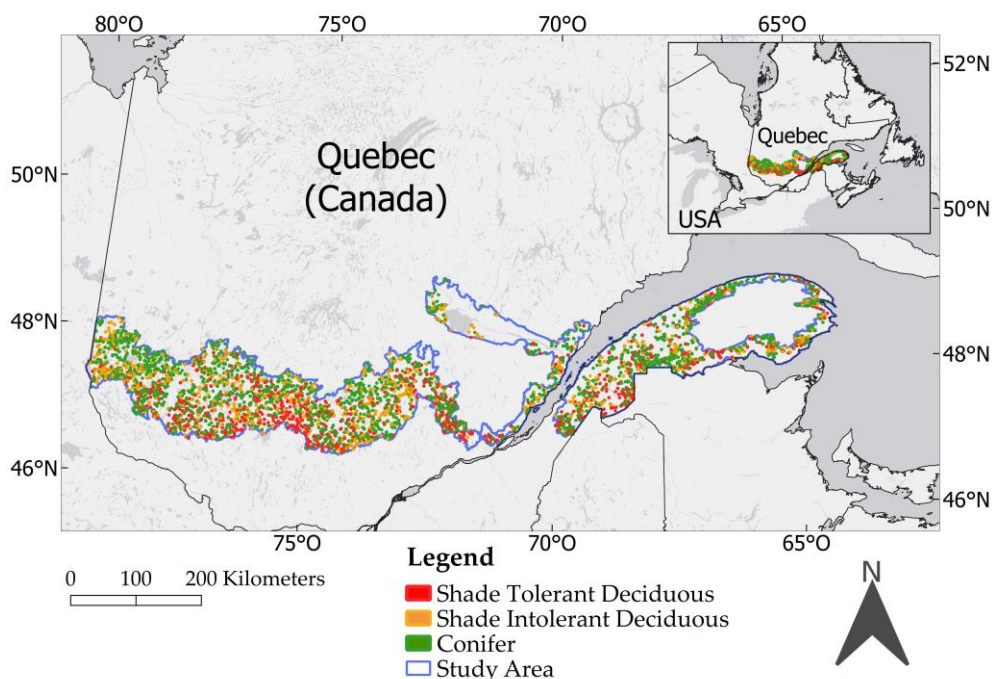


Figure 1. Field Ground Sample Plots Distribution by Stand Type (Color Dots) in the Balsam Fir-Yellow Birch Forest of Quebec (Study Area)

To ensure an adequate sample size, literature suggests a minimum of 375 plots of 0.04 ha per stand type to achieve stability in the results [53] and the applied selection criteria allowed this threshold to be exceeded for all stand types (Table 1)

Table 1. Distribution of plots by stand type and observed basal area.

Dominant species composition	Number of plots	Mean Basal Area (m ² /ha)
Shade-tolerant deciduous	2 590	24.22 (0.18 – 85.56)
Shade-intolerant deciduous	3 217	23.77 (1.15 – 61.51)
Conifers	6 699	24.62 (0.20 – 107.52)

Total	12 506	24.32 (0.18 – 107.52)
--------------	---------------	------------------------------

2.2.3. Independent Variables

Independent variables in the study are split into two groups: those used to build the LiDAR Dendrometric Map (LDM) [13,18,52] and additional data sources aimed at improving the accuracy of basal area prediction.

The source data for the LDM were provided by the MNRF and include LiDAR-derived, MNRF custom-generated, and ecoforest map variables. Details on LiDAR data acquisition parameters are presented in Section 2.2.1. (Airborne LiDAR Data and Basal Area prediction (Predicted Basal Area)).

Selection of the additional variables followed criteria of geolocated availability, continuous numerical type, coverage across the entire study area, and ability to explain differences or similarities between stand types. The additional set includes species composition from the MNRF ecoforest map, climatic data (1981-2010 mean) derived from geographical location and the BIOSIM software tool [31,54], as well as geographic coordinates (latitude, longitude). These additional datasets are derived from the ecoforest map database [52].

Table 2 contains a detailed description of all variables.

Table 2. Description and value ranges of independent variables used in the study.

Variable	Description	Values
DENS_CLASS* ¹²³⁴ ₇	Canopy density. Calculated as the proportion of LiDAR points that do not reach the ground (%)	0 to 95
DH* ²³⁴⁵⁷	Dominant height. average height of the highest LiDAR return in each 5×5 m pixel within the 11.26 m plot. (m)	0.964 to 29.331
H_CLASS* ¹²³⁴⁵⁶⁷	Height above ground of the tallest tree. LDM polygon 1-m scale classification computed from DH (m).	4 to 30
LGDH*	Log ₁₀ (DH)	
RDI*	Relative density index indicating stand quality [17]	0.000875 to 1.212438
R_RDI_DH* ¹²³⁴⁵⁶⁷	RDI/ DH	
LGRDI* ²³⁴⁵⁶⁷	Log ₁₀ (RDI)	
FF10* ²³⁴⁵⁶⁷	Form factor used to estimate basal area from volume and dominant height. Predicted with LiDAR dendrometric map	0 to 0.415
N10* ¹²³⁴⁷	Estimated number of merchantable stems (trees) per hectare. Predicted with LiDAR dendrometric map	0 to 2938.899
ALTITUDE*	Altitude (m) derived from digital terrain model	0 to 890
LGALT* ²³⁴⁵⁶⁷	Log ₁₀ (ALTITUDE)	0 to 2.938
dSPEC_PR* ¹²³⁴⁵⁶⁷	Dominant species of the stand according to the ecoforest map where PR correspond to EN (<i>Picea mariana</i>), EB (<i>Picea glauca</i>), SB (<i>Abies balsamea</i>), BP (<i>Betula papyrifera</i>), BJ (<i>Betula alleghaniensis</i>), ES (<i>Acer saccharum</i>), EO (<i>Acer rubrum</i>), PE (<i>Populus spp.</i>), TO (<i>Thuja occidentalis</i>), PG (<i>Pinus banksiana</i>), AF (other deciduous species), AR (other coniferous species), ML (<i>Larix laricina</i>), FN (<i>Fraxinus nigra</i>) and PB (<i>Pinus strobus</i>). Factor/ categorical value.	"EN", "EB", "SB", "BP", "BJ", "ES", "EO", "PE", "TO", "PG", "AF", "AR", "ML", "FN", "PB"
pXXX ¹²³⁴⁵⁶⁷	Proportion of the basal area in the stand cover by species (%) where XXX corresponds to WSP ¹⁴⁶⁷ (<i>Picea glauca</i>), BSP ¹⁴⁶ (<i>Picea mariana</i>), BFI ²³⁴⁶⁷ (<i>Abies balsamea</i>), JPI ¹⁴⁷ (<i>Pinus banksiana</i>), TAM ⁴⁷ (<i>Larix laricina</i>), FSPL ²³⁴⁷ (fir, spruce, pine, and larch), WPI ⁷ (<i>Pinus strobus</i>), OPI ⁷ (other pine species, mostly <i>Pinus resinosa</i>), OCON ¹²³⁴⁵⁶⁷ (other conifers, mostly <i>Thuja occidentalis</i>), TCON ¹²³⁴⁷ (all conifers), WBI ¹³⁴⁶⁷ (<i>Betula papyrifera</i>), YBI ¹²³⁴⁷ (<i>Betula alleghaniensis</i>), ASPE ¹⁴⁶⁷ (<i>Populus spp.</i>), RMA ⁴⁷ (<i>Acer rubrum</i>),	0 to 100

	SMA ³⁴⁷ (<i>Acer saccharum</i>), ODEC ⁴⁷ (other deciduous, mostly <i>Quercus rubra</i> , <i>Fraxinus nigra</i> and <i>Fagus grandifolia</i>)	
DEGRE_DAY ¹³⁴⁷	Annual sum of daily average temperatures above 5°C (Tavg – 5) for days where Tavg > 5°C (°C)	977.93 to 1660.48
PRECI_GRS ¹⁴⁶⁷	Total precipitation during the growing season (GRS) (mm)	380.32 to 664.68
PRECI_SNOW ¹⁴⁷	Annual total of snow precipitation converted to liquid equivalent, considering only days where Tavg < 0°C (mm)	229.93 to 574.15
PP_SNOW ¹³⁴⁷	Proportion of total precipitation in the form of snow (%)	27.95 to 45.75
TMIN_AN ³⁴⁵⁶⁷	Annual average of daily minimum temperatures (°C)	-6.34 to -0.60
TMOY_AN ⁴	Annual average temperature (°C)	-0.42 to 4.11
TMAX_AN ¹³⁴⁵⁶⁷	Annual average of daily maximum temperatures (°C)	5.35 to 9.64
TMOY_GRS ¹²³⁴⁵⁶⁷	Average temperature during the growing season (°C)	12.49 to 14.84
TMOY_JULY ³⁴⁵⁶⁷	Average temperature in July (°C)	14.26 to 18.51
FF_TOT ¹⁴⁵⁶⁷	Total number of frost-free days (Tmin > 0°C)	137 to 193
CONSE_FF ⁴⁶⁷	Number of consecutive frost-free days (Tmin > 0°C)	49 to 137
DAY_GRS ⁴⁶⁷	Number of days in the growing season (GRS)	115 to 166
FDAY_FF ¹⁴⁶⁷	First day of the consecutive frost-free days period (Julian day)	135 to 179
VPD_UTI ¹³⁴⁷	Cumulative vapor pressure deficit for June, July, and August (Julian days 152 to 243) (mbar)	1075.13 to 1551.57
ARID_TOT ¹⁴⁵⁶⁷	Sum of monthly water deficits based on the difference between monthly precipitation and Thornthwaite potential evapotranspiration (0 if negative) (cm)	0.49 to 5.88
RADIA_T ⁴⁷	Annual sum of energy emitted by solar radiation (MJ/m ²)	4580.70 to 5414.55
RADIA_GRS ⁴⁷	Energy emitted by solar radiation during the growing season (GRS) (MJ/m ²)	2327.24 to 2921.49
LATITUDE ¹²³⁴⁵⁶⁷	Geographic coordinate that specifies the north-south position of a point on Earth. (dd.ddddd)	46.66424 to 49.22475
LONGITUDE ¹³⁴⁵⁶⁷	Geographic coordinate that specifies the east-west position of a point on Earth. (dd.ddddd)	-79.51612 to -64.32079

* Variables used to build LDM

¹²³⁴⁵⁶⁷Details on the meaning of the index numbers are provided in the Results section 3.2.

LGDH and RDI are not considered in variable selection, because they are directly correlated to other variables. LGDH to DH and RDI to LGRDI.

The relative density index (RDI) and the form factor (FF10) are custom variables generated by the MNRF. RDI is calculated by first calibrating a self-thinning line using quantile regression on the number of stems taller than 5 m from ground sample plots. RDI values are then derived by taking the ratio of the observed stem count to the predicted values from the self-thinning line. Finally, a log-linear model with LiDAR-derived metrics is developed to generate an RDI raster for the study area [17]. The form factor is generated using a bank of measured taper coefficients for each species and then adapted to a 5m x 5m raster [55].

Species composition follows the MNRF ecoforest map of Quebec, which is based on photo-interpretation of high-resolution aerial imagery. Trained photo-interpreters delineate forest stands based on stand resemblance, and species distribution is assessed using basal area cover proportion [56].

Geolocated meteorological data are generated through the BIOSIM software tool [31,54], which was initially developed to analyze insect behavior. BIOSIM utilizes historical meteorological data

collected by weather stations across Canada and extrapolates weather trends between these stations to generate geolocated climate data.

Geographical coordinates consist of the latitude and longitude of the center of each sample plot.

The use of the variables is presented in Section 2.3.2. (Identification of basal area predictors).

2.3. Data Analysis

2.3.1. Asses the BA Accuracy of the LDM

The accuracy assessment of the LDM for predicting basal area (predicted BA) was conducted at the plot level using ground sample plots for comparison (observed BA). To ensure consistency with the ground sample plots, the 5 m × 5 m raster data were converted to match the format of the 11.28 m circular plots. Predicted BA for each plot was obtained by overlaying the plots onto the raster pixels and calculating a weighted average of the BA values from the intersecting pixels.

Once the data were adjusted to the appropriate format, analysis was performed using R statistical software. Accuracy was assessed through statistical analysis using the coefficient of determination (R^2) (Equation (1)), bias (Equation (2)), root mean square error percentage (RMSE) (Equation (3)), mean absolute error (MAE) (Equation (4)), and the fraction of variance of the "expected response" (r_{ER}^2) (Equation (5)) [57].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (\bar{x} - x_i)^2} \quad (1)$$

$$bias = \frac{1}{N} \sum_{i=1}^n (y_i - x_i) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - x_i)^2} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (4)$$

$$r_{ER}^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

Where n is the number of field plots, x_i is the ground field inventory (observed value), \bar{x} is the mean of the field inventory, y_i is the predicted value and \bar{y} is the mean of the predicted value.

To evaluate the influence of species composition on model performance, stands were analyzed not only collectively but also by dominant species. This stratification allowed the assessment of how species dominance affected the effectiveness of the LiDAR Dendrometric Map (LDM). Predictions were evaluated for stands dominated by *Betula papyrifera* (white birch), *Populus spp.* (aspen), *Betula alleghaniensis* (yellow birch), *Acer rubrum* (red maple), *Acer saccharum* (sugar maple), *Fraxinus spp.* (ash), *Picea glauca* (white spruce), *Picea mariana* (black spruce), *Larix laricina* (larch), *Pinus strobus* (white pine), *Pinus banksiana* (jack pine), *Abies balsamea* (balsam fir), and *Thuja occidentalis* (white cedar). In addition, stands were grouped into three stand types (coniferous, shade-intolerant deciduous, and shade-tolerant deciduous) to examine differences in model performance across general stand categories.

2.3.2. Identification of Basal Area Predictors (Independent Variables)

Feature selection is an important step in model improvement. Using fewer variables is more restrictive, as it limits the model to only the most critical predictors, potentially simplifying the model

but risking the loss of valuable information. On the other hand, including more variables is less restrictive, providing the model with a broader range of data, which can capture more complex patterns but may also introduce noise or redundancy [58]. To select the optimal predictors among the variables presented in Tables 2, three variable selection methods were used to identify subsets of relevant variables for the development of the predictive model: stepwise forward and backward selection, Boruta method, and the Least Absolute Shrinkage and Selection Operator (Lasso). Feature selection for each method was performed in R Studio.

Stepwise forward and backward selection [59]: We implemented both forward selection (adding variables one by one) and backward elimination (removing less significant variables) using the Akaike Information Criterion (AIC) to optimize model fit. The process followed these steps:

1. Base model: A linear regression model was initialized with only an intercept.
2. Subsets models: Models including similar predictors subsets were defined.
3. Stepwise selection: For each subset, the `step()` function in R was used to iteratively add or remove variables based on AIC, balancing model complexity and predictive power. The most significant variables from each subset were retained.
4. Final selection: After analyzing the individual subsets, a final stepwise selection was performed on the combined set of shortlisted variables to further refine the model. This final iteration ensured that only the most predictive and non-redundant variables were retained.

Boruta method [60]: Boruta is a wrapper algorithm built around the Random Forest classifier, designed to assess the importance of each predictor by comparing it to randomly permuted "shadow" features. We implemented Boruta using the Boruta package in R. The model was trained with BA as the response variable and all predictor variables as inputs. The algorithm iteratively shuffled feature values and evaluated their importance based on how well they contributed to model predictions. Predictors were categorized as Confirmed, Tentative, or Rejected, based on whether their importance scores significantly exceeded those of shadow features. To refine the selection process, we applied `TentativeRoughFix()`, which resolves ambiguous (Tentative) variables by reclassifying them as either Confirmed or Rejected. The final list of selected predictors consisted of those with nonzero importance scores after this adjustment. We then extracted and ranked the most influential variables based on their mean importance values

Lasso regression [61,62]: Lasso performs both regularization and feature selection by adding an L1 penalty to the regression model, which forces some coefficients to shrink to exactly zero, effectively eliminating less relevant variables. Before applying Lasso, all predictor variables were standardized (mean = 0, standard deviation = 1) to ensure a fair comparison and prevent scale differences from influencing the selection process. We used the `glmnet` package in R, which performs Lasso regression by solving:

$$\min \sum (y_i - \hat{y}_i)^2 + \lambda \sum |\beta_j|$$

Where y_i represents the observed values, \hat{y}_i represent the predicted values, β_j are the model coefficients and λ is a regularization parameter controlling the level of shrinkage. Optimal λ values were determined via k-fold cross-validation using `cv.glmnet()`. Two values were considered: λ_{\min} , which minimizes the cross-validated mean squared error (MSE) and retains more predictors, and λ_{1se} , within one standard error of the minimum, producing a more parsimonious model with fewer variables. Final models were refitted using each λ , generating a list of coefficients. Nonzero coefficients indicate the predictors retained by the model, and their magnitude and sign reflect the relative contribution and direction of each variable on the predicted outcome.

Four primary subsets were generated from the feature selection methods: one from stepwise selection, two from Lasso (λ_{1se} and λ_{\min}), and one from Boruta selection. Additional subsets were constructed based on the variable importance rankings from Lasso and Boruta to match the number of variables selected by stepwise and Lasso λ_{1se} . This procedure resulted in three additional subsets, for a total of seven subsets. These subsets were later compared in the modeling approaches section to evaluate their performance under the best-performing modeling approach.

2.3.3. Modelling Approaches

Three modeling approaches were applied to the variables selected through the stepwise method: linear parametric models, non-linear parametric models, and automated machine learning. Each model was trained on 80% of the dataset and evaluated on the remaining 20% to assess predictive accuracy.

(1) Linear parametric models fitted a linear regression with observed basal area values as the response variable and variables selected by the stepwise method cited in section 2.3.2 above as the explanatory variable. The model can be expressed as follows:

$$Y = A \times X + B \times Z + \varepsilon \quad (6)$$

Where Y is the dependent variable (observed basal area), X is the predicted basal area, Z are the variables explanatory, A and B are the general parameters of the equation and ε is the error term.

(2) Non-linear parametric models [63], quadratic, logarithmic, square root, and reciprocal applying the same methodology as in method one. The performance of each model was tested. The linear model and the best-performing non-linear model were retained for comparison with the other approach.

(3) Using Automatic Machine Learning H₂O (H₂O AutoML) with R package H₂O[64]. It is an open-source tool for automating the predictive analytics workflow. Widely used in academia and industry as well [65]. The H₂O AutoML test different models such as generalized Linear Model (GLM), a default random forest (DRF), gradient boosting machine (GBM), a near default Deep Neutral Nets (DNN) and Stacked Ensemble Model (SEM) that integrates the previous models to determine the optimal combination of prediction algorithms all in one function [66]. The testing process involves running the AutoML function with a time limit of 60 seconds to determine the combination with the best predictive performance.

Model performance was compared across the three approaches and against the LDM. The best-performing approach was subsequently applied to the six additional variable subsets generated through Lasso and Boruta feature selection in order to identify the most effective feature selection strategy.

3. Results

3.1. LiDAR Dendrometric Map Accuracy Assessment

For practical purposes (AAC calculations, harvest planning, etc.), forest inventory must provide useful information by species or groups of species. Table 3 displays the average observed and predicted basal area for each of the three stand types and dominant species along with the computed statistics.

Table 3. Performance Results of the LiDAR Dendrometric Map Sorted by Main Tree Species Composition.

Stand Type	Species	N	BA _{Obs} (m ² /ha)	BA _{Pred} (m ² /ha)	Bias (m ² /ha)	MAE (m ² /ha)	RMSE (m ² /ha)	R ²	r ² _{ER}
Shade-tolerant deciduous		2590	24.22	27.01	2.79	6.01	7.43	24.60%	10.30%
	Yellow birch	1430	24.5	26.9	2.41	5.96	7.43	24.30%	15.06%
	Red maple	337	24.0	24.7	0.68	5.52	6.94	36.30%	33.24%

Sugar maple	802	24.0	28.5	4.41	6.35	7.65	22.63%	-17.74%
Ash	21	17.0	17.4	0.43	4.84	5.66	39.99%	19.03%
Shade-intolerant deciduous	3217	23.77	24.43	0.66	5.30	6.83	46.87%	46.13%
White birch	2192	23.8	24.8	1.06	5.09	6.43	45.35%	43.44%
Aspen	1025	23.8	23.6	-0.18	5.74	7.61	49.83%	49.77%
Coniferous	6699	24.62	23.26	-1.36	5.27	7.38	53.38%	51.30%
White spruce	388	27.2	24.1	-3.04	6.10	7.87	50.89%	41.85%
Black spruce	2194	21.4	21.7	0.29	4.30	5.70	60.34%	60.24%
Larch	109	21.1	23.0	1.84	4.83	6.69	56.02%	51.70%
White pine	65	23.3	23.9	0.52	6.28	8.11	41.20%	40.89%
Jack pine	515	20.6	21.4	0.80	4.43	5.65	55.92%	54.01%
Balsam fir	3047	26.2	24.0	-2.20	5.39	7.41	51.38%	46.41%
White cedar	381	34.8	28.1	-6.65	10.08	14.26	27.06%	5.67%

Bias indicates that on average, the LDM underestimated the basal area for coniferous stands while it overestimated it for the deciduous stands. Shade-tolerant deciduous stands showed slightly higher Bias (2.79 m²/ha) and MAE (6.01 m²/ha) than the other two stand types. According to RMSE, the LDM is slightly better at predicting BA for shade-intolerant deciduous than for the other stand types. R² and the r²_{ER} showed that the independent variables in the LDM explained a greater proportion of the variability of the BA for the coniferous (53.88%) and shade-intolerant deciduous (46.78%) stands than for the shade-tolerant deciduous stands (26.58%).

Figure 2 illustrates the relationship between ground-observed data and the LiDAR Dendrometric Map. The closer the line approaches a 1:1 ratio, the more accurate the results are.

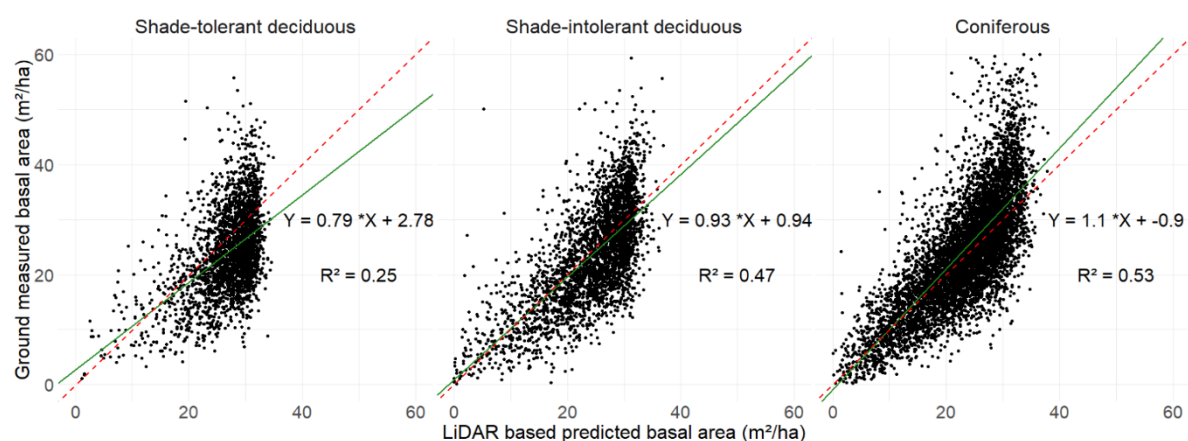


Figure 2. Correlation between ground sample plots measured basal area compared to predicted basal area from the LiDAR Dendrometric Map for the three stand types. The red points represent the 1:1 ratio, while the green line indicates the fitted linear relationship.

Figure 2 illustrates that shade-tolerant deciduous stands exhibit a correlation deviating from the 1:1 line, as indicated by the equation $Y = 0.79x + 2.78$ (a). Conversely, shade-intolerant and coniferous

stands demonstrate regressions closer to the 1:1 line, with equations of $Y = 0.93x + 0.94$ (b) and $Y = 1.1x - 0.9$ (c), respectively.

3.2. Independent Variables Analysis

Stepwise selection method

A total of 25 variables were retained, including localization data, species composition percentages, and meteorological variables. The selected variables are indicated with the index 1 in Table 2.

Boruta method

Figure 3 depicts the Boruta variable importance chart. Blue boxes represent the maximum, mean, and minimum Z score of a shadow attribute. Green and red boxplots represent Z score of confirmed and rejected variables respectively.

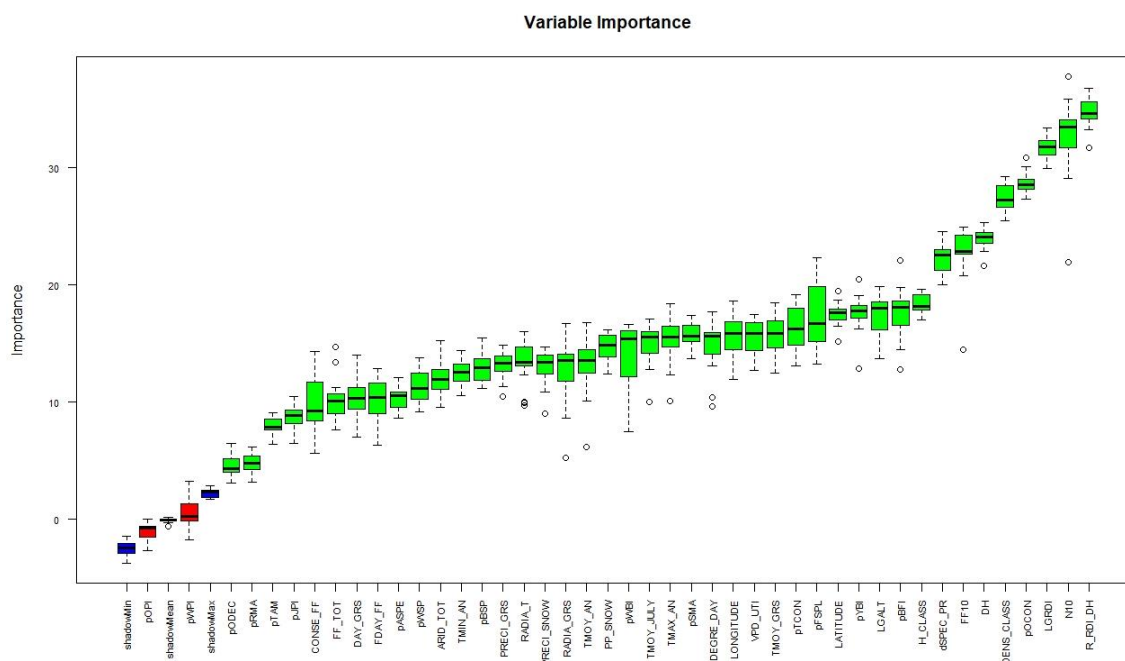


Figure 3. Distribution of variables according to their degree of importance as judged by the Boruta method.

Out of the 44 variables, two were rejected (pOPI and pWPI) and 42 were confirmed.

Lasso Method

Figure 4 and Table 4 illustrate the results obtained from the Lasso method.

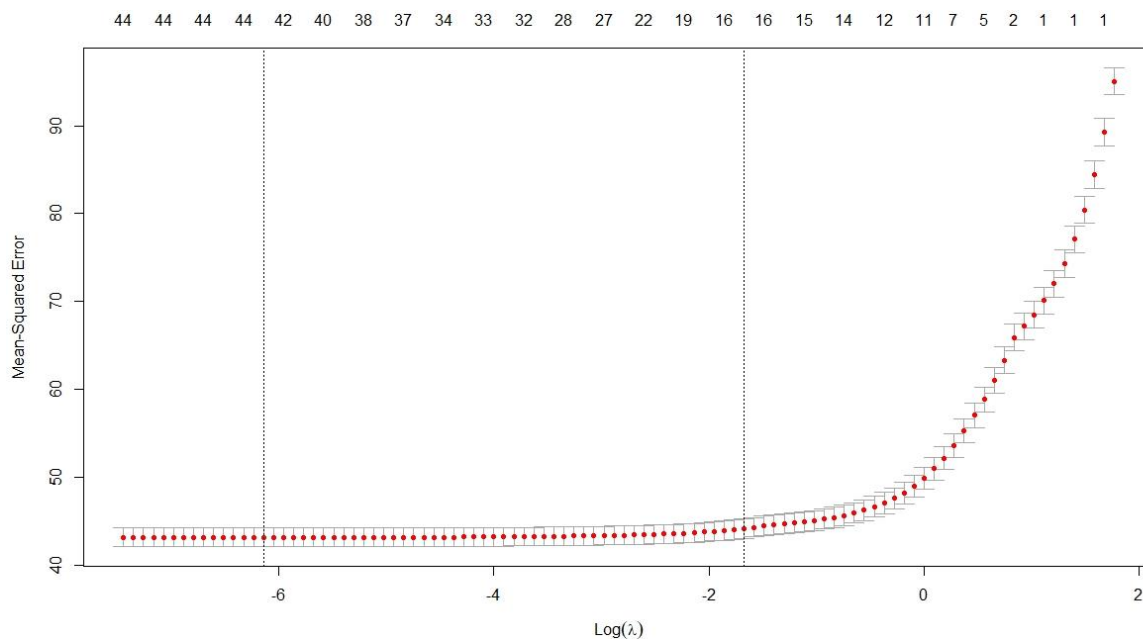


Figure 4. Lasso-generated graph illustrates the optimal number of variables (top axe) to incorporate into the model, between the available variables.

The two vertical lines correspond to λ_{\min} , which minimizes the cross-validated mean squared error (MSE) and retains more predictors, and λ_{1se} , within one standard error of the minimum, producing a more parsimonious model with fewer variables. According to these values, the optimal number of variables falls within the range of 16 to 43.

The method utilized ranks variables based on their coefficients, indicating the change in the response variable for a one-unit increase in the corresponding predictor variable, while holding all other variables constant. Because variables were set on the same scale, the absolute values of the coefficients directly correlate with their importance in the model. Table 4 shows the variable rank by coefficient value.

Table 4. Variables Best 25 Selected by Lasso Regression Ranked by Absolute Coefficient Value.

Rank	Variable	Coefficient	Abs (Coefficient)
1	R_RDI_DH	1037.24	1037.24
2	FF10	15.93	15.93
3	LGALT	-2.60	2.60
4	TMOY_GRS	-2.30	2.30
5	TMAX_AN	1.70	1.70
6	LATITUDE	1.01	1.01
7	LGRDI	-0.92	0.92
8	H_CLASS	-0.63	0.63
9	ARID_TOT	-0.63	0.63
10	TMOY_JULY	-0.38	0.38
11	TMIN_AN	-0.32	0.32
12	pOCON	0.24	0.24
13	DH	-0.12	0.12
14	LONGITUDE	0.12	0.12
15	dSPEC_PR	0.11	0.11
16	FF_TOT	-0.10	0.10
17	DAY_GRS	0,09	0,09
18	pWSP	0,07	0,07

19	CONSE_FF	0,06	0,06
20	pBFI	0,05	0,05
21	pWBI	0,04	0,04
22	pBSP	0,04	0,04
23	pASPE	0,03	0,03
24	PRECI_GRS	-0,03	0,03
25	FDAY_FF	-0,03	0,03

Table 4 presents the ranking of the 25 most influential variables identified by the Lasso regression; beyond this point, coefficient magnitudes gradually decreased from PP_SNOW (0.03) to N10 (-0.0003). The only variable with a coefficient of 0 is TMOY_AN, which means it was not considered to have any influence on the model and was excluded from the lasso selection.

Because each feature selection method produced a different number of selected variables, subsets were created to allow a fair comparison by using the same number of variables across methods. Since the stepwise method does not provide a variable ranking, subsets were generated only for Boruta and Lasso. Three rankings were considered: the 16 best variables according to Lasso λ_{1se} (within one standard error of the minimum), the 25 best variables corresponding to the number selected by the stepwise method, and the corresponding number of variables selected by each method. This procedure resulted in seven subsets, with the selected variables for each method presented in Table 2 along with their respective indices.

¹ Stepwise method, 25 selected variables

² Boruta method, best 16 variables

³ Boruta method, best 25 variables

⁴ Boruta method, 42 selected variables

⁵ Lasso method, λ_{1se} 16 variables

⁶ Lasso method, best 25 variables

⁷ Lasso method, λ_{min} 43 variables

3.3. Model Improvement with Three Approaches

3.3.1. Linear Parametric, Non-Linear Parametric and H20 AutoML Models

For comparison, all new modeling approaches were developed using the same set of variables selected through the stepwise backward selection method (Leboeuf et al., 2022).

Linear Model

Equation (7) displays the results of the optimized linear model.

$$BA_{corrected} = 3.64 + 959.03 * R_{RDI_{DH}} - 1.57 * TMOY_{GRS} + 1.39 * TMAX_{AN} + 1.31 * LATITUDE - 0.57 * H_{CLASS} + 0.16 * pOCON - 0.13 * ARID_{TOT} + B * factor(dSPEC_{PR}) + \sum_i \beta_i X_i \quad (7)$$

Where $BA_{corrected}$ is the prediction of the model, $factor(dSPEC_{PR})$ represents a coefficient B assigned to each species, which is only applied when that species is the dominant one in the stand. Other predictors are also present in the model but are not explicitly displayed in the formula due to their lower relative importance $\sum_i \beta_i X_i$. Results are shown in Table 5.

Non-Linear Model

Out of all the non-linear models tested, the quadratic regression model produced the best results. Equation (8) displays the optimized quadratic base on the same approach as the linear model.

$$BA_{corrected} = -24.92 + 834.14 * R_{RDI_{DH}} + 23.68 * R_{RDI_{DH}}^2 - 0.73 * H_{CLASS} + 0.14 * pOCON - 3.14 * TMOY_{GRS} + 0.55 * TMAX_{AN} + 0.40 * ARID_{TOT} + 2.02 * LATITUDE + factor(dSPEC_{PR}) + \sum_i \beta_i X_i \quad (8)$$

Where $BA_{corrected}$ is the prediction of the model, $factor(dSPEC_{PR})$ represents a coefficient B assigned to each species, which is only applied when that species is the dominant one in the stand.

Other predictors are also present in the model but are not explicitly displayed in the formula due to their lower relative importance $\sum_i \beta_i X_i$. Results are shown in Table 5.

H2O AutoML Model

The H2O AutoML function does not provide a detailed description of the best resulting model because its primary goal is to automate the model selection and tuning process rather than to generate interpretable model outputs. The best-performing model often corresponds to a Stacked Ensemble, which combines multiple base learners (e.g., GBM, DRF, DNN, GLM) to maximize predictive accuracy. This ensemble model integrates the predictions of several algorithms through a meta-learner, making the internal structure complex and not easily summarized in a single set of equations or coefficients.

The leaderboard indicates that the Stack Ensemble Model achieved the highest performance among all AutoML candidates, followed by GBM, and outperformed DRF, DNN, and GLM. Results for the best model are presented in Table 5 under H2O AutoML.

For all three modeling approaches, the best performance for shade-tolerant species was achieved by training models exclusively on data from shade-tolerant deciduous stands, using the same input variables.

3.3.2. Comparing Models Performance: New Approaches vs. Initial Model

Table 5. New Modelling approaches performances using the stepwise variable selection, compared to the LiDAR Dendrometric Map, for all commercial species and shade-tolerant species.

Approaches	All the Commercial Species					Shade-Tolerant Species				
	N	MAE (m ² /ha)	Bias	R ²	r ² _{ER}	N	MAE (m ² /ha)	Bias	R ²	r ² _{ER}
LiDAR Dendrometric Map	2560	5.47	-0.02	45.61%	45.59%	515	6.03	2.70	26.58%	13.37%
H2O AutoML	2560	4.78	-0.05	59.75%	59.64%	515	5.13	0.01	31.18%	31.01%
Linear	2560	5.03	0.06	55.93%	55.81%	515	5.21	0.03	30.99%	30.80%
Quadratic	2560	5.06	-0.30	54.62%	54.41%	515	5.20	0.03	30.60%	30.43%

Table 5 shows that the H2O AutoML method achieved the best overall performance across both groups. For all commercial species, r²_{ER} increased from 45.59 % (LDM) to 59.75 %, with MAE reduced from 5.47 m²/ha to 4.78 m²/ha and bias remained close to zero (-0.05). For shade-tolerant species, r²_{ER} improved from 13.37 % to 31.01 %, MAE decreased from 6.03 m²/ha to 5.13 m²/ha, and bias was also minimized from 2.70 m²/ha to 0.01 m²/ha. Linear and quadratic models also provided considerable improvements, with the linear model performing slightly better for both all commercial species and shade-tolerant species.

In order to further improve model performance, subsequent evaluations were conducted using the other variable subsets with the H2O AutoML, the best-performing modeling approach.

3.3.3. Testing the variable selection impacts on the best performing model (H2O AutoML)

To select the inputs for automatic machine learning, a comparison between metrics selection approaches was conducted and is presented in Table 6.

Table 6. Variable Selection Results: Comparison of performance for Boruta, Stepwise, and Lasso Methods with Different Number of Inputs.

Method	Number of Inputs	r ² _{ER}
Stepwise	25	59.64%
Boruta	16	59.3%
	25	59.65%

	42	59,48%
Lasso	16	58.54%
	25	59.56%
	43	59.69%

Based on the results presented in Table 6, using only 16 input variables is insufficient to maximize the explained variance, while increasing the number of inputs beyond 25 does not lead to a significant improvement in predictive accuracy. No significant differences in performance were observed between the feature selection methods. The stepwise method with 25 selected variables was therefore chosen to generate the final model. This configuration provides performance comparable to both the 25-variable Boruta subset and the 25-variable Lasso subset, with the added advantage that all 25 inputs were retained by the stepwise method, whereas Boruta and Lasso retain 42 and 43 variables, respectively, without providing additional predictive gain.

3.3.4. Testing Best Model with best variable selection

A graphical representation illustrating the correction's impact using AutoML on the distribution of data, comparing observed and predicted basal area in shade-tolerant deciduous, is depicted in Figure 5.

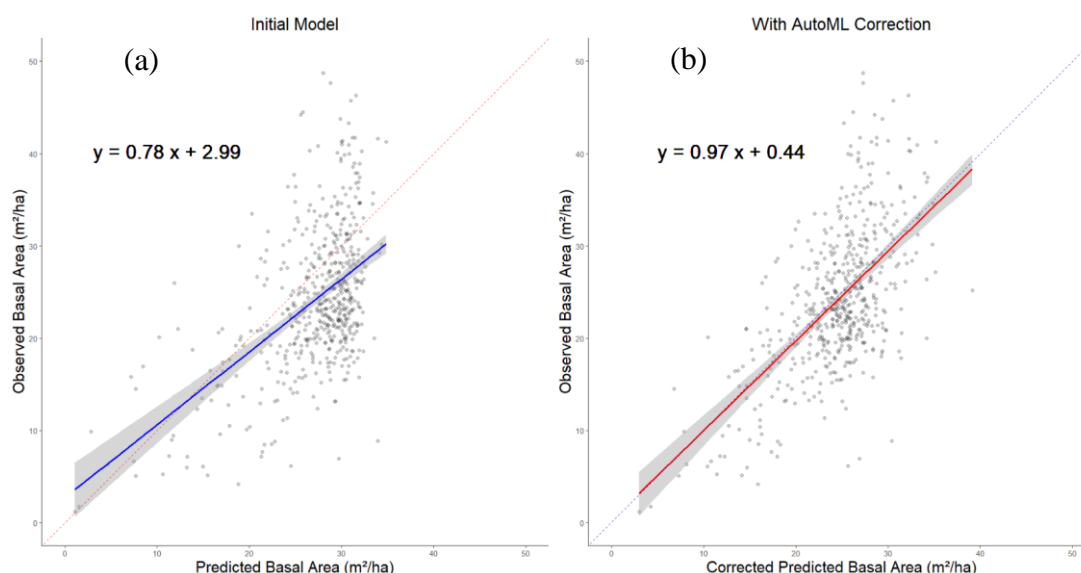


Figure 5. Observed stand basal area as a function of the stand basal area predicted by initial model (a) and the corrected one following H2O AutoML method (b) in shade-tolerant deciduous situations.

We observe that the results are centered on the 1:1 line in the Auto ML corrected model; however, the data points are still widely scattered around this line, highlighting the lack of accuracy.

Results for each stand type show the improvement from the initial model to the best model (Table 7). The test dataset is used for the comparison, which represents 20% of the total data, ensuring trustworthy results by using data that was not part of the model training.

Table 7. Performance Results of the Initial LiDAR Parametric Model and the best model issue with H2O AutoML Correction on the Test Data by Stand Type.

Stand type	N	Initial Model					Best Model				
		Bias	MAE	RMSE	R ²	r ² _{ER}	Bias	MAE	RMSE	R ²	r ² _{ER}
Shade-tolerant deciduous	515	2.70	6.03	7.45	26.58%	13.37%	0.01	5.13	6.64	31.18%	31.01%
Shade-intolerant deciduous	624	0.96	5.31	7.07	46.78%	45.51%	0.42	4.81	6.37	55.88%	55.69%
Coniferous	1421	-1.43	5.33	7.70	53.88%	51.56%	-0.28	4.64	6.43	66.89%	66.24%
All commercial	2560	-0.02	5.47	7.50	45.61%	45.59%	-0.05	4.78	6.46	59.75%	59.64%

Improvements were observed across all categories, with the most significant corrections seen in shade-tolerant deciduous stands in terms of bias, MAE and r²_{ER}. Coniferous species received the most significant correction in terms of RMSE and R². Despite the improvements, shade-tolerant deciduous stands still exhibit a lower performance indicator than the other stands in the initial model.

4. Discussion

This project has provided essential confirmation that ALS-based approaches, such as the LDM developed by the Quebec government's forest inventory program, still require refinement before they can be extended to temperate forests with complex stands, like those in the southern regions of Quebec. In particular, shade-tolerant deciduous stands have been identified as the main challenge requiring further research and development. The incorporation of geolocalized data independent of ground samples, together with the original metrics of the LDM, combined with the use of automated machine learning, has led to notable improvements. Although these enhancements have not yet yielded competitive precision for shade-tolerant deciduous stands, the approach developed in this project opens a new avenue for generating data that can be integrated with future advancements in the field. The approach adopted in this project lays the foundation for leveraging automated machine learning to fully exploit the potential of large datasets and incorporate new inputs derived from recent advances in forest remote sensing.

When assessing the LDM model's precision in predicting plot basal area across different stand types, it yielded R² values of 53.4% in coniferous stands, 45.2% in shade-intolerant deciduous stands, and only 24.6% in shade-tolerant deciduous stands. The low R² in the latter indicates a weak correlation between the predicted and observed basal area. Although shade-tolerant deciduous stands have the lowest R², the RMSE is higher in coniferous stands, this suggests that while the model captures overall trends better in conifers (higher R²), individual plot predictions for conifers are more variable, leading to larger absolute errors.

Further analysis revealed that dominant species composition significantly affects BA prediction accuracy, consistent with previous studies [23–25]. Stands dominated by sugar maple and yellow birch showed particularly poor performance, with R² values of 22.6% and 24.3%, and overestimations of 4.4 m²/ha and 2.4 m²/ha, respectively. The r²_{ER} of -17.7% for the sugar maple indicates that the LiDAR model performs worse than a trivial baseline, due to systematic overestimation of BA. This result highlights a clear performance issue of the model in this forest type. Stands dominated by white cedar also performed poorly, with an R² of 27.1% and an underestimation of 6.6 m²/ha. This bias contributes to the higher RMSE observed in conifer stands. These species-specific biases highlight the importance of considering species composition in the model.

Given that other studies have reported R² values exceeding 90% for basal area estimation [19,20], we examined the relationship between ground observations and predictions graphically to identify potential weaknesses in the model. The model appears resistant to predicting values above 35 m²/ha, suggesting a saturation effect driven by model constraints that limit overestimating basal area at the pixel level. While this limitation may be beneficial for large-scale estimations, it restricts accuracy at

the plot level, suggesting that the 5m × 5m data provided by the LDM is not well-suited for fine-scale predictions where higher variability in BA is typically observed.

To complement the species information already considered in the LDM, an alternative metric - the proportion of basal area covered by species- was tested alongside LDM inputs, geospatial climate variables, and location coordinates. Six variables were consistently identified as most important across all input selection methods: R_RDI_DH, H_CLASS, dSPEC_PR and pOCON, TMOY_GRS and LATITUDE. Among these, R_RDI_DH and H_CLASS confirmed their strong relationship with BA, whereas species-related metrics such as dSPEC_PR and pOCON emphasize the role of species composition and help explain the poor performance of the LDM in cedar-dominated stands [67]. Specifically, the former captures species-specific height-to-diameter ratios, while the latter reflects structural divergence in stands such as cedar. Climatic and geographic variables (TMOY_GRS, LATITUDE) further demonstrate how environmental conditions shape forest structure. Conversely, the rejection of redundant predictors such as TMOY_AN illustrates the model's ability to prioritize the most informative variables.

Integrating these variables into the three modeling approaches improved the results. For shade-tolerant species, the initial R^2 of 26.6% increased to 31.2% (AutoML), 31.0% (linear regression), and 30.6% (quadratic regression). For all commercial species, the initial R^2 of 45.6% improved to 59.8%, 55.9%, and 54.6% under AutoML, linear regression, and quadratic regression, respectively. Given its superior performance, the AutoML modeling approach was selected for the subsequent analyses. The results showed that using 16 input variables was insufficient to achieve optimal accuracy, whereas increasing the number of inputs beyond 25 did not yield substantial gains. No significant difference was observed between the model selection methods. In the comparison between AutoML (25 splitwise-selected variables) and the LDM model (Figure 5), the elimination of the prediction plateau at 35 m²/ha in the AutoML outputs suggests that this approach better captures stand dynamics and reduces systematic bias. The closer alignment of predicted versus observed values with the 1:1 line further indicates an improved capacity to generalize across conditions. However, despite the gains observed in shade-tolerant deciduous stands, model precision remains limited compared to the performance levels achieved in coniferous and shade-intolerant deciduous stands, underscoring the persistent challenges of modeling heterogeneous forest structures.

This research supports the MNRF's efforts to identify optimal LiDAR-based approaches for southern Quebec's forest inventory. It confirms that the LDM requires further refinement to handle structurally irregular, hardwood-dominated forests. The improvements made here offer guidance for future aerial LiDAR applications in complex forest environments.

To enhance BA prediction in shade-tolerant deciduous stands, new structural variables must be explored. Satellite imagery has shown mixed results [68–70], and canopy complexity remains a challenge [71]. The height-BA relationship may be insufficient, prompting investigation into variables like slope angle, stem size distribution [72], and vertical foliage profiles [73]. Promising LiDAR-derived metrics include mean height filtered by intensity thresholds and leaf area density variation [24,74]. With ongoing advances, species-level identification via aerial LiDAR may soon be feasible [75,76], offering transformative inputs for models highly sensitive to dominant species.

This study faced several challenges that may have influenced model performance and interpretation. A key limitation was the need to obtain a large volume of observed data corresponding to MNRF validation plots, where GPS imprecision and human interpretation bias may have introduced spatial and semantic errors. Despite the careful selection process of ground sample plots, it was not possible to verify whether unrecorded disturbances (e.g., harvesting, windthrow, etc.) occurred between the field data collection period and LiDAR acquisition, potentially degrading the quality of the training dataset. Moreover, some variables used in the model were designed for broad-scale applications and lacked the spatial specificity required for fine-scale plot-level predictions. For instance, tree species composition percentages are currently derived from photo-interpretation over sectors spanning several hectares. This coarse resolution may introduce discrepancies when applied to 400 m² plots, where species composition can vary significantly. The

available LiDAR dataset had already been transformed into predefined summary metrics, with a single value per plot. This constrained the capacity to extract new structural information directly from the point cloud. As a result, model improvements had to rely on external inputs independent of LiDAR, such as climate variables or geolocation. Incorporating custom LiDAR-derived variables—tailored to local canopy structure and vertical heterogeneity—could enhance the model's sensitivity to subtle structural variations. Another limitation stems from the use of automated machine learning (AutoML), particularly the Stacked Ensemble Model, which does not provide interpretable model parameters. While feature importance can be assessed, the internal logic of the ensemble remains opaque. This reinforces the importance of careful input selection, as model transparency is limited. Additionally, a common issue in machine learning regression models was observed: low values were generally overestimated, whereas high values were underestimated [77]. To address this, the Regression of Observed on Estimated values (ROE) method was applied to correct the bias [78], but the improvements were marginal. This suggests that further investigation is needed to determine whether H2O's AutoML framework exhibits structural bias in extreme values.

In summary, the study has advanced our understanding of the link between stand type and ALS-based model accuracy and highlighted potential improvements.

5. Conclusions

This study has demonstrated that shade-tolerant deciduous stands remain a major challenge for airborne LiDAR-based approaches. Evaluating the LDM across three distinct stand types confirmed that ALS-based models require improvement to accurately predict basal area in southern Quebec, particularly in structurally complex hardwood-dominated forests. The incorporation of new variables through both parametric models and automated machine learning led to notable improvements, especially when using the H2O AutoML framework. However, the objective of achieving precision levels for shade-tolerant deciduous stands comparable to those observed in shade-intolerant deciduous and coniferous stands has not yet been met. Key input variables were identified to improve the model, including species-specific basal area proportions, geolocation coordinates, and meteorological metrics. These findings underscore the importance of input selection and structural sensitivity in LiDAR-based modeling. Future research should prioritize enhancing spatial alignment between field plots and LiDAR data, improving species composition inputs through higher-resolution mapping, and investigating bias-correction techniques as well as model interpretability in AutoML frameworks. These refinements are essential to strengthen the robustness, precision, and scalability of LiDAR-based forest inventory models, particularly in heterogeneous and species-diverse environments.

References

1. Ministère des Ressources naturelles et des Forêts *Ressources et industrie forestière du Québec - Portrait statistique 2023*; Direction du développement et de l'innovation de l'industrie; 29th ed.; Gouvernement du Québec, 2024; ISBN 978-2-555-00005-6.
2. FAO *Global Forest Resources Assessment 2020. Main Report*; FAO: Rome, Italy, 2020; ISBN 978-92-5-132974-0.
3. Sun, H.G.; Zhang, J.G.; Duan, A.G.; He, C.Y. A Review of Stand Basal Area Growth Models. *For. Stud. China* **2007**, *9*, 85–94, doi:10.1007/S11632-007-0014-2.
4. Vogt, J.T.; Koch, F.H. The Evolving Role of Forest Inventory and Analysis Data in Invasive Insect Research. *Am. Entomol.* **2016**, *62*, 46–58, doi:10.1093/AE/TMV072.
5. Comeau, P.G.; Wang, J.R.; Letchford, T. Influences of Paper Birch Competition on Growth of Understory White Spruce and Subalpine Fir Following Spacing. *Can. J. For. Res.* **2003**, *33*, 1962–1973, doi:10.1139/X03-117.
6. Sharma, M.; Oderwald, R.G.; Amateis, R.L. A Consistent System of Equations for Tree and Stand Volume. *For. Ecol. Manag.* **2002**, *165*, 183–191, doi:10.1016/S0378-1127(01)00616-8.
7. Fang, Z.; Bailey, R.L.; Shiver, B.D. A Multivariate Simultaneous Prediction System for Stand Growth and Yield with Fixed and Random Effects. *For. Sci.* **2001**, *47*, 550–562, doi:10.1093/FORESTSCIENCE/47.4.550.

8. Nyland, R.D.; Ray, D.G.; Yanai, R.D.; Briggs, R.D.; Zhang, L.; Cymbala, R.J.; Twery, M.J. Early Cohort Development Following Even-Aged Reproduction Method Cuttings in New York Northern Hardwoods. *Can. J. For. Res.* **2000**, *30*, 67–75, doi:10.1139/X99-190.
9. Pothier, D.; Savard, F. *Actualisation Des Tables de Production Pour Les Principales Espèces Forestières Québécoises*; Ministère des Ressources naturelles; gouvernement du Québec: Québec, 1998; ISBN 2-551-18982-9.
10. Eerikäinen, K. A Site Dependent Simultaneous Growth Projection Model for Pinus Kesiya Plantations in Zambia and Zimbabwe. *For. Sci.* **2002**, *48*, 518–529, doi:10.1093/FORRESTSCIENCE/48.3.518.
11. Pearse, G.D.; Dash, J.P.; Persson, H.J.; Watt, M.S. Comparison of High-Density LiDAR and Satellite Photogrammetry for Forest Inventory. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 257–267, doi:10.1016/J.ISPRSJPRS.2018.06.006.
12. Naeset, E. Area-Based Inventory in Norway—From Innovation to an Operational Reality. In *Forestry Applications of Airborne Laser Scanning; Managing Forest Ecosystems*; Springer: Dordrecht, 2014; Vol. 27, pp. 215–240 ISBN 978-94-017-8663-8.
13. Ministère des Forêts, de la Faune et des Parcs *Méthodologie Des Compilations Forestières Du 4e Inventaire Écoforestier Du Québec Méridional : Cas Particulier Des Estimations k-NN*; Direction des inventaires forestiers; gouvernement du Québec, 2017; ISBN 978-2-550-78110-3.
14. Ministère des Ressources naturelles et des Forêts Inventaire écoforestier du Québec méridional Available online: <https://www.quebec.ca/agriculture-environnement-et-ressources-naturelles/forets/recherche-connaissances/inventaire-forestier/types/quebec-meridional> (accessed on 29 January 2025).
15. Apostol, B.; Chivulescu, S.; Ciceu, A.; Petrila, M.; Pascu, I.S.; Apostol, E.N.; Leca, Ş.; Lorent, A.; Tănase, M.; Badea, O. Data Collection Methods for Forest Inventory: A Comparison between an Integrated Conventional Equipment and Terrestrial Laser Scanning. *Ann. For. Res.* **2018**, *61*, 189–189, doi:10.15287/afr.2018.1189.
16. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273, doi:10.1093/JOF/109.5.267.
17. Leboeuf, A.; Riopel, M.; Munger, D.; Fradette, M.-S.; Bégin, J. Modeling Merchantable Wood Volume Using Airborne LiDAR Metrics and Historical Forest Inventory Plots at a Provincial Scale. *Forests* **2022**, *13*, 985, doi:10.3390/F13070985.
18. Ministère des Ressources naturelles et des Forêts *Carte Dendrométrique LiDAR : Méthode et Utilisation - 2e Édition*; Direction des inventaires forestiers; 3rd ed.; gouvernement du Québec, 2025; ISBN 978-2-555-01344-5.
19. Thomas, V.; Treitz, P.; McCaughey, J.H.; Morrison, I. Mapping Stand-Level Forest Biophysical Variables for a Mixedwood Boreal Forest Using Lidar: An Examination of Scanning Density. *Can. J. For. Res.* **2011**, *36*, 34–47, doi:10.1139/X05-230.
20. Silva, C.A.; Klauber, C.; Hudak, A.T.; Vierling, L.A.; Fennema, S.J.; Corte, A.P.D. Modeling and Mapping Basal Area of Pinus Taeda L. Plantation Using Airborne LiDAR Data. *An. Acad. Bras. Ciênc.* **2017**, *89*, 1895–1905, doi:10.1590/0001-3765201720160324.
21. Chen, Q.; Gong, P.; Baldocchi, D.; Tian, Y.Q. Estimating Basal Area and Stem Volume for Individual Trees from Lidar Data. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1355–1365, doi:10.14358/PERS.73.12.1355.
22. Woods, M.; Pitt, D.; Penner, M.; Lim, K.; Nesbitt, D.; Etheridge, D.; Treitz, P. Operational Implementation of a LiDAR Inventory in Boreal Ontario. *For. Chron.* **2011**, *87*, 512–528, doi:10.5558/tfc2011-050.
23. Spriggs, R.A.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A Critique of General Allometry-Inspired Models for Estimating Forest Carbon Density from Airborne LiDAR. *PLOS ONE* **2019**, *14*, doi:10.1371/JOURNAL.PONE.0215238.
24. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.P. Generalizing Predictive Models of Forest Inventory Attributes Using an Area-Based Approach with Airborne LiDAR Data. *Remote Sens. Environ.* **2015**, *156*, 322–334, doi:10.1016/J.RSE.2014.10.004.
25. Vandendaele, B.; Fournier, R.A.; Vepakomma, U.; Pelletier, G.; Lejeune, P.; Martin-ducup, O. Estimation of Northern Hardwood Forest Inventory Attributes Using Uav Laser Scanning (Uls): Transferability of Laser

- Scanning Methods and Comparison of Automated Approaches at the Tree- and Stand-level. *Remote Sens.* **2021**, *13*, doi:10.3390/rs13142796.
26. Allouis, T.; Durrieu, S.; Vega, C.; Couteron, P. Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees From LiDAR Data: Contribution of Full-Waveform Signals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 924–934, doi:10.1109/JSTARS.2012.2211863.
 27. Sharma, M. Comparing Height-Diameter Relationships of Boreal Tree Species Grown in Plantations and Natural Stands. *For. Sci.* **2016**, *62*, 70–77, doi:10.5849/forsci.14-232.
 28. Wang, X.; Fang, J.; Tang, Z.; Zhu, B. Climatic Control of Primary Forest Structure and DBH–Height Allometry in Northeast China. *For. Ecol. Manag.* **2006**, *234*, 264–274, doi:10.1016/j.foreco.2006.07.007.
 29. Zhang, Z.; Papaik, M.J.; Wang, X.; Hao, Z.; Ye, J.; Lin, F.; Yuan, Z. The Effect of Tree Size, Neighborhood Competition and Environment on Tree Growth in an Old-Growth Temperate Forest. *J. Plant Ecol.* **2017**, *10*, 970–980, doi:10.1093/jpe/rtw126.
 30. Sylvain, J.-D.; Drolet, G.; Kiriazis, N.; Thiffault, É.; Anctil, F. Assessing the Hydroclimatic Sensitivity of Tree Species in Northeastern America through Spatiotemporal Modelling of Annual Tree Growth. *Agric. For. Meteorol.* **2024**, *355*, doi:10.1016/j.agrformet.2024.110079.
 31. Régnière, Jacques.; Cooke, Barry.; Bergeron, Vincent. *BioSIM: A Computer-Based Decision Support Tool for Seasonal Planning of Pest Management Activities. User's Manual*; Natural Resources Canada; Natural Resources Canada, Laurentian Forestry Centre: Sainte-Foy, Québec, 1995; ISBN 978-1-100-23464-9.
 32. Kim, J.; Popescu, S.C.; Lopez, R.R.; Wu, X.B.; Silvy, N.J. Vegetation Mapping of No Name Key, Florida Using Lidar and Multispectral Remote Sensing. *Int. J. Remote Sens.* **2020**, *41*, 9469–9506, doi:10.1080/01431161.2020.1800125.
 33. Zhao, Q.; Yu, S.; Zhao, F.; Tian, L.; Zhao, Z. Comparison of Machine Learning Algorithms for Forest Parameter Estimations and Application for Forest Quality Assessments. *For. Ecol. Manag.* **2019**, *434*, 224–234, doi:10.1016/J.FORECO.2018.12.019.
 34. Liu, Z.; Peng, C.; Work, T.; Candau, J.N.; Desrochers, A.; Kneeshaw, D. Application of Machine-Learning Methods in Forest Ecology: Recent Progress and Future Challenges. *Environ. Rev.* **2018**, *26*, 339–350, doi:10.1139/ER-2018-0034.
 35. Naik, P.; Dalponte, M.; Bruzzone, L. Automated Machine Learning Driven Stacked Ensemble Modeling for Forest Aboveground Biomass Prediction Using Multitemporal Sentinel-2 Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3442–3454, doi:10.1109/JSTARS.2022.3232583.
 36. LeDell, E.; Poirier, S. *H2O AutoML: Scalable Automatic Machine Learning*; 7th ICML Workshop on Automated Machine Learning; 2020;
 37. Hutter, F.; Kotthoff, L.; Vanschoren, J. *Automated Machine Learning: Methods, Systems, Challenges*; Springer: Cham, 2019; ISBN 978-3-030-05318-5.
 38. Ministère des Ressources naturelles et des Forêts *LISEZ MOI - Description d'un Jeu de Données : Placettes-Échantillons Permanente*; Direction des inventaires forestiers: Québec, 2022; p. 23;.
 39. Ministère des Ressources naturelles et des Forêts *LISEZ-MOI - Description d'un jeu de données : Placettes-échantillons temporaire du quatrième inventaire*; Direction des inventaires forestiers: Québec, 2024; p. 28;.
 40. Ministère des Ressources naturelles et des Forêts *LISEZ MOI - Description d'un Jeu de Données : Placettes-Échantillons Temporaires Du Cinquième Inventaire*; Direction des inventaires forestiers: Québec, 2024; p. 22;.
 41. Vanclay, J.K.; Skovsgaard, J.P. Evaluating Forest Growth Models. *Ecol. Model.* **1997**, *98*, 1–12, doi:10.1016/S0304-3800(96)01932-1.
 42. Pretzsch, H. Forest Dynamics, Growth and Yield: From Measurement to Model. *For. Dyn. Growth Yield Meas. Model* **2010**, 1–664, doi:10.1007/978-3-540-88307-4/COVER.
 43. Stage, A.R. How Forest Models Are Connected to Reality: Evaluation Criteria for Their Use in Decision Support. *Can. J. For. Res.* **2003**, *33*, doi:10.1139/X02-203.
 44. Ministère des Ressources naturelles et des Forêts *Classification écologique du territoire québécois*; Direction des inventaires forestiers; gouvernement du Québec, 2021; ISBN 978-2-550-89426-1.
 45. Ministère des Ressources naturelles du Québec *Rapport de classification écologique - Sapinière à Bouleau jaune de l'est*; Direction des inventaires forestiers; Gouvernement du Québec: Québec, 1999; ISBN 2-551-34331-X.

46. Ministère des Ressources naturelles du Québec *Rapport de classification écologique - Sapinière à Bouleau jaune de l'ouest*; Direction des inventaires forestiers; Gouvernement du Québec: Québec, 1998; ISBN 2-551-19035-5.
47. Guillemette, F.; Bédard, S. *Sylviculture des peuplements à dominance de feuillus nobles au Québec - hors-série*; Ministère des ressources naturelles et de la faune, Direction de la recherche forestière: Québec, 2006; p. 112;.
48. Ministère des Forêt, de la Faune et des Parcs *Norme d'inventaire écoforestier : Placette-échantillons permanentes - 4e inventaire*; Direction des inventaires forestiers; Gouvernement du Québec, secteur des forêts: Québec, 2016; ISBN 978-2-550-72875-7.
49. Ministère des Ressources naturelles et des Forêts *Norme d'inventaire écoforestier : Placettes-échantillons temporaires - 5e inventaire*; Direction des inventaires forestiers; Gouvernement du Québec, secteur des forêts: Québec, 2025; ISBN 978-2-555-01399-5.
50. Ministère des Ressources naturelles et des Forêts *Norme d'inventaire écoforestier : Placette-échantillons permanentes - 5e inventaire*; Direction des inventaires forestiers; Gouvernement du Québec, secteur des forêts: Québec, 2025; ISBN 978-2-555-01400-8.
51. Ministère des Forêt, de la Faune et des Parcs *Norme d'inventaire écoforestier : Placettes-échantillons temporaires - 4e inventaire*; Direction des inventaires forestiers; Gouvernement du Québec, secteur des forêts: Québec, 2016; ISBN 978-2-550-72899-3.
52. Ministère des Ressources naturelles et des Forêts *Cartographie du cinquième inventaire écoforestier du Québec méridional - Méthodes et données associées*; Direction des inventaires forestiers; 4th ed.; Gouvernement du Québec, secteur des forêts, 2025; ISBN 978-2-550-97714-8.
53. Levick, S.R.; Hessenmöller, D.; Schulze, E.D. Scaling Wood Volume Estimates from Inventory Plots to Landscapes with Airborne LiDAR in Temperate Deciduous Forest. *Carbon Balance Manag.* **2016**, *11*, 1–14, doi:10.1186/S13021-016-0048-7/FIGURES/8.
54. Ministère des Ressources naturelles et des Forêts *Guide d'utilisation de La Carte Écoforestière Originale et Résultats d'inventaire Courants*; Direction des inventaires forestiers; 8th ed.; gouvernement du Québec, 2024; ISBN 978-2-555-00104-6.
55. Schneider, R.; Fortin, M.; Saucier, J.-P. *Équations de défilement en forêt naturelle pour les principales essences commerciales du Québec*; Mémoire de recherche forestière; Ministère des Ressources naturelles et de la Faune, Direction de la recherche forestière: Québec, 2013; ISBN 978-2-550-66646-2.
56. Ministère des Ressources naturelles et des Forêts *Cartographie du cinquième inventaire écoforestier du Québec méridional - méthode et données associées*; Direction des inventaires forestiers; 4th ed.; gouvernement du Québec: Québec, Canada, 2025; ISBN 978-2-550-97714-8.
57. Pospisil, D.A.; Bair, W. The Unbiased Estimation of the Fraction of Variance Explained by a Model. *PLoS Comput. Biol.* **2021**, *17*, doi:10.1371/JOURNAL.PCBI.1009212.
58. Theng, D.; Bhojar, K.K. Feature Selection Techniques for Machine Learning: A Survey of More than Two Decades of Research. *Knowl. Inf. Syst.* **2024**, *66*, 1575–1637, doi:10.1007/s10115-023-02010-5.
59. Whittingham, M.J.; Stephens, P.A.; Bradbury, R.B.; Freckleton, R.P. Why Do We Still Use Stepwise Modelling in Ecology and Behaviour? *J. Anim. Ecol.* **2006**, *75*, 1182–1189, doi:10.1111/J.1365-2656.2006.01141.X.
60. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13, doi:10.18637/JSS.V036.I11.
61. Muthukrishnan, R.; Rohini, R. LASSO: A Feature Selection Technique in Predictive Modeling for Machine Learning. *2016 IEEE Int. Conf. Adv. Comput. Appl. ICACA 2016* **2017**, 18–20, doi:10.1109/ICACA.2016.7887916.
62. Ranstam, J.; Cook, J.A. LASSO Regression. *Br. J. Surg.* **2018**, *105*, 1348–1348, doi:10.1002/BJS.10895.
63. Seber, G.A.F.; Wild, C.J. Nonlinear Regression. **1989**, doi:10.1002/0471725315.
64. Goretzko, D.; Bühner, M. One Model to Rule Them All? Using Machine Learning Algorithms to Determine the Number of Factors in Exploratory Factor Analysis. *Psychol. Methods* **2020**, *25*, 776–786, doi:10.1037/MET0000262.
65. Shinde, P.P.; Shah, S. A Review of Machine Learning and Deep Learning Applications. *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018* **2018**, doi:10.1109/ICCUBEA.2018.8697857.

66. H2O H2O AutoML: Automatic Machine Learning Available online: <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html> (accessed on 11 February 2025).
67. Rijal, B.; Weiskittel, A.R.; Kershaw, J.A. Development of Regional Height to Diameter Equations for 15 Tree Species in the North American Acadian Region. *For. Int. J. For. Res.* **2012**, *85*, 379–390, doi:10.1093/forestry/cps036.
68. Brown, S.; Narine, L.L.; Gilbert, J. Using Airborne Lidar, Multispectral Imagery, and Field Inventory Data to Estimate Basal Area, Volume, and Aboveground Biomass in Heterogeneous Mixed Species Forests: A Case Study in Southern Alabama. *Remote Sens.* **2022**, *14*, doi:10.3390/rs14112708.
69. Lahssini, K.; Teste, F.; Dayal, K.R.; Durrieu, S.; Ienco, D.; Monnet, J.-M. Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2022–2022, doi:10.1109/JSTARS.2022.3175609i.
70. Monnet, J.M.; Chirouze, É.; Mermin, É. Estimation de Paramètres Forestiers Par Données LiDAR Aéroporté et Imagerie Satellitaire Rapideye : Étude de Sensibilité. *Rev. Française Photogramm. Teledetection* **2015**, 71–79, doi:10.52638/RFPT.2015.544.
71. Hastings, J.H.; Ollinger, S.V.; Ouimette, A.P.; Sanders-Demott, R.; Palace, M.W.; Ducey, M.J.; Sullivan, F.B.; Basler, D.; Orwig, D.A. Tree Species Traits Determine the Success of LiDAR-Based Crown Mapping in a Mixed Temperate Forest. *Remote Sens.* **2020**, *12*, 309–309, doi:10.3390/rs12020309.
72. Leclère, L.; Lejeune, P.; Bolyn, C.; Latte, N. Estimating Species-Specific Stem Size Distributions of Uneven-Aged Mixed Deciduous Forests Using ALS Data and Neural Networks. *Remote Sens.* **2022**, *14*, doi:10.3390/rs14061362.
73. Racine, E.B.; Coops, N.C.; Bégin, J.; Myllymäki, M. Tree Species, Crown Cover, and Age as Determinants of the Vertical Distribution of Airborne LiDAR Returns. *Trees - Struct. Funct.* **2021**, *35*, 1845–1861, doi:10.1007/S00468-021-02155-2.
74. Lim, K.; Treitz, P.; Baldwin, K.; Morrison, I.; Optech, J. LiDAR Remote Sensing of Biophysical Properties of Tolerant Northern Hardwood Forests. *Can. J. Remote Sens.* **2003**, *29*, doi:10.5589/m03-025.
75. Sylvain, J.-D.; Drolet, G.; Thiffault, É.; Ancil, F. High-Resolution Mapping of Tree Species and Associated Uncertainty by Combining Aerial Remote Sensing Data and Convolutional Neural Networks Ensemble. *Int. J. Appl. Earth Obs. Geoinformation* **2024**, *131*, 103960, doi:10.1016/j.jag.2024.103960.
76. Yu, X.; Hyyppä, J.; Litkey, P.; Kaartinen, H.; Vastaranta, M.; Holopainen, M.; Waser, L.T.; Wynne, R.H.; Thenkabail, P.S. Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning. *Remote Sens.* **2017**, *9*, 108–108, doi:10.3390/rs9020108.
77. Zhang, G.; Lu, Y. Bias-Corrected Random Forests in Regression. *J. Appl. Stat.* **2012**, *39*, 151–160, doi:10.1080/02664763.2011.578621.
78. Belitz, K.; Stackelberg, P.E. Evaluation of Six Methods for Correcting Bias in Estimates from Ensemble Tree Machine Learning Regression Models. *Environ. Model. Softw.* **2021**, *139*, 105006–105006, doi:10.1016/J.ENVSOF.2021.105006.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.