

## Article

# Estimating Stem Diameter Distributions with Airborne Laser Scanning Metrics and Derived Canopy Surface Texture Metrics

Xavier Gallagher-Duval <sup>1</sup>, Olivier R. van Lier <sup>2,\*</sup>  and Richard A. Fournier <sup>1</sup>

<sup>1</sup> Department of Applied Geomatics, Centre d'Applications et de Recherche en Télédétection (CARTEL), Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada

<sup>2</sup> Canadian Forest Service—Canadian Wood Fibre Centre, Natural Resources Canada, Corner Brook, NL A2H 5G4, Canada

\* Correspondence: olivier.vanlier@nrcan-rncan.gc.ca

**Abstract:** This study aimed to determine the optimal approach for estimating stem diameter distributions (SDD) from airborne laser scanning (ALS) data using point cloud metrics ( $M_{als}$ ), a canopy height model (CHM) texture metrics ( $M_{tex}$ ), and a combination thereof ( $M_{comb}$ ). We developed area-based models (i) to classify SDD modality and (ii) predict SDD function parameters, which we tested for 5 modelling techniques. Our results demonstrated little variability in the performance of SDD modality classification models (mean overall accuracy: 72%; SD: 2%). Our best SDD function parameter models were generally fitted with  $M_{comb}$ , with  $R^2$  improvements up to 0.25. We found the variable Correlation, originating from  $M_{tex}$ , to be the most important predictor within  $M_{comb}$ . Trends in the performance of the predictor groups were mostly consistent across the modelling techniques within each parameter. Using an Error Index (EI), we determined that differentiating modality prior to estimating SDD improved the accuracy of estimates for bimodal plots (~12% decrease in EI), which was trivially not the case for unimodal plots (<1% increase in EI). We concluded that (i) CHM texture metrics can be used to improve the estimate of SDD parameters and that (ii) differentiating for modality prior to estimating SDD is especially beneficial in stands with bimodal SDD.

**Keywords:** airborne laser scanning; texture; stem diameter distributions; forest inventory; boreal forest



**Citation:** Gallagher-Duval, X.; van Lier, O.R.; Fournier, R.A. Estimating Stem Diameter Distributions with Airborne Laser Scanning Metrics and Derived Canopy Surface Texture Metrics. *Forests* **2023**, *14*, 287. <https://doi.org/10.3390/f14020287>

Academic Editors: Jianping Wu, Zhongbing Chang and Xin Xiong

Received: 10 January 2023  
Revised: 27 January 2023  
Accepted: 31 January 2023  
Published: 2 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the last decade, much effort has been devoted to modelling and mapping forest inventory attributes from airborne laser scanning data (ALS) to the point where these data are being used operationally over large, continuous areas internationally (e.g., [1–3]). ALS can provide precise and reliable predictions of many stand-mean values of biophysical attributes (e.g., biomass, volume, height, and DBH [4–6]), as well as distributions thereof (e.g., stem diameter, height, and volume distributions [7–9]). Stem diameter is the most frequently modelled distribution found in the literature (e.g., [10–17]) as it provides insights on stand structure, the basis for understanding the stand's ecological and economic value. Stem diameter distributions can be used to describe forest dynamics [18], carbon stock, biomass, and wood volumes [19,20], and are known to be correlated with species diversity [21,22]. This information is an important aid for forest managers, who are planning silvicultural strategies [23] and assessing the commercial value of given stands.

Numerous functions have been described in the literature to fit stem diameter distributions (SDD). The early works of Bailey and Dell (1973) [24] proposed the Weibull probability density function (PDF) as a diameter distribution model. Since then, many studies have evaluated the effectiveness of other statistical functions. Hafley and Schreuder (1977) [25] found Johnson's  $S_B$  function to outperform the Weibull in terms of quality of fit of the distributions. Similarly, and more recently, Gorgoso-Varela et al. (2021) [26] compared the Weibull (2P and 3P), Johnson's  $S_B$ , beta, generalized beta, and gamma-2P functions, and although the Weibull (2P and 3P) and Johnson's  $S_B$  yielded the poorest fits

to the data, they concluded that all six assessed PDF produced reasonable results. Similarly, Consenza et al. (2019) [27] demonstrated that Johnson's  $S_B$  presented comparable performances to the Weibull for two forest types: slightly better for a *Eucalyptus globulus* plantation and slightly worse for a *Pinus radiata* plantation. Although these studies have demonstrated that the Weibull is not always the best-fitting PDF, it has been most widely used in forestry (e.g., [7,14,28,29]), namely due to the function's flexibility in shape and relative simplicity of mathematical implementation [24]. The Weibull, however, is better suited to represent homogenous stands, given that it contains only one mode. Recent studies have demonstrated improvements in SDD predictions by fitting the bimodal SDD of heterogeneous stands to two PDF in structurally diverse forests [28,30]. The accuracy in representing SDD is therefore inevitably dependent on the forest structure being assessed.

Parametric and non-parametric approaches have been used to model SDD regardless of the distribution's modality (e.g., [8,19,20,31]). As PDFs are multivariate, it is often necessary to use multiple models developed with methods that can handle high-dimensional space [32]. Although many approaches to predict SDD have been proposed in recent decades, the current trends have been based on the PDF parameter prediction [28,33–35] and recovery methods [12,36–38]. In the 1990s, studies found  $k$ -nearest neighbor ( $k$ -NN) regression to be more accurate and flexible than methods based on parametric distributions in predicting stand-level diameter distributions [39,40]. With the advent of ALS,  $k$ -NN approaches were implemented using the area-based approach [41] to produce sub-stand-level diameter predictions with similar results [42–44]. Although  $k$ -NN estimation has long been used to predict SDD, the large amounts of training data required can limit its application. Many other approaches have also been proposed. For example, Kangas and Maltamo (2000) [45] suggested a model that first predicted diameters at 12 percentiles, then the basal area diameter distribution was interpolated using a rational spline. Liu et al. (2009) [46] later assessed the percentile-based approach [47] against five other methods in predicting parameters for SDD represented by a Weibull function for white spruce plantations and found the percentile-based parameter recovery method performed best. In another study, Bollandsås and Naesset (2007) [19] proposed to use partial least squares regression to effectively predict diameters at percentiles of basal area in uneven-sized Norway spruce stands. In a most recent study, Strunk and McGaughey (2023) [36] compared post-stratification, ordinary least squares regression,  $k$ -NN, and random forest to predict diameter class-specific volumes and found that random forest produced overall better results for a managed southern white pine forest. The complex distributions associated with more heterogeneous forest structures are, however, often better represented within a Finite Mixture Model (FMM) by combining two or more PDFs [28,33,34,48]. For example, Mulverhill et al. (2018) [34] developed maximum likelihood estimation models for both unimodal and bimodal SDD to appropriately characterize the simple and irregular distributions found in stands of boreal mixedwood forests (Canada). Though the estimation approaches continue to evolve, no consensus on a singularly favoured modelling method has yet been established.

Spatially explicit and exhaustive characterizations of SDD are made possible with remote sensing. Tarp-Johansen (2002) [49] used a 3D model and digital aerial photographs to estimate stem diameters for monospecific English oak (*Quercus robur* L.) stands in Denmark. With the development of ALS, Gobakken and Næsset (2004) [50] used various ALS height metrics to estimate Weibull parameters accurately ( $R^2$  ranging between 0.6–0.9 with an RMSE of 0.15) to predict SDD for the boreal forest in southeast Norway. Multi-source remote sensing data can also be combined to improve prediction accuracy. Peuhkurinen et al. (2018) [30] combined ALS data and SPOT5 imagery to make accurate predictions (Reynold's Error Index for all plots ranged from 17.99 to 122.94) of SDD for coniferous boreal forests of Russia's Perm Region with the non-parametric  $k$ -Most Similar Neighbour method. In addition to height metrics, intensity metrics can be derived from ALS data, thereby providing indications of the strength of backscattered energy. Shang et al. (2017) [51] used ALS height and intensity metrics to predict SDD for a hardwood forest in

Ontario, Canada. They found that combining intensity and height metrics improved the model's performance beyond employing either height-only or intensity-only metrics.

Texture metrics that are derived from remote sensing can provide additional information regarding canopy structure that is independent of spectral features regarding spatial variations [52]. Haralick's Grey-Level Co-occurrence Matrix (GLCM) [53] is one common approach to calculating texture features from a given raster surface. GLCM uses second-order statistics, which are defined as the probability of observing a certain pair of pixel values within a predefined angle and observation window size [54]. Studies have demonstrated that texture metrics derived from optical data can be used successfully to predict forest attributes for a range of forest types (e.g., for boreal and Great Lakes—St. Lawrence forests of Canada [55]; temperate forests of Ontario, Canada [8]; boreal forests of Finland [56]). Dube and Mutanga (2015) [57] compared aboveground biomass models for three medium-density plantation forest species in South Africa that were derived from Landsat-8 spectral bands, spectral band ratios, vegetation indices, texture bands, and texture band ratios. The study demonstrated that models developed from multiple texture band ratios yielded the highest  $R^2$ . Several studies have incorporated canopy height model (CHM)-derived texture metrics in predicting forest attributes. Ozdemir and Donoghue (2013) [58] used CHM-derived texture metrics to explain tree diversity for a broad range of stand types (pure conifer, mixed conifer, pure deciduous, mixed deciduous, and conifer, different age classes) and found that the combination of ALS metrics with texture metrics explained up to 85% of the measured tree height diversity. Niemi and Vauhkonen (2016) [59] demonstrated that using texture metrics improved prediction of total stem volume and basal area over models that were developed solely from ALS metrics for boreal forests in southern Finland. Similarly, van Ewijk et al. (2019) [55] found that combining ALS, CHM texture, and intensity metrics improved  $R^2$  by 0.19 for the prediction of stem density when compared with models that were developed solely with ALS metrics.

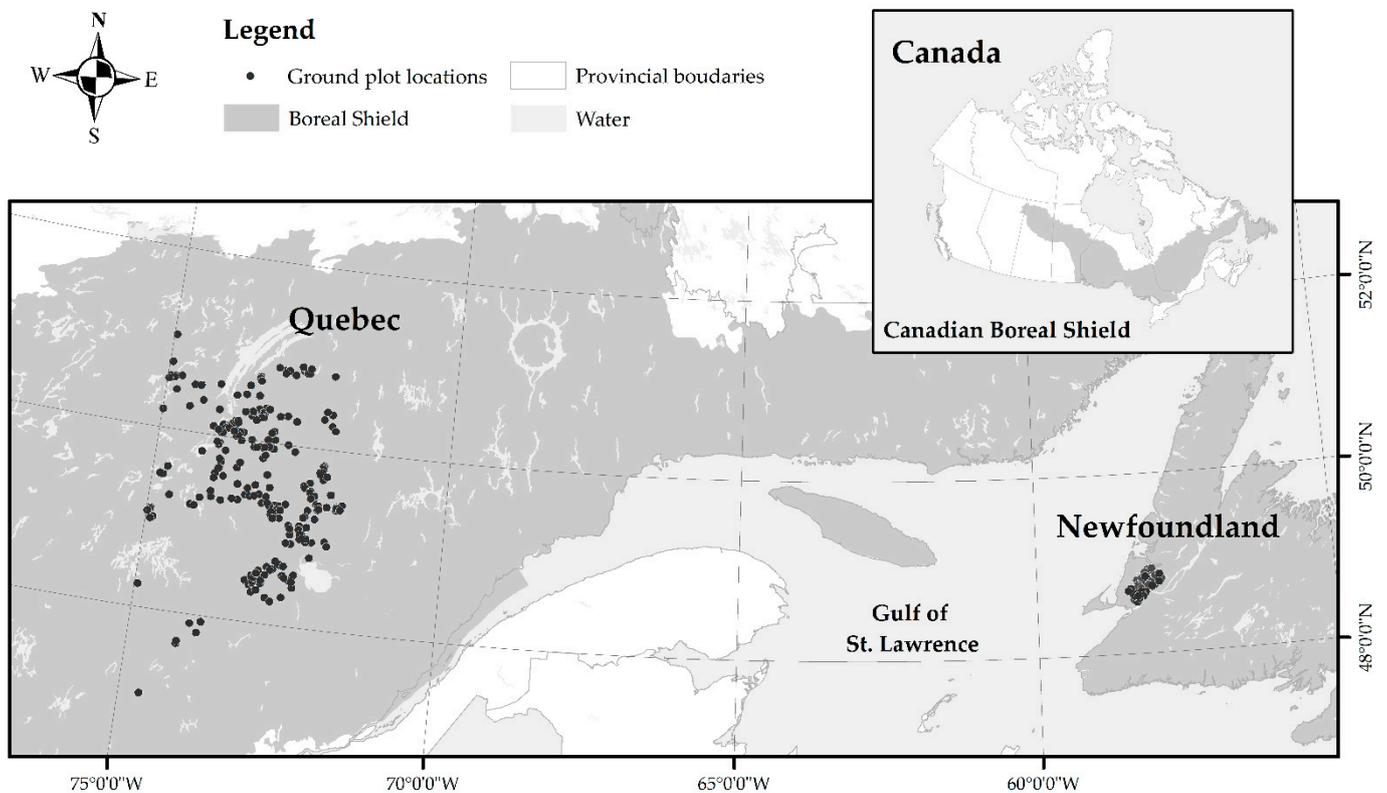
The studies provide meaningful insight into potential improvements for predicting forest attributes using a variety of modelling approaches and predictor variables that are derived from remote sensing data. To date, no studies have specifically examined whether the inclusion of canopy surface texture metrics can improve the characterization of SDD from ALS data. In this study, we compared the accuracy of SDD predictions that were modelled independently from commonly used ALS metrics, CHM-derived texture metrics, and a combination of the two using multiple statistical modelling techniques. We first hypothesized that models using texture-derived metrics would more accurately predict SDD parameters than ones using ALS metrics alone. Second, based upon past research, we hypothesized that developing differentiated modality-specific models (unimodal or bimodal) would improve SDD predictions. We tested these hypotheses by developing two modelling approaches: the first considers a priori knowledge regarding the modality of the SDD, while the second considers all SDD to be unimodal. We then evaluated the contribution of texture metrics in both approaches and determined which approach is best suited for estimating SDD in the eastern boreal forests of Quebec and western Newfoundland.

## 2. Materials and Methods

### 2.1. Study Area

Two study areas were selected based on their similarity in forest composition: both are conifer-dominated and lie within the eastern extent of the North American boreal forest [60] (Figure 1). The forests are comprised of balsam fir (*Abies balsamea* (L.) Miller), black spruce (*Picea mariana* [Miller] Britton), white spruce (*Picea glauca* [Moench] Voss), paper or white birch (*Betula papyrifera* Marshall), yellow birch (*Betula alleghaniensis* Britton) and, to a lesser extent, tamarack, or eastern larch (*Larix laricina* [Du Roi] K. Koch). Balsam fir and white spruce-dominated mixed stands are found south of the 50th parallel in our first study area (123,140 km<sup>2</sup>), located in the province of Quebec. As we move north, the presence of black spruce increases until it completely dominates the landscape above the 52nd parallel. The second study area (977 km<sup>2</sup>) is in the most eastern extent of the Boreal Shield Ecozone, in the

province of Newfoundland and Labrador, and is dominated by balsam fir. The climate at both sites is favorable for forest growth due to abundant precipitation and warm summers. The primary silvicultural treatments practiced in these areas are pre-commercial thinning and clear-cut harvesting, which generally yield even-aged, homogeneous forest stands.



**Figure 1.** Plot distribution across two sites within the eastern Boreal Shield, Canada.

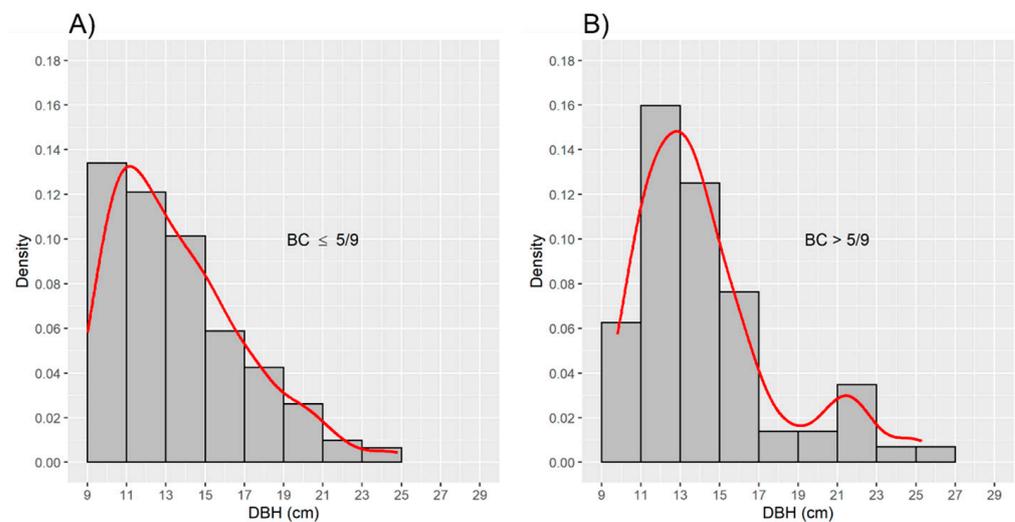
## 2.2. Ground Plots

Fixed-area circular plots were established with radii of 11.28 m where species, diameter at breast height (DBH), height, and status (live or dead) were recorded for all merchantable trees (trees  $\geq 9$  cm DBH). We retained plots having a total basal area  $\geq 75\%$  associated with balsam fir or black spruce with a presence of  $\leq 10\%$  hardwoods. We then identified and removed outlier plots by performing a multivariate local outlier factor analysis with the R package *DMwR* [61]. The analysis was based upon mean DBH and gross merchantable volume, together with the shape and scale parameters of a fitted Weibull function. We differentiated the SDD of each retained plot as unimodal or bimodal using the Bimodality Coefficient (BC) [62], given that its validity has been demonstrated in boreal forest environments [34] (Figure 2). The BC is proportional to the ratio between squared skewness and uncorrected kurtosis [63]. We associated plots having BC values  $\leq 5/9$  with unimodal distributions, while bimodal distributions were associated with BC values  $> 5/9$  [64]. In total, we retained 307 plots differentiated as unimodal and 120 as bimodal for the analysis of our hypotheses.

## 2.3. ALS Data and Metrics

All ALS data were acquired within 2 years of ground-plot measurements between 2012 and 2016. We calculated the mean point densities from plot locations to be  $5.8 \text{ points m}^{-2}$  and  $4.9 \text{ points m}^{-2}$  for the Quebec and Newfoundland sites, respectively. We created a CHM at a  $1 \text{ m} \times 1 \text{ m}$  resolution from first returns that were classified as vegetation using a natural neighbor interpolation. Binning cell assignment was set to the maximum value, and zeros replaced negative values. We calculated ALS metrics that are commonly used

to describe the height, structure, and density of the canopy using the *lidR* package [65] in the R programming environment [66], using only returns  $\geq 2$  m that were classified as vegetation. We calculated the GLCM edge (contrast and dissimilarity) and patch interior texture metrics from the CHM, i.e., correlation, homogeneity, mean, and angular second moment [67]. We considered three window sizes,  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$ , for the GLCM texture feature calculations and determined that the  $3 \times 3$  window produced metrics that explained the most variation in our response variables (i.e., Weibull parameters). We computed the GLCM features in all directions and limited the number of grey levels to 32. We then averaged the  $1 \text{ m} \times 1 \text{ m}$  resolution texture feature values for each ground plot location to produce associated metrics of texture. To evaluate our hypotheses, we grouped the predictor variables into three sets of ALS metrics based upon: (i) point cloud metrics ( $M_{\text{als}}$ ); (ii) CHM texture metrics ( $M_{\text{tex}}$ ); and (iii) a combination thereof ( $M_{\text{comb}}$ ) (Table 1).



**Figure 2.** Example of Stem Diameter Distribution (SDD) differentiated according to the Bimodality Coefficient (BC) as (A) unimodal and (B) bimodal, and respectively fitted with a Weibull distribution and a Finite Mixture Model (red lines).

**Table 1.** Description of metrics and associated groupings used as predictor variables: ALS metrics ( $M_{\text{als}}$ ), texture metrics ( $M_{\text{tex}}$ ), and combined ALS and texture metrics ( $M_{\text{comb}}$ ).

Group	Metric	Units	Description
$M_{\text{als}}$	MAX	m	Maximum height
	MEAN	m	Mean height [68]
	P25, P75, P90	m	Height percentiles. E.g., P25 is the height of the 25th percentile. [69]
	SKEW		Skewness
	VAR		Variance [68]
	COVAR	%	Coefficient of variation: standard deviation/mean [70]
	VDR		Vertical Distribution Ratio: (MAX-MEAN)/MAX [71]
	VCI		Vertical Complexity Index [72]
	ENT		Entropy: normalized Shannon diversity index [73]
	RI		Rumple Index of roughness [74]
	D2, D5, D8	%	Proportion of all vegetation returns found in sections divided within the range of heights of all returns for each plot. [75]
	COVER		Ratio of the number of vegetated returns above 2 m to the total number of ground and vegetated returns [76]
	LPI		Light Penetration Index, Ground returns/(Ground returns + Canopy returns). [69]
	LPI1st		Light Penetration Index (first returns): Ground first returns/(Ground returns + Canopy returns) [77]
FR		First return ratio: number of first return heights below a specified height threshold/total number of first return heights [68]	

Table 1. Cont.

Group	Metric	Units	Description
	RR		All return ratio: all returns < 2 m/all returns [78]
	LAI		Sum of Leaf Area Density [68]
	cvLAI		Coefficient of variation of Leaf Area Density [68]
$M_{\text{tex}}$	CON		Contrast (edge texture) [67]
	COR		Correlation (interior textures) [67]
	DIS		Dissimilarity (edge textures) [67]
	HOM		Homogeneity (interior textures) [67]
	MEAN		Mean (interior textures) [67]
$M_{\text{comb}}$			Combination of all metrics ( $M_{\text{als}}$ and $M_{\text{tex}}$ )

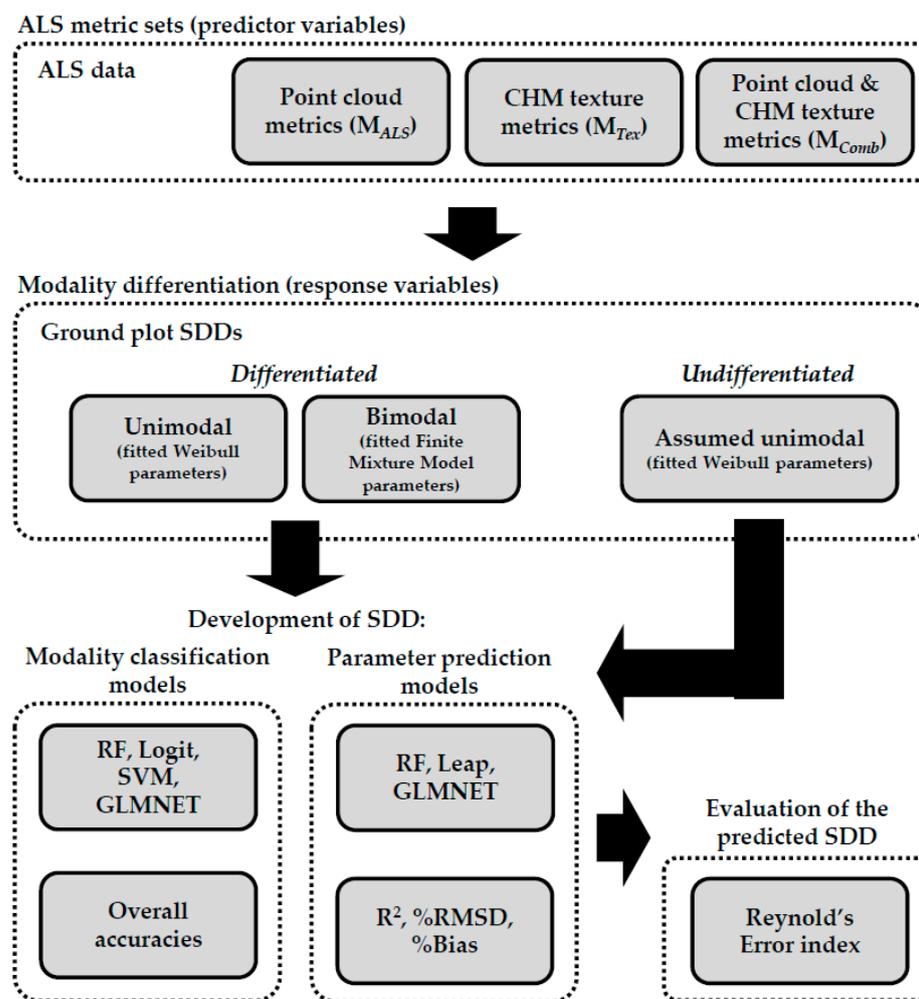
#### 2.4. Overview of the Methods

Figure 3 provides an overview of the methodological approach of the study. We used the ground-plot data to develop area-based models (i) to classify SDD modality and (ii) to predict SDD function parameters. We first defined three sets of ALS metrics from the ground plot locations ( $M_{\text{als}}$ ,  $M_{\text{tex}}$ , and  $M_{\text{comb}}$ ). We then created three ground plot datasets: the first two, unimodal and bimodal, were differentiated based on the modality of the SDD, while the third group was undifferentiated and assumed all plots were unimodal. Within each of the differentiated modality groups, we randomly selected 70% of plots for model development and used the remaining 30% as test cases. We developed models using 70% of the model development data for training and the remaining 30% for evaluating model performances. We generated three sets of models for each of the ground-plot groups using the ALS metrics sets. We used the modality and associated Weibull parameters as response variables for the SDD modality classification models and the SDD parameter prediction models, respectively. We implemented our best-performing models on our reserved test case data and analyzed the contribution of the CHM texture metrics to both groups of models (classification and prediction). Finally, we compared the predicted SDD that was obtained from the differentiated and undifferentiated modality models to assess whether modality differentiation improved the prediction of SDD in our data. All calculations were performed in R [66].

#### 2.5. Development of SDD Modality Classification Models

We developed classification models to classify the modality of SDD using the differentiated SDD modality plot datasets (unimodal and bimodal). We constructed models independently using the three metrics groups ( $M_{\text{als}}$ ,  $M_{\text{tex}}$ , and  $M_{\text{comb}}$ ) as predictor variables. Herein, we evaluated four statistical techniques: random forest (RF); generalized linear model (Logit); support vector machine (SVM); and generalized linear model through penalized maximum likelihood (GLMNET), which uses the elastic net penalty that mixes the lasso and ridge penalties [79]. These contained internal feature selection mechanisms for selecting the best predictors and models with the *caret* package [80]. We developed the RF models with the *randomForest* package [81] and optimized the parameter *mtry*, which controls the number of predictors that were randomly picked at each split, by testing five values, viz., 1, 2, 3, 4, and 5. Logit models were developed with the *MASS* package [82] and used stepwise model selection based upon the Akaike Information Criterion (AIC). We defined the family parameter as a binomial and conducted no grid search for parameter optimization. SVM models were developed with the *kernelab* package [83] and used a radial basis function. We tuned two parameters for SVM, **sigma**, which controls the rigidity of the decision boundaries, and **C**, which controls the influence of misclassification. The values for **sigma** were  $2^{-25}$ ,  $2^{-20}$ ,  $2^{-15}$ ,  $2^{-10}$ ,  $2^{-5}$ , and  $2^0$ , while those for **C** were  $2^0$ ,  $2^1$ ,  $2^2$ ,  $2^3$ ,  $2^4$ , and  $2^5$ . Finally, GLMNET models were developed with the *glmnet* package [84]. GLMNET corresponds to a ratio between model regularization levels L1 and L2, affecting the penalty coefficient, and allows the selection of relevant predictors [85]. The two parameters that

were tuned were **lambda**, which controls the overall strength of the penalty, and **alpha**, which controls the gap between the L1 and L2 regularization. We tested **alpha** values ranging from 0 to 1 with 0.1 increments and the following **lambda** values: 0.0001, 0.1112, 0.2223, 0.3334, 0.4445, 0.5556, 0.6667, 0.7778, 0.8889, and 1. We repeated cross-validation five times, using 70% of the model development data for training and 30% for validation. Finally, we averaged the overall accuracies within each technique and ALS metric group and applied the best performing models to our test case dataset and assessed the contribution of CHM texture metrics.



**Figure 3.** Overview of the methodological approach for assessing the contribution of CHM texture metrics and modality differentiation in predicting stem diameter distribution (SDD) parameters. CHM = Canopy Height Model; RF = Random Forest; Logit = generalized linear model with stepwise feature selection; SVM = Support Vector Machine; GLMNET = Generalized linear model through penalized maximum likelihood; Leap = Best subset regression with branch-and-bound algorithm; R<sup>2</sup> = Coefficient of determination; %RMSD = relative root-mean-squared deviation expressed as a percentage of the mean; %Bias = relative Bias expressed as a percentage of the mean.

### 2.6. Development of SDD Prediction Models

We developed three sets of models to predict SDD function parameters using (i) differentiated unimodal, (ii) differentiated bimodal, and (iii) undifferentiated SDD modality plot datasets. Using the differentiated unimodal plot data, we fitted a truncated Weibull function over the measured SDD and estimated the two function parameters (i.e., shape and scale) using the *fitdistrplus* package [86]. We implemented the same analysis for the undifferentiated plot data, for which all plots were treated as having a unimodal SDD

distribution. From the differentiated bimodal plot data, we fitted a FMM composed of two Weibull functions over the SDD. The first Weibull related to smaller stem diameters relative to the second Weibull, which described the probability distribution of larger stems. The FMM can be represented by either the scale and shape, or the mean and standard deviation, of each of the two Weibull components and their associated proportions. We estimated the parameters of each function using the *mixR* package [87]. We assessed three modelling techniques within each model set, which included feature selection that was based on optimizing the root-mean-square deviation (RMSD) using the *caret* package. Again, the three metric groups ( $M_{als}$ ,  $M_{tex}$ , and  $M_{comb}$ ) were used independently as predictor variables. The maximization option for RMSD was set to FALSE to ensure that the best combination of parameters produced the lowest RMSD. The first technique that was used was RF from the *randomForest* package. Again, the only optimized parameter with grid search was *mtry*, with values 1, 2, 3, 4, and 5. The second technique was GLMNET, with two parameters to optimize, i.e., **alpha** and **lambda**. The **alpha** that was tested ranged from 0 to 1 in 0.1 increments; **lambda** values were 0.0001, 0.1112, 0.2223, 0.3334, 0.4445, 0.5556, 0.6667, 0.7778, 0.8889, or 1. We implemented the third and final technique, i.e., best subset regression with branch-and-bound algorithm (LEAP) [88], with the R package *leaps* [89]. This best subset regression used the branch-and-bound algorithm [90], which solves and optimizes combinatorial problems to select the best subset of predictors. In this study, we defined the number of predictors allowed in each subset to range between 2 and 6 predictors.

We evaluated the best-tuned models from the repeated 5-time cross-validation with the reserved test case dataset not used for model development. We compared the coefficient of determination ( $R^2$ ), the absolute and relative RMSD (Equations (1) and (2)), and the absolute and relative bias (Equations (3) and (4)) for both the model development and test case datasets to assess our two hypotheses:

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

$$RMSD\% = \frac{RMSD}{\bar{y}} \times 100 \quad (2)$$

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

$$Bias\% = \frac{Bias}{\bar{y}} \times 100 \quad (4)$$

where  $y_i$  is the observed value,  $\hat{y}_i$  is the predicted value for case  $i$ ,  $n$  is the number of observations, and  $\bar{y}$  is the mean.

To evaluate the composition of metrics used in the best-performing models developed with  $M_{comb}$ , we calculated the associated variable importance. Since methods to characterize variable importance are dependent on the modelling technique implemented, we first scaled values between 0 and 100 to finally derive an average for each parameter modelled. For random forest models, we calculated the variable importance as the percent increase in mean square error (noted %IncMSE) [91]. For GLMNET models, we scaled variable coefficients as a representation of variable importance since they are proportionally indicative of the variables' importance [85] due to the penalization that reduces the coefficients of less-important variables [84]. Finally, we calculated variable importance for LEAP models as the absolute value of the t-statistic for each parameter in the final model [80].

## 2.7. Evaluation of the Predicted SDD

The quality of the predicted SDD was estimated with the Reynolds Error Index (EI) [92]. To do so, we predicted the SDD's parameters with the models demonstrating the highest  $R^2$  and lowest RMSD% for the unimodal, bimodal, and undifferentiated plots from both model development and test case datasets. We then grouped the predicted tree DBH into

2-centimetre-wide bins to limit variability at larger intervals [93]. Finally, we evaluated the goodness-of-fit between the predicted SDD and the observed SDD of each plot with EI as follows:

$$EI = \sum_{i=1}^m 100 \left| \frac{f_{refi} - f_{alsi}}{N_{ref}} \right| \quad (5)$$

where  $m$  is the total number of bins,  $f_{refi}$  is the reference stem count for DBH bin  $i$ ,  $f_{alsi}$  is the predicted stem count for DBH bin  $i$ , and  $N_{ref}$  is the true stem count of all DBH bins. EI values ranged between 0 and 200, where an EI of 0 indicated a perfect fit between predicted and observed SDD and an EI of 200 indicated a completely different SDD. To assess the effects of modality differentiation, we averaged the EI from all plots that had been derived independently for both the differentiated (unimodal and bimodal) and undifferentiated modelling approaches.

### 3. Results

#### 3.1. SDD Modality Classification Models

Table 2 denotes the overall accuracies of the modality classification models using the three ALS metric sets as predictor variables and four modelling techniques for both model development and test case datasets. During model development, we observed  $M_{als}$  and  $M_{comb}$  to perform best using RF and GLMNET (overall accuracy of 74%). Surprisingly, the  $M_{tex}$  predictor set was used in both the best (using Logit) and worst (using RF) performing models in our test case. Overall, we observed little variability in the overall model accuracies regardless of the ALS predictor variable set or modelling technique used during model development or in our test case (mean: 72%; SD: 2% in both scenarios).

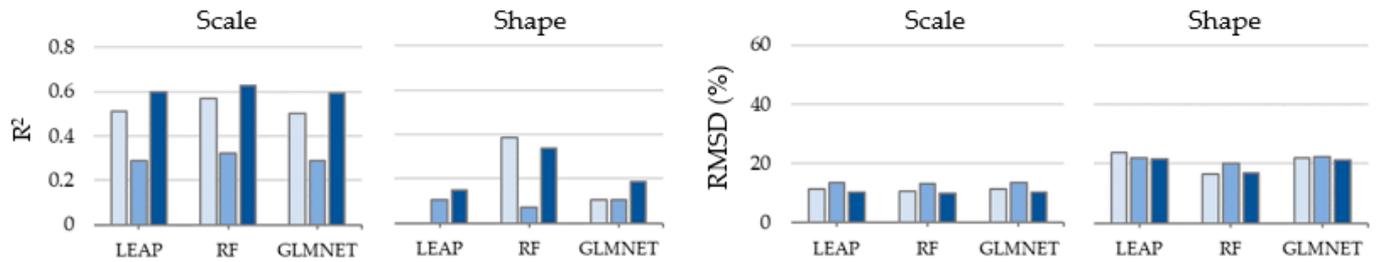
**Table 2.** Overall accuracies (%) of the SDD modality differentiation models using predictor variables that were derived from the three ALS metrics sets ( $M_{als}$ ,  $M_{tex}$ , and  $M_{comb}$ ) for both model development and test case datasets.

ALS Metric Set	RF	SVM	Logit	GLMNET
Model development				
$M_{als}$	74	72	71	74
$M_{tex}$	73	72	68	68
$M_{comb}$	74	71	70	74
Test case				
$M_{als}$	72	73	72	71
$M_{tex}$	66	72	74	73
$M_{comb}$	71	71	72	71

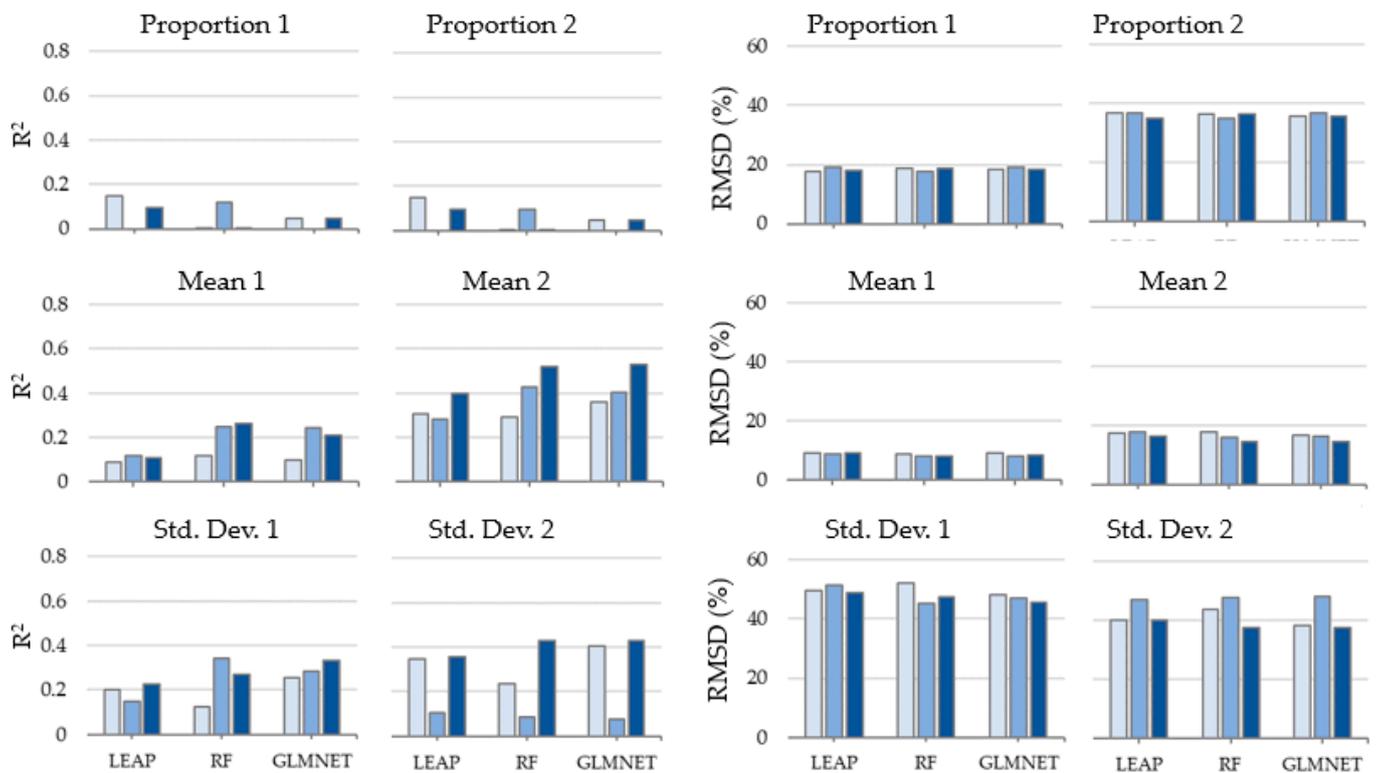
#### 3.2. SDD Prediction Models

We developed model sets to estimate probability distribution function parameters from the differentiated unimodal, differentiated bimodal, and undifferentiated SDD modality plot datasets. We developed models within each model set using the three ALS metrics sets ( $M_{als}$ ,  $M_{tex}$ , and  $M_{comb}$ ) and three modelling techniques (RF, GLMNET, and LEAP). The model performance measures ( $R^2$ , RMSD%) that were derived from cross-validation are presented as Supplementary Material (Figure S1), as we observed for the most part the same trends in results with our case study illustrated in Figure 4. The results of our test case show that the proportion of the variance in the parameters describing the differentiated unimodal SDD were variable ( $R^2$ : 0–0.62). We observed associated errors ranging between 9.9% and 13.4% and 16.4% and 23.8% for models predicting scale and shape, respectively. For both parameters, the results indicate, with one exception (Shape  $\sim f(M_{als})$  using RF), that models developed with  $M_{comb}$  consistently outperformed models that were developed with either  $M_{als}$  or  $M_{tex}$ . Both parameters were best predicted with RF; scale was best predicted using  $M_{comb}$  ( $R^2$ : 0.62; RMSD%: 9.9%), while shape, using  $M_{als}$  ( $R^2$ : 0.39; RMSD%: 16.4%).

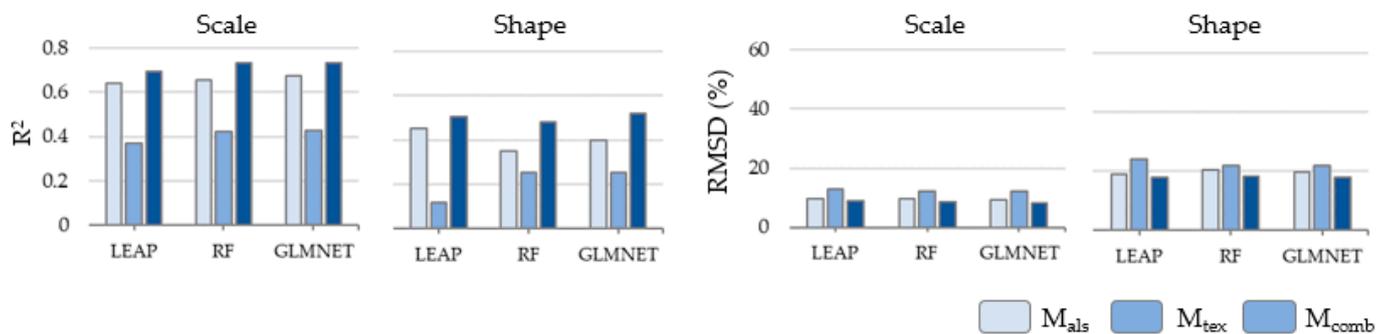
### Differentiated unimodal



### Differentiated bimodal



### Undifferentiated



**Figure 4.** Coefficient of determination ( $R^2$ ) and relative root-mean-squared deviation (RMSD%) that was derived from the application of the SDD prediction models to the test case data using the differentiated unimodal, differentiated bimodal, and undifferentiated SDD modality plot groupings; three ALS metrics sets ( $M_{als}$ ,  $M_{tex}$ , and  $M_{comb}$ ) and three modelling techniques (RF, GLMNET, LEAP) were used.

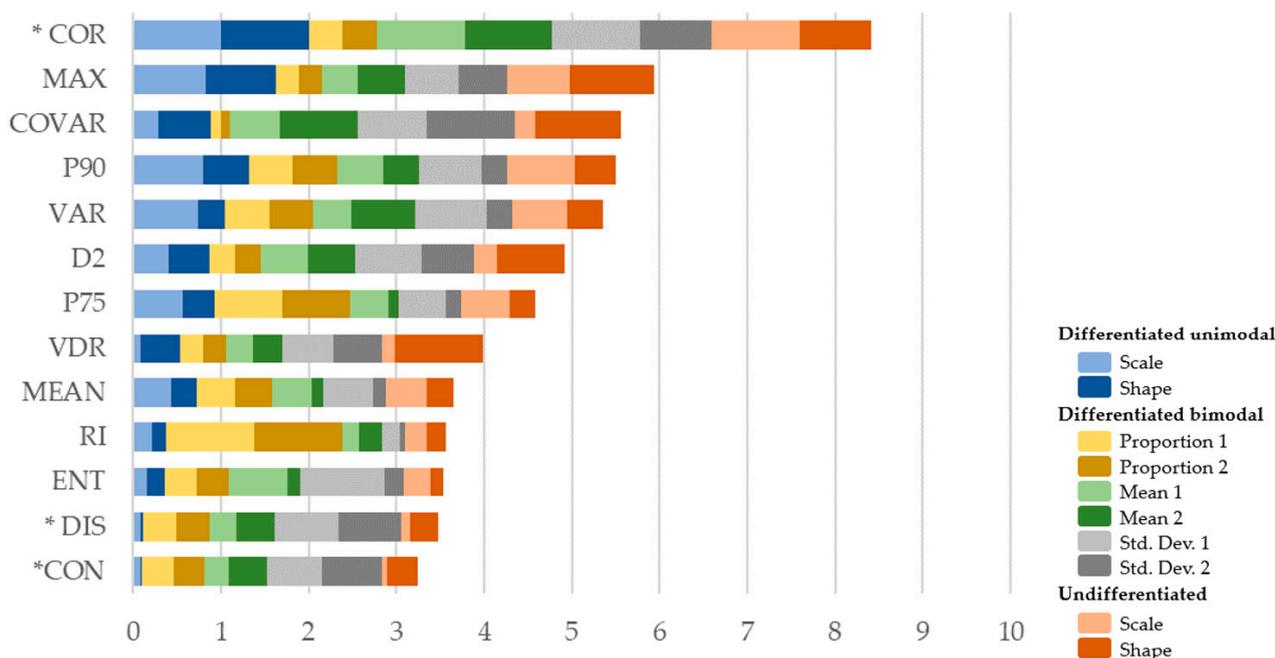
The performance of models that were developed using the differentiated bimodal SDD modality plot data were again variable ( $R^2$ : 0–0.53; RMSD%: 8.2%–52.1%). The results indicated that the FMM could not be represented by the parameters' scale and shape; the parameter shape of the first Weibull component could not be predicted given that the resulting models could never explain any of the variation in the parameter around its mean ( $R^2$ : 0), regardless of the ALS metric set or modelling approach. We therefore used the parameters mean and standard deviation to describe each component of the FMM. As expected, variation in the two proportion parameters was very poorly explained, if at all, by the predictor sets ( $R^2$ : 0–0.15), with associated errors ranging from 17.5% to 36.9%. As expected, the two Weibull component proportions of the FMMs were poorly predicted, with the best predictions modeled with RF using  $M_{\text{als}}$  ( $R^2$ : 0.15, 0.15; RMSD%: 17.5% and 33.8% for the proportions of the first and second components, respectively). The parameter mean was best predicted using  $M_{\text{comb}}$  for both components ( $R^2$ : 0.27, 0.53; RMSD%: 8.2%, 14.4%; using LEAP and GLMNET for means 1 and 2, respectively). Of note, GLMNET only marginally outperformed LEAP for the mean of the second FMM component (increase in  $R^2 < 0.01$ , decrease in RMSD%  $< 0.13\%$ ), both using  $M_{\text{comb}}$ . Standard deviation was best predicted with LEAP using  $M_{\text{tex}}$  for the first Weibull component ( $R^2$ : 0.34; RMSD%: 45.13%) and  $M_{\text{comb}}$  for the second ( $R^2$ : 0.43; RMSD%: 37.6%) with either LEAP or GLMNET.

The development of models using the undifferentiated modality SDD plot data involved applying the unimodal fitting analysis to all plots, regardless of modality. Herein, models performed better for the scale parameter ( $R^2$ : 0.37–0.73; RMSD%: 8.4%–12.9%) than for shape ( $R^2$ : 0.12–0.52; RMSD%: 17.7%–23.9%). We consistently observed improvements in model performance associated with models that have been developed with  $M_{\text{comb}}$ . Scale was best predicted with LEAP ( $R^2$ : 0.73; RMSD%: 8.4%), while shape was best predicted with GLMNET ( $R^2$ : 0.52; RMSD%: 17.7%). For these models, we observed a mean increase in  $R^2$  of 0.08 (SD: 0.03) and a mean decrease in RMSD% of 1.3% (SD: 0.6%) with models that were developed using  $M_{\text{comb}}$  over those developed using  $M_{\text{als}}$ .

Analysis of the variable importance indicated that the correlation metric from  $M_{\text{tex}}$  is holistically the most important predictor within  $M_{\text{comb}}$  (Figure 5). The most important predictors thereafter are, for the majority, from  $M_{\text{als}}$ . In summary, we generally observed higher  $R^2$  and lower RMSD% to be associated with models that were developed with  $M_{\text{comb}}$  compared with those using  $M_{\text{als}}$  or  $M_{\text{tex}}$ , regardless of the parameter being modelled or modelling technique being used. We found the variable correlation, originating from  $M_{\text{tex}}$ , to be the most important predictor within  $M_{\text{comb}}$ . Relative biases remained very low regardless of the parameter being modelled, the ALS metric set that was used, or the modelling approach that was employed (min.: –8.8; max.: 9.2; mean: 1.0; SD: 2.7 in absolute values of bias; data not shown). We observed no trend in the performance of the modelling techniques across all parameters.

### 3.3. Goodness-of-Fit of the Predicted SDD

We applied the best model within each model set independently to each plot and calculated mean Error Indices (EIs) from the predicted SDD parameters for both the model development and test case datasets (Table 3). We observed the same trends in both datasets. Surprisingly, we observed an increase in EI by applying differentiated unimodal models to unimodal plots, although the increase is negligible ( $< 1\%$ ). Differentiating modalities prior to estimating SDD most improved the accuracy of estimates for bimodal plots ( $\sim 12\%$  decrease in EI). Of the 120 plots that were used to test our models, 50 (41.7%) had a better EI when derived from differentiated modality model predictions (31 and 19 plots within the differentiated unimodal and bimodal plots, respectively). Overall, we observed a marginally better fit ( $\sim 4\%$  decrease in EI) for SDD that were estimated from the differentiated modality model set in comparison with those estimated from the undifferentiated modality model set. The results therefore indicate improvements in SDD predictions by using differentiated modality-specific models, namely for heterogeneous (bimodal) stands.



**Figure 5.** Cumulative variable importance values for metrics used in the best SDD parameter models which used  $M_{comb}$  during model development. Individual values represent the average variable importance across the three modelling techniques within each parameter and was scaled between 0 and 1. Only metrics with a cumulative value > 3 are shown. Asterisk (\*) denotes metrics originating from  $M_{tex}$ .

**Table 3.** Plot-level Reynold’s Error Index means for each ground plot dataset and model set. EI values ranged between 0 and 200, where an EI of 0 indicated a perfect fit between predicted and observed SDD, which an EI of 200 indicated a completely different SDD.

Plot Dataset	n	Model Set	
		Differentiated	Undifferentiated
Model development			
Differentiated as unimodal	215	50.4	50.3
Differentiated as bimodal	92	65	74
Undifferentiated modality	307	54.8	57.4
Test case			
Differentiated as unimodal	88	50.8	50.5
Differentiated as bimodal	32	59.1	67
Undifferentiated modality	120	53	54.9

#### 4. Discussion

From our first hypothesis, we expected models that were developed with CHM texture metrics to outperform SDD prediction models developed solely with ALS metrics. This expectation was based upon previous studies that related CHM texture metrics ( $M_{tex}$ ) to properties of the growing stock, such as the spatial pattern of trees [94], and furthermore, demonstrated that their inclusion as predictors in modelling forest attributes improved predictions over using ALS metrics alone [55,58,59]. For example, van Ewijk et al. (2019) [55] tested multiple predictor sets using height metrics with combinations of CHM texture and intensity metrics and found that the addition of texture metrics improved prediction accuracy for basal area, quadratic mean DBH, and stem density. To our knowledge, no published studies have directly assessed the contribution of CHM texture metrics in estimating SDD using ALS data. Hence, the innovative aspects of our study make direct comparisons with past research challenging, especially regarding the attributes that we

assessed (i.e., SDD modality and parameters), together with the CHM texture metrics that were included in our analyses. Nevertheless, our study demonstrated comparable results in classifying SDD modality with Zhang et al. (2019) [33] and Mulverhill et al. (2018) [34] using  $M_{\text{als}}$  (range in overall accuracies: 71%–73% vs. 49%–76% and 47%–78%, respectively). Our results for estimating SDD were generally comparable with those presented in Mulverhill et al. (2018) [34] for the differentiated unimodal distributions' modelled parameters, albeit with consistently lower error. Consistent with Thomas et al. (2008) [28] and Zhang et al. (2019) [33], the second component of the FMM that was associated with differentiated bimodal distributions was better predicted than the first. As highlighted by Thomas et al. (2008) [28], the main drawback of FMM is the increase in parameters that are needed to describe it. With the increase in modelled parameters, it becomes unlikely that each can be predicted accurately with  $M_{\text{als}}$ . Apart from the proportions associated with the FMM's components, the parameters of the differentiated bimodal distributions were best predicted with  $M_{\text{comb}}$ . Unlike Zhang et al. (2019) [33] and Mulverhill et al. (2018) [34], who developed models solely from  $M_{\text{als}}$ , our best SDD prediction models were generally developed with  $M_{\text{comb}}$ . Therefore, we could confirm our first hypothesis given that our study demonstrated that SDD prediction models developed with  $M_{\text{comb}}$  usually outperformed those developed with  $M_{\text{als}}$  (Figure 4). Inevitably, the contribution of CHM texture metrics will be dependent on the complexity of the forest environment assessed. Further research is warranted to determine the consistency of these results across varied forest types.

Our second hypothesis stated that developing differentiated modality-specific models (i.e., unimodal or bimodal) would improve SDD predictions for heterogeneous stands in our study site. The literature demonstrates improvements in estimating SDD with approaches that differentiate stand modality over approaches that do not (e.g., [33,34]). Our results indicated a similar trend. Yet, when interpreted globally, the improvements were marginal (~4 decrease in EI). Surprisingly, within our differentiated plot datasets, we observed that SDD was marginally better predicted by the undifferentiated modality model set that was intended for unimodal plots. Notably, and in support of our hypothesis, we observed SDD to be better predicted by the differentiated bimodal model set for bimodal plots (mean EI of 59.1 vs. 67.0). Our results therefore support the idea that developing model sets based on the modality of stands can improve SDD predictions for bimodal stands. Given this, we can confirm our hypothesis that differentiating for modality prior to estimating SDD improved the accuracy of estimates for the bimodal SDD conifer stands of our study site.

The accurate differentiation of the SDD modalities was assumed in our analyses, and therefore, potential errors in differentiation would directly impact model performances. Of the multiple available approaches to differentiate SDD modalities, we implemented BC as it has been successfully implemented in similar studies (e.g., [34]). Yet, it should be noted that BC is directly influenced by the kurtosis and, more so, by the skewness of a given distribution [64]. A distribution with high skewness and low kurtosis can inflate BC and subsequently differentiate the distribution as bimodal. Left-skewed distributions are observed when larger diameter trees dominate, while right-skewed distributions are associated with stands that are dominated by smaller diameter trees. Both situations will yield, however, a skewness value greater than zero. The closer that observed skewness is to zero, the more homogeneous the distribution will be and the stand can be described as having an even-aged distribution [48]. Freeman and Dale (2013) [63] evaluated the effect of the skewness, the proportion, and the distance between the modes on the BC value. In their study, BC produced 21% of false positives where simulated unimodal distributions had BC values greater than the bimodality threshold of 5/9 and were subsequently classified as bimodal. The BC relies upon the basic assumption that bimodality involves an increase in distribution asymmetry; therefore, an increase in skewness within a unimodal context can increase the BC and produce misclassification. Furthermore, the BC is not calibrated to proportion size; a small proportion in either component of a bimodal distribution can also produce false positives when the former is combined with a small distance between associated means. Of the 124 (92 in model development and 32 in test case) plots that

were differentiated as bimodal in our study, 81 had skewness estimates  $> 1$  and, thus, can be considered substantially skewed. Furthermore, the proportions that were associated with the second component of the bimodal distributions of our bimodal plots were low, as were the distances between the observed means (mean 5.8 cm). Given these results, it is possible that the combination of these factors could have inflated the BC and, therefore, mis-differentiated plots as bimodal. We can advance this as a plausible explanation, given the observed better fit for SDD that was estimated from the differentiated modalities model set was minimal (decrease in RI  $\sim 4\%$ ). These effects on the BC suggest that relying solely on this differentiation method may not be advisable for all forest types. Zhang et al. (2019) [33] used a combination of the Gini Coefficient and the asymmetry of the Lorenz curve to differentiate SDD modality, given that both measures are related to stand heterogeneity and the skewness of the diameter distribution [28,68]. Additional research is required to determine the optimal approach for differentiating the modality of SDD for a given forest type.

Nevertheless, the research presented here is important for several reasons. First, the methodology is used to differentiate the SDD modality and to develop the modality classification model, which can be used by foresters to improve the differentiation of stand structure types and to select the most appropriate models for accurately estimating diameter distributions across large ALS coverages. Second, we demonstrated that models fitted with  $M_{\text{comb}}$  yielded higher  $R^2$  and lower RMSD% in comparison with those using solely  $M_{\text{als}}$ , thereby indicating that textural metrics contain additional information useful for the estimation of SDD.

## 5. Conclusions

In this study, we demonstrated that SDD probability function parameters were generally best estimated using a combination of ALS and texture metrics, thereby emphasizing the additional information contained in CHM texture metrics. As expected, we confirmed that developing modality-specific models improved SDD predictions for bimodal distributions, which, surprisingly, was not the case for unimodal distributions. For forest managers who rely on timely and detailed information, more accurate assessments of the distribution of diameters across a land base can therefore be made by differentiating modalities and adding texture metrics to modelling and mapping efforts. These results may provide for operational efficiencies in modelling and mapping SDD in these balsam fir or spruce-dominated forest environments.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/f14020287/s1>, Figure S1: Average of 5 repeated cross-validation performance measures ( $R^2$ , RMSD%) derived during model development using the various SDD modality plot groupings, ALS metric sets and modelling techniques.

**Author Contributions:** Conceptualization, X.G.-D., O.R.v.L. and R.A.F.; methodology, X.G.-D., O.R.v.L. and R.A.F.; validation, X.G.-D.; formal analysis, X.G.-D.; investigation, X.G.-D. and O.R.v.L.; resources, O.R.v.L. and R.A.F.; data curation, X.G.-D., O.R.v.L. and R.A.F.; writing—original draft preparation, X.G.-D. and O.R.v.L.; writing—review and editing, X.G.-D., O.R.v.L. and R.A.F.; visualization, X.G.-D. and O.R.v.L.; supervision, R.A.F. and O.R.v.L.; project administration, R.A.F.; funding acquisition, R.A.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Natural Resources Canada's Canadian Forest Service—Canadian Wood Fibre Centre; and the Assessment of Wood Attributes using Remote Sensing Project (National Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant PJ-462973-14, grantee N.C. Coops, UBC); in collaboration with Corner Brook Pulp and Paper Limited; and the Newfoundland and Labrador Department of Fisheries and Land Resources.

**Data Availability Statement:** The data underlying this article will be shared on reasonable request to the corresponding author.

**Acknowledgments:** This research was mainly developed in the Centre d’Applications et de Recherche en T  l  d  tection of the Universit   de Sherbrooke, Canada. We thank Faron Knott and Kim Childs of Corner Brook Paper Limited for their input and assistance with the project. We thank the journal’s associate editor and anonymous reviewers for their constructive feedback and suggestions for improving the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Hyypp  , J.; Hyypp  , H.; Leckie, D.; Gougeon, F.; Yu, X.; Maltamo, M. Review of Methods of Small-Footprint Airborne Laser Scanning for Extracting Forest Inventory Data in Boreal Forests. *Int. J. Remote Sens.* **2008**, *29*, 1339–1366. [[CrossRef](#)]
2. N  sset, E. Area-Based Inventory in Norway—From Innovation to an Operational Reality. In *Forestry Applications of Airborne Laser Scanning*; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 215–240.
3. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [[CrossRef](#)]
4. Badreldin, N.; Sanchez-Azofeifa, A. Estimating Forest Biomass Dynamics by Integrating Multi-Temporal Landsat Satellite Images with Ground and Airborne LiDAR Data in the Coal Valley Mine, Alberta, Canada. *Remote Sens.* **2015**, *7*, 2832–2849. [[CrossRef](#)]
5. Luther, J.E.; Fournier, R.A.; van Lier, O.R.; Bujold, M. Extending ALS-Based Mapping of Forest Attributes with Medium Resolution Satellite and Environmental Data. *Remote Sens.* **2019**, *11*, 1092. [[CrossRef](#)]
6. Mielcarek, M.; Kamińska, A.; Stereńczak, K. Digital Aerial Photogrammetry (DAP) and Airborne Laser Scanning (ALS) as Sources of Information about Tree Height: Comparisons of the Accuracy of Remote Sensing Methods for Tree Height Estimation. *Remote Sens.* **2020**, *12*, 1808. [[CrossRef](#)]
7. Cao, L.; Zhang, Z.; Yun, T.; Wang, G.; Ruan, H.; She, G. Estimating Tree Volume Distributions in Subtropical Forests Using Airborne LiDAR Data. *Remote Sens.* **2019**, *11*, 97. [[CrossRef](#)]
8. Spriggs, R.A.; Coomes, D.A.; Jones, T.A.; Caspersen, J.P.; Vanderwel, M.C. An Alternative Approach to Using LiDAR Remote Sensing Data to Predict Stem Diameter Distributions across a Temperate Forest Landscape. *Remote Sens.* **2017**, *9*, 944. [[CrossRef](#)]
9. Tompalski, P.; Coops, N.C.; White, J.C.; Wulder, M.A. Enriching ALS-Derived Area-Based Estimates of Volume through Tree-Level Downscaling. *Forests* **2015**, *6*, 2608–2630. [[CrossRef](#)]
10. Zhang, L.; Gove, J.H.; Liu, C.; Leak, W.B. A Finite Mixture of Two Weibull Distributions for Modeling the Diameter Distributions of Rotated-Sigmoid, Uneven-Aged Stands. *Can. J. For. Res.* **2001**, *31*, 1654–1659. [[CrossRef](#)]
11. Cao, Q.V. Predicting Parameters of a Weibull Function for Modeling Diameter Distribution. *For. Sci.* **2004**, *50*, 682–685. [[CrossRef](#)]
12. Siipilehto, J.; Meht  talo, L. Parameter Recovery vs. Parameter Prediction for the Weibull Distribution Validated for Scots Pine Stands in Finland. *Silva Fenn.* **2013**, *47*, 22. [[CrossRef](#)]
13. McGarrigle, E.; Kershaw Jr, J.A.; Lavigne, M.B.; Weiskittel, A.R.; Ducey, M. Predicting the Number of Trees in Small Diameter Classes Using Predictions from a Two-Parameter Weibull Distribution. *Forestry* **2011**, *84*, 431–439. [[CrossRef](#)]
14. Poudel, K.P.; Cao, Q.V. Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions. *For. Sci.* **2013**, *59*, 243–252. [[CrossRef](#)]
15. Palah  , M.; Pukkala, T.; Trasobares, A. Modelling the Diameter Distribution of *Pinus Sylvestris*, *Pinus Nigra* and *Pinus Halepensis* Forest Stands in Catalonia Using the Truncated Weibull Function. *For. Int. J. For. Res.* **2006**, *79*, 553–562. [[CrossRef](#)]
16. Duan, A.G.; Zhang, J.G.; Zhang, X.Q.; He, C.Y. Stand Diameter Distribution Modelling and Prediction Based on Richards Function. *PLoS ONE* **2013**, *8*, e62605. [[CrossRef](#)]
17. Guo, H.; Lei, X.; You, L.; Zeng, W.; Lang, P.; Lei, Y. Climate-Sensitive Diameter Distribution Models of Larch Plantations in North and Northeast China. *For. Ecol. Manag.* **2022**, *506*, 119947. [[CrossRef](#)]
18. West, G.B.; Enquist, B.J.; Brown, J.H. A General Quantitative Theory of Forest Structure and Dynamics. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7040–7045. [[CrossRef](#)]
19. Martin Bolland  s, O.; N  sset, E. Estimating Percentile-Based Diameter Distributions in Uneven-Sized Norway Spruce Stands Using Airborne Laser Scanner Data. *Scand. J. For. Res.* **2007**, *22*, 33–47. [[CrossRef](#)]
20. Rana, P.; Vauhkonen, J.; Junttila, V.; Hou, Z.; Gautam, B.; Cawkwell, F.; Tokola, T. Large Tree Diameter Distribution Modelling Using Sparse Airborne Laser Scanning Data in a Subtropical Forest in Nepal. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 86–95. [[CrossRef](#)]
21. Xu, Q.; Hou, Z.; Maltamo, M.; Tokola, T. Calibration of Area Based Diameter Distribution with Individual Tree Based Diameter Estimates Using Airborne Laser Scanning. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 65–75. [[CrossRef](#)]
22. Fries, C.; Johansson, O.; Pettersson, B.; Simonsson, P. Silvicultural Models to Maintain and Restore Natural Stand Structures in Swedish Boreal Forests. *For. Ecol. Manag.* **1997**, *94*, 89–103. [[CrossRef](#)]
23. Packal  n, P.; Maltamo, M. Estimation of Species-Specific Diameter Distributions Using Airborne Laser Scanning and Aerial Photographs. *Can. J. For. Res.* **2008**, *38*, 1750–1760. [[CrossRef](#)]
24. Bailey, R.L.; Dell, T.R. Quantifying Diameter Distributions with the Weibull Function. *For. Sci.* **1973**, *19*, 97–104. [[CrossRef](#)]

25. Hafley, W.L.; Schreuder, H.T. Statistical Distributions for Fitting Diameter and Height Data in Even-Aged Stands. *Can. J. For. Res.* **1977**, *7*, 481–487. [[CrossRef](#)]
26. Gorgoso-Varela, J.J.; Ponce, R.A.; Rodríguez-Puerta, F. Modeling Diameter Distributions with Six Probability Density Functions in Pinus Halepensis Mill. Plantations Using Low-Density Airborne Laser Scanning Data in Aragón (Northeast Spain). *Remote Sens.* **2021**, *13*, 2307. [[CrossRef](#)]
27. Nepomuceno Cosenza, D.; Soares, P.; Guerra-Hernández, J.; Pereira, L.; González-Ferreiro, E.; Castedo-Dorado, F.; Tomé, M. Comparing Johnson's S B and Weibull Functions to Model the Diameter Distribution of Forest Plantations through ALS Data. *Remote Sens.* **2019**, *11*, 2792. [[CrossRef](#)]
28. Thomas, V.; Oliver, R.D.; Lim, K.; Woods, M. LiDAR and Weibull Modeling of Diameter and Basal Area. *For. Chron.* **2008**, *84*, 866–875. [[CrossRef](#)]
29. Hao, Y.; Widagdo, F.R.A.; Liu, X.; Quan, Y.; Liu, Z.; Dong, L.; Li, F. Estimation and Calibration of Stem Diameter Distribution Using UAV Laser Scanning Data: A Case Study for Larch (*Larix Olgensis*) Forests in Northeast China. *Remote Sens. Environ.* **2022**, *268*, 112769. [[CrossRef](#)]
30. Peuhkurinen, J.; Tokola, T.; Plevak, K.; Sirparanta, S.; Kedrov, A.; Pyankov, S. Predicting Tree Diameter Distributions from Airborne Laser Scanning, SPOT 5 Satellite, and Field Sample Data in the Perm Region, Russia. *Forests* **2018**, *9*, 639. [[CrossRef](#)]
31. Maltamo, M.; Malinen, J.; Kangas, A.; Härkönen, S.; Pasanen, A.M. Most Similar Neighbour-Based Stand Variable Estimation for Use in Inventory by Compartments in Finland. *Forestry* **2003**, *76*, 449–463. [[CrossRef](#)]
32. Mauro, F.; Frank, B.; Monleon, V.J.; Temesgen, H.; Ford, K.R. Prediction of Diameter Distributions and Tree-Lists in Southwestern Oregon Using LiDAR and Stand-Level Auxiliary Information. *Can. J. For. Res.* **2019**, *49*, 775–787. [[CrossRef](#)]
33. Zhang, Z.; Cao, L.; Mulverhill, C.; Liu, H.; Pang, Y.; Li, Z. Prediction of Diameter Distributions with Multimodal Models Using LiDAR Data in Subtropical Planted Forests. *Forests* **2019**, *10*, 125. [[CrossRef](#)]
34. Mulverhill, C.; Coops, N.C.; White, J.C.; Tompalski, P.; Marshall, P.L.; Bailey, T. Enhancing the Estimation of Stem-Size Distributions for Unimodal and Bimodal Stands in a Boreal Mixedwood Forest with Airborne Laser Scanning Data. *Forests* **2018**, *9*, 95. [[CrossRef](#)]
35. Saad, R.; Wallerman, J.; Lämås, T. Estimating Stem Diameter Distributions from Airborne Laser Scanning Data and Their Effects on Long Term Forest Management Planning. *Scand. J. For. Res.* **2015**, *30*, 186–196. [[CrossRef](#)]
36. Strunk, J.L.; McGaughey, R.J. Stand Validation of Lidar Forest Inventory Modeling for a Managed Southern Pine Forest. *Can. J. For. Res.* **2023**, *53*, 1–19. [[CrossRef](#)]
37. Peuhkurinen, J.; Mehtätalo, L.; Maltamo, M. Comparing Individual Tree Detection and the Areabased Statistical Approach for the Retrieval of Forest Stand Characteristics Using Airborne Laser Scanning in Scots Pine Stands. *Can. J. For. Res.* **2011**, *41*, 583–598. [[CrossRef](#)]
38. Arias-Rodil, M.; Diéguez-Aranda, U.; Álvarez-González, J.G.; Pérez-Cruzado, C.; Castedo-Dorado, F.; González-Ferreiro, E. Modeling Diameter Distributions in Radiata Pine Plantations in Spain with Existing Countrywide LiDAR Data. *Ann. For. Sci.* **2018**, *75*, 36. [[CrossRef](#)]
39. Tomppo, E.; Katila, M. Satellite Image-Based National Forest Inventory of Finland for Publication in the Igarss'91 Digest. In Proceedings of the IGARSS'91 Remote Sensing: Global Monitoring for Earth Management, Espoo, Finland, 3–6 June 1991; Volume 3, pp. 1141–1144.
40. Maltamo, M.; Kangas, A. Methods Based on K-Nearest Neighbor Regression in the Prediction of Basal Area Diameter Distribution. *Can. J. For. Res.* **1998**, *28*, 1107–1115. [[CrossRef](#)]
41. Næsset, E. Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [[CrossRef](#)]
42. Bollandsås, O.M.; Maltamo, M.; Gobakken, T.; Næsset, E. Comparing Parametric and Non-Parametric Modelling of Diameter Distributions on Independent Data Using Airborne Laser Scanning in a Boreal Conifer Forest. *Forestry* **2013**, *86*, 493–501. [[CrossRef](#)]
43. Peuhkurinen, J.; Maltamo, M.; Malinen, J. Estimating Species-Specific Diameter Distributions and Saw Log Recoveries of Boreal Forests from Airborne Laser Scanning Data and Aerial Photographs: A Distribution-Based Approach. *Silva Fenn.* **2008**, *42*, 625–641. [[CrossRef](#)]
44. Strunk, J.L.; Gould, P.J.; Packalen, P.; Poudel, K.P.; Andersen, H.-E.E.; Temesgen, H. An Examination of Diameter Density Prediction with K-NN and Airborne Lidar. *Forests* **2017**, *8*, 444. [[CrossRef](#)]
45. Kangas, A.; Maltamo, M. Percentile Based Basal Area Diameter Distribution Models for Scots Pine, Norway Spruce and Birch Species. *Silva Fenn.* **2000**, *34*, 371–380. [[CrossRef](#)]
46. Liu, C.; Beaulieu, J.; Prigent, G.; Zhang, S.Y. Applications and Comparison of Six Methods for Predicting Parameters of the Weibull Function in Unthinned Picea Glauca Plantations. *Scand. J. For. Res.* **2009**, *24*, 67–75. [[CrossRef](#)]
47. Borders, B.E.; Souter, R.A.; Bailey, R.L.; Ware, K.D. Percentile-Based Distributions Characterize Forest Stand Tables. *For. Sci.* **1987**, *33*, 570–576. [[CrossRef](#)]
48. Zhang, L.; Liu, C. Fitting Irregular Diameter Distributions of Forest Stands by Weibull, Modified Weibull, and Mixture Weibull Models. *J. For. Res.* **2006**, *11*, 369–372. [[CrossRef](#)]
49. Tarp-Johansen, M.J. Stem Diameter Estimation from Aerial Photographs. *Scand. J. For. Res.* **2002**, *17*, 369–376. [[CrossRef](#)]

50. Gobakken, T.; Næsset, E. Estimation of Diameter and Basal Area Distributions in Coniferous Forest by Means of Airborne Laser Scanner Data. *Scand. J. For. Res.* **2004**, *19*, 529–542. [[CrossRef](#)]
51. Shang, C.; Treitz, P.; Caspersen, J.; Jones, T. Estimating Stem Diameter Distributions in a Management Context for a Tolerant Hardwood Forest Using ALS Height and Intensity Data. *Can. J. Remote Sens.* **2017**, *43*, 79–94. [[CrossRef](#)]
52. Hou, Z.; Xu, Q.; Tokola, T. Use of ALS, Airborne CIR and ALOS AVNIR-2 Data for Estimating Tropical Forest Attributes in Lao PDR. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 776–786. [[CrossRef](#)]
53. Haralick, R.M.; Shanmugam, K.; Its'Haq, D. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
54. Tuceryan, M.; Jain, A.K. Texture Analysis. In *Handbook of Pattern Recognition and Computer Vision*; Chen, C.H., Pau, L.F., Wang, P.S.P., Eds.; World Scientific: Singapore, 1999; pp. 207–248.
55. van Ewijk, K.; Treitz, P.; Woods, M.; Jones, T.; Caspersen, J. Forest Site and Type Variability in ALS-Based Forest Resource Inventory Attribute Predictions over Three Ontario Forest Sites. *Forests* **2019**, *10*, 226. [[CrossRef](#)]
56. Tuominen, S.; Pekkarinen, A. Performance of Different Spectral and Textural Aerial Photograph Features in Multi-Source Forest Inventory. *Remote Sens. Environ.* **2005**, *94*, 256–268. [[CrossRef](#)]
57. Dube, T.; Mutanga, O. Investigating the Robustness of the New Landsat-8 Operational Land Imager Derived Texture Metrics in Estimating Plantation Forest Aboveground Biomass in Resource Constrained Areas. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 12–32. [[CrossRef](#)]
58. Ozdemir, I.; Donoghue, D.N.M. Modelling Tree Size Diversity from Airborne Laser Scanning Using Canopy Height Models with Image Texture Measures. *For. Ecol. Manag.* **2013**, *295*, 28–37. [[CrossRef](#)]
59. Niemi, M.T.; Vauhkonen, J. Extracting Canopy Surface Texture from Airborne Laser Scanning Data for the Supervised and Unsupervised Prediction of Area-Based Forest Characteristics. *Remote Sens.* **2016**, *8*, 582. [[CrossRef](#)]
60. Rowe, J.S. *Forest Regions of Canada. Based on W. E. D. Halliday's "A Forest Classification for Canada" 1937*; Publication No 1300; Department of the Environment, Canadian Forestry Service: Ottawa, ON, Canada, 1972.
61. Torgo, L. *Data Mining with R: Learning with Case Studies*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017; ISBN 978-1482234893.
62. Ellison, A.M. Effect of Seed Dimorphism on the Density-Dependent Dynamics of Experimental Populations of *Atriplex Triangularis* (Chenopodiaceae). *Am. J. Bot.* **1987**, *74*, 1280–1288. [[CrossRef](#)]
63. Freeman, J.B.; Dale, R. Assessing Bimodality to Detect the Presence of a Dual Cognitive Process. *Behav. Res. Methods* **2013**, *45*, 83–97. [[CrossRef](#)]
64. Pfister, R.; Schwarz, K.A.; Janczyk, M.; Dale, R.; Freeman, J. Good Things Peak in Pairs: A Note on the Bimodality Coefficient. *Front. Psychol.* **2013**, *4*, 700. [[CrossRef](#)]
65. Roussel, J.-R.; Auty, D. LidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. Available online: <https://cran.r-project.org/web/packages/lidR/index.html> (accessed on 15 May 2020).
66. R Core Team R: A Language and Environment for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 22 May 2020).
67. Hall-Beyer, M. Practical Guidelines for Choosing GLCM Textures to Use in Landscape Classification Tasks over a Range of Moderate Spatial Scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338. [[CrossRef](#)]
68. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.-P.P. Generalizing Predictive Models of Forest Inventory Attributes Using an Area-Based Approach with Airborne LiDAR Data. *Remote Sens. Environ.* **2015**, *156*, 322–334. [[CrossRef](#)]
69. Peduzzi, A.; Wynne, R.H.; Fox, T.R.; Nelson, R.F.; Thomas, V.A. Estimating Leaf Area Index in Intensively Managed Pine Plantations Using Airborne Laser Scanner Data. *For. Ecol. Manag.* **2012**, *270*, 54–65. [[CrossRef](#)]
70. Pope, G.; Treitz, P. Leaf Area Index (LAI) Estimation in Boreal Mixedwood Forest of Ontario, Canada Using Light Detection and Ranging (LiDAR) and Worldview-2 Imagery. *Remote Sens.* **2013**, *5*, 5040–5063. [[CrossRef](#)]
71. Goetz, S.; Steinberg, D.; Dubayah, R.; Blair, B. Laser Remote Sensing of Canopy Habitat Heterogeneity as a Predictor of Bird Species Richness in an Eastern Temperate Forest, USA. *Remote Sens. Environ.* **2007**, *108*, 254–263. [[CrossRef](#)]
72. van Ewijk, K.Y.; Treitz, P.M.; Scott, N.A. Characterizing Forest Succession in Central Ontario Using Lidar-Derived Indices. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 261–269. [[CrossRef](#)]
73. Pretzsch, H. Description and Analysis of Stand Structures. In *Forest Dynamics, Growth and Yield*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 223–289. ISBN 9783540883067.
74. Jenness, J.S. Calculating Landscape Surface Area from Digital Elevation Models. *Wildl. Soc. Bull.* **2004**, *32*, 829–839. [[CrossRef](#)]
75. 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. [[CrossRef](#)]
76. Beets, P.N.; Reutebuch, S.; Kimberley, M.O.; Oliver, G.R.; Pearce, S.H.; McGaughey, R.J. Leaf Area Index, Biomass Carbon and Growth Rate of Radiata Pine Genetic Types and Relationships with LiDAR. *Forests* **2011**, *2*, 637–659. [[CrossRef](#)]
77. Solberg, S.; Brunner, A.; Hanssen, K.H.; Lange, H.; Næsset, E.; Rautiainen, M.; Stenberg, P. Mapping LAI in a Norway Spruce Forest Using Airborne Laser Scanning. *Remote Sens. Environ.* **2009**, *113*, 2317–2327. [[CrossRef](#)]
78. Hopkinson, C.; Chasmer, L. Testing LiDAR Models of Fractional Cover across Multiple Forest Ecozones. *Remote Sens. Environ.* **2009**, *113*, 275–288. [[CrossRef](#)]

79. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]
80. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
81. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
82. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002.
83. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20. [[CrossRef](#)]
84. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
85. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning: Data Mining Inference and Prediction*, 2nd ed.; Springer Series in Statistics New York; Springer: New York, NY, USA, 2001; Volume 1, ISBN 978-0-387-21606-5.
86. Delignette-Muller, M.L.; Dutang, C. Fitdistrplus: An R Package for Fitting Distributions. *J. Stat. Softw.* **2015**, *64*, 1–34. [[CrossRef](#)]
87. Yu, Y. MixR: An R Package for Finite Mixture Modeling for Both Raw and Binned Data. *J. Open Source Softw.* **2022**, *7*, 4031. [[CrossRef](#)]
88. Furnival, G.M.; Wilson, R.W. Regressions by Leaps and Bounds. *Technometrics* **1974**, *16*, 499–511. [[CrossRef](#)]
89. Lumley, T. Leaps: Regression Subset Selection, Based on Fortran Code by Alan Miller, R Package Version 3.1. Available online: <https://cran.r-project.org/web/packages/leaps/leaps.pdf> (accessed on 18 May 2020).
90. Land, A.H.; Doig, A.G. An Automatic Method of Solving Discrete Programming Problems. *Econometrica* **1960**, *28*, 497–520. [[CrossRef](#)]
91. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling Spatial Patterns of Fire Occurrence in Mediterranean Europe Using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. [[CrossRef](#)]
92. Reynolds, M.R.; Burk, T.E.; Huang, W.-C. Goodness-of-Fit Tests and Model Selection Procedures for Diameter Distribution Models. *For. Sci.* **1988**, *34*, 373–399. [[CrossRef](#)]
93. Coomes, D.A.; Duncan, R.P.; Allen, R.B.; Truscott, J. Disturbances Prevent Stem Size-Density Distributions in Natural Forests from Following Scaling Relationships. *Ecol. Lett.* **2003**, *6*, 980–989. [[CrossRef](#)]
94. Pippuri, I.; Kallio, E.; Maltamo, M.; Peltola, H.; Packalén, P. Exploring Horizontal Area-Based Metrics to Discriminate the Spatial Pattern of Trees and Need for First Thinning Using Airborne Laser Scanning. *Forestry* **2012**, *85*, 305–314. [[CrossRef](#)]

**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.