Prediction of Diameter Distributions with Multimodal Models Using LiDAR Data in Subtropical Planted Forests

Zhengnan Zhang [1] , **Lin Cao** [1,*] , **Christopher Mulverhill** [2] , **Hao Liu** [1] , **Yong Pang** [3]
and Zengyuan Li [3]

[1] Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University,
   Nanjing 210037, China; Zhangzhengnan_njfu@hotmail.com (Z.Z.); Liuhao_njfu@hotmail.com (H.L.)
[2] Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada;
   cmulv@mail.ubc.ca
[3] Institute of Forest Resources Information Technique, Chinese Academy of Forestry, Beijing 100091, China;
   pangy@ifrit.ac.cn (Y.P.); zengyuan.li@ifrit.ac.cn (Z.L.)
* Correspondence: lincao@njfu.edu.cn; Tel.: +86-137-7665-8458

check for
updates

Abstract: Tree diameter distributions are essential for the calculation of stem volume and biomass, as well as simulation of growth and yield and to understand timber assortments. Accurate and reliable prediction of tree diameter distributions is critical for optimizing forest structure compositions, scheduling silvicultural operations and promoting sustainable management. In this study, we investigated the potential of airborne Light Detection and Ranging (LiDAR) data for predicting tree diameter distributions using a bimodal finite mixture model (FMM) and a multimodal k-nearest neighbor (KNN) model (compared to the unimodal Weibull model (UWM)) over a subtropical planted forest in southern China. To do so, we first evaluated the capability of various LiDAR predictions (i.e., the bimodality coefficient (BC) and Lorenz-based indicators) to stratify forest structural types into unimodal and multimodal stands. Once the best LiDAR prediction for the differentiation was determined, the parameters of UWM (in non-specific and species-specific models) and FMM (in structure-specific models) were estimated by LiDAR-derived metrics and the tree diameter distributions of stands were generated by the estimated LiDAR parameters. When KNN was applied for constructing diameter distributions, optimal KNN strategies, including number of neighbors k, response configurations and imputation methods (i.e., Most Similar Neighbor (MSN) and Random Forest (RF)) for different species were heuristically determined. Finally, the predictive performance of estimated LiDAR the parameters of UWM, FMM and KNN for predicting diameter distributions were assessed. The results showed that LiDAR-predicted Lorenz-based indicators performed best for differentiation. Parameters of UWM and FMM were predicted well and the species-specific models had higher accuracies than the non-specific models. Overall, RF imputation from KNN with an optimal response set (i.e., DBH) were was stable than MSN imputation when k = 5 neighbors. In addition, the inclusion of bimodal FMM for differentiated all plots generally produced a more accurate result (Mean e_R = 40.85, Mean e_P = 0.20) than multimodal KNN (Mean e_R = 52.19, Mean e_P = 0.26), whereas the UWM produced the lowest performance (Mean e_R = 52.31, Mean e_P = 0.26). This study demonstrated the benefits of multimodal models with LiDAR for estimating diameter distributions for supporting forest inventory and sustainable forest management in subtropical planted forests.

Keywords: diameter distribution; LiDAR; Weibull; finite mixture model; KNN; Lorenz curves; planted forests

1. Introduction

As an indispensable component of the world's forest resources, planted forests make up approximately 6.94% of global forest areas [1]. Planted forests make significant contributions to society and the environment and have the potential to provide not only timber products (e.g., wood, fiber and fuel), but also non-timber forest services and ecological functions [2,3], such as carbon sequestration, rehabilitation of degraded land, and species diversity conservation [3,4]. China currently has the largest planted forest area at 91.8 million ha, approximately 33.03% of total planted forests in the world [1,5]. Subtropical planted forests account for approximately 9% of planted forest areas in the world [5] and are regarded as an important ecosystem with higher species richness and higher productivity, compared to boreal or temperate forests [6]. Differing from natural or semi-natural forests, subtropical planted forests usually endure intensive harvest or other human disturbance, due to commercial operations. China, especially in the southeastern region, also contains the largest proportion of low latitude humid subtropical forests, which play an essential role in establishing an ecological balance and improving environmental quality [7,8]. Attributes describing forest structure can be an imperative support tool in forest management planning [9]. Thus, timely, effectively, and accurately acquiring up-to-date and reliable information on the state and structure attributes of China's subtropical planted forests is crucial for enhancing forest productivity, understanding forest ecological function, supporting forest management decisions, improving silvicultural treatments (e.g., thinning, timber harvesting and assortment) and promoting sustainable management [1,10–12].

Tree diameter distributions of forest stands are critical to describing forest structures and functions [13–16]. For example, a diameter distribution is an essential component for calculation of stand volume, biomass, and carbon stocks [14,17,18], stand-level timber assortment [17,19], assessment of ecological biodiversity [20,21], and forest growth and yield [22]. Accurately predicting diameter distributions of planted forest stands is of great importance to inform forest managers for understanding forest structure and succession [23], updating forest inventories [24] and scheduling optimal silvicultural treatments [25], thereby facilitating efficient use of forest resources and supporting sustainable management [26]. Traditionally, forest inventories for acquiring diameter distribution have relied on field measurements, which is time consuming, spatially constrained, and labor intensive [27]. As a result, remote sensing has become a key instrument to enhance traditional forest inventories and inform spatial forest management planning [27–29]. Airborne Light Detection and Ranging (LiDAR) has been widely regarded as an information-rich tool for accurately characterizing 3D forest structure and providing spatially consistent wall-to-wall estimations of forest attributes, such as diameter at breast height (DBH), tree height, basal area, volume, stem density and biomass in an area-based approach (ABA) [30–32].

In recent years, due to strong correlation with LiDAR-derived forest structure attributes (e.g., DBH, basal area, volume, stem density, etc.) [33], airborne LiDAR has shown potential in predicting diameter distributions using a variety of strategies [24,27,28,34–39]. One such strategy, using parametric distributions, is based on the assumption that the diameter distribution's probability density can be characterized by a theoretical probability density function (PDF) [22,35], e.g., Weibull, beta, gamma and Log-normal [27]. The Weibull model has been widely applied for modeling diameter distributions, due to its flexibility and fewer parameters [40]. The ABA method for estimating diameter distributions by airborne LiDAR employs LiDAR metrics as predictors to estimate density function parameters [22], such as the scale and shape parameters, in the case of a Weibull model. Gobakken and Næsset [39] were among the first to predict diameter distributions based on estimated Weibull parameters using airborne LiDAR data over a boreal forest in southeast Norway. In studies, such as this, diameter distributions derived from a unimodal Weibull (UWM) model could accurately be predicted by LiDAR data. Another parametric approach used to infer diameter distributions is a parameter recovery method (PRM), which is an indirect approach that first estimates other distribution-related attributes and then uses these estimates to predict parameters of one PDF, e.g., a Weibull model. The PRM approach can be parameterized by the field-inventoried forest attributes [41–44], e.g., stand age, DBH, height,

and volume, or by LiDAR metrics [45]. However, for irregular stand structures, such as uneven-aged stands, mixed-species stands, and old-growth stands, the use of these UWM and PRM may lead to an oversimplified characterization of diameter distributions and the irregularities of distributions are more or less ignored [38,46]. For these structural types, an alternative method for modeling a multimodal or irregular distribution of diameters is a finite mixture model (FMM), which is made up of two or more component distributions (typically Weibull) [47,48]. Thomas et al. [36] compared an FMM approach using two component Weibull distributions and a UWM to predict diameter distributions with LiDAR data in the boreal forest stands of central Ontario. Results suggested that the FMM approach provided more effective predictions of diameter distribution than the UWM in irregular stands. Mulverhill et al. [48] used the FMM approach to predict diameter distributions in bimodal stands that were differentiated by LiDAR metrics in a boreal mixedwood forest. They found that the FMMs provided more accurate estimates for bimodal stands compared with unimodal Weibull models.

A non-parametric strategy with an effective ability to predict multimodal diameter distributions is a k-Nearest Neighbor (KNN) imputation model [22,33,35]. KNN is a reliable strategy in more irregular forest stand structures, because KNN does not rely on any underlying probability distribution [49–51] and can accurately describe the local variability of diameter distributions [37,52]. The most frequently employed imputation method is a Most Similar Neighbor (MSN), in which the k neighbor reference observations are determined by distances computed in a projected canonical space [53]. Packalén and Maltamo [27] employed the MSN method to impute diameter distributions and compared the performance of MSN to a Weibull model in Finnish boreal forests. They found that the diameter distribution based on MSN imputed tree-lists outperformed the Weibull distribution. Maltamo et al. [38] also applied airborne LiDAR data to predict stand diameter distributions in south-eastern Norway. Their results reported the utility of non-parametric MSN method for prediction of diameter distributions. Recently, Random Forest (RF) [54] has been adapted as a KNN imputation method [55] in which numerous classification trees fitted from a random sample of reference data are concatenated to calculate the proportion of shared terminal nodes across all the response variables for the purpose of "identifying nearest neighbors" [56], and therefore represent an essentially different approach from the other KNN imputation methods [57]. Penner et al. [58] found the potential of this imputation method to predict tree diameter distributions in a northeastern Ontario boreal forest. Strunk et al. [35] also evaluated RF imputation strategies for prediction of tree diameter distributions with tree species using LiDAR data in a boreal forest. The utility of RF imputation also been reported by Shang et al. [59].Using LiDAR data for predicting diameter distributions are usually based on UWM approaches that are difficult to implement for complex forest stands, whereas studies related to using multimodal FMM or KNN approach to predict multimodal diameter distributions are few. Furthermore, previous studies for estimating LiDAR-derived diameter distributions are typically limited to boreal forests of Europe [33,37–39,51,60,61] or boreal and temperate forests of North America [36,48,58], whereas few studies focus on subtropical planted forests. Predicting diameter distributions of subtropical planted forests aid in operational planning of timber harvest and other silvicultural treatments. In this study, we predicted diameter distributions using multimodal FMM and KNN approaches compared to UWM based on airborne LiDAR data in a subtropical planted forest of Southern China. The aims of this study are: (1) To evaluate the capability of LiDAR for discriminating unimodal and multimodal stands by examining the modality of diameter distributions; (2) to implement multimodal FMM and KNN models for predicting tree diameter distributions, and to investigate the sensitivity of various KNN strategies, including the number of neighbors (i.e., k), response configurations and imputations approaches (i.e., MSN and RF) for KNN-imputed diameter distributions; and (3) to assess the performance of FMM and KNN models for predicting diameter distributions by comparing to those estimated by UWM.

2. Materials and Methods

2.1. Study Area

This study was conducted in Gaofeng Forest, a state-operated planted forest located near the town of Nanning city of Guangxi Province in southern China, approximately between latitudes 22°49′ N, 23°5′ N and longitudes 108°7′ E, 108°38′ E (Figure 1). The area of the study site is approximately 5200 ha with the elevation range between approximately 80 m and 460 m above sea level. The climate of this study site belongs to a typical south-subtropical monsoon with annual mean temperature, mean precipitation means a relative humidity and annual total sunshine of approximately, of 21.6 °C, 1304.2 mm, 79%, 1500–1700 h. The highest monthly precipitation occurs in summer. The main soil type is lateritic red soil developed by sand shale, and the average thickness of the soil layer is above 80 cm. The Gaofeng Forest is a south-subtropical plantation type with two main habitats: Conifer-dominated and broadleaf-dominated. Planted broadleaf tree species are mainly swamp mahogany (*Eucalyptus robusta* Smith), along with Timor white gum (*Eucalyptus urophylla* ST Blake) and flooded gum (*Eucalyptus grandis* L.). The dominant coniferous tree species is Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.). Other species include *Acacia confuse* Merr., *Magnolia glance* Blume. and *Pinus massoniana* Lamb.
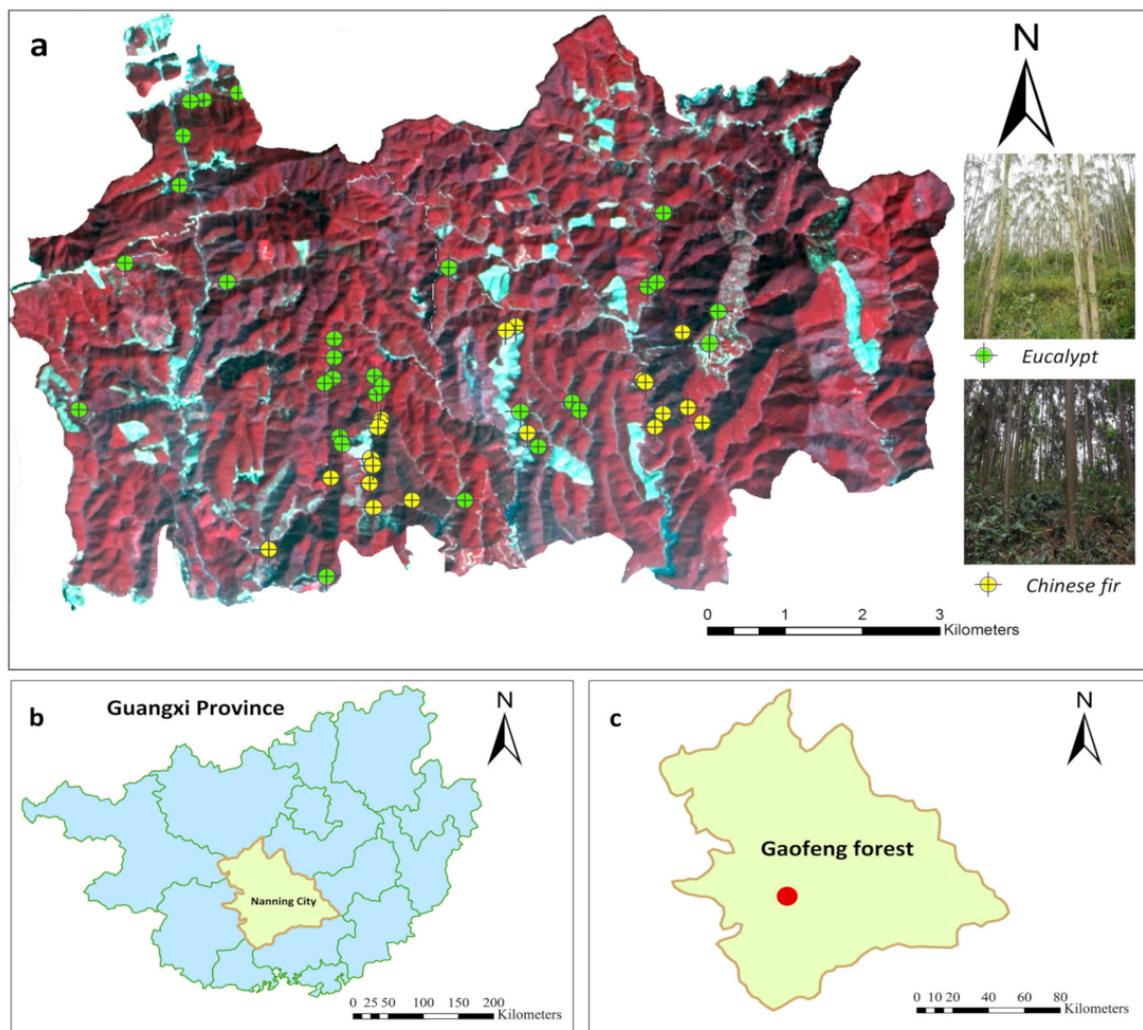

Figure 1. The location of the study site in Gaofeng Forest and sample plots distribution of two main tree species (Chinese fir and Eucalypt). (**a**) The Landsat 8 satellite image (June 2017) of study site; (**b**) the location of Nanning city; (**c**) the location of the Gaofeng forest.

2.2. Field Data

The field surveys for ground data acquisition of the study site was carried out from 16 January to 3 February 2018. 49 fixed-area square sample plots (20 × 20 m) within the study site were used in this research. All plots were divided into those dominated by Chinese fir (n = 20) and those dominated by Eucalypt (n = 29). The center position of each plot and each plot corner were registered using a Trimble juno T41/5 Handheld GNSS (Global Navigation Satellite System) instrument (including a GPS Collector) (Trimble, Sunnyvale, CA, USA), and corrected with high-precision real-time corrected differential post-processing signals received from Trimble NetR9 GNSS Reference Receiver equipped with Trimble Zephyr Geodetic Model 2 GNSS Antenna, resulting in a positional of accuracy of less than 0.5 m [62]. The plot directions were recorded by forest compass, and the border lengths were measured by PI tape. For all live trees with a diameter at breast height (DBH) over 5 cm, species, DBH, height, height to crown base, crown width in both cardinal directions were measured. In addition, small trees (DBH < 5 cm) and dead wood were also tallied for total stem density, but not used in volume calculations.

Table 1 shows a summary of plot-level DBH, mean height (H), stem density (N), basal area (G) and volume (V). The individual tree volume with each plot was calculated according to the local provincial species volume equations, which were based on DBH and tree height as predictive variables. Then, plot-level volume was summed by the calculated tree volume within each plot.

Table 1. Summary of field parameters.

Parameters	Chinese Fir Forests (n = 20)			Eucalypt Forests (n = 29)		
	Range	**Mean**	**SD**	**Range**	**Mean**	**SD**
DBH/cm	11.41–23.34	16.48	3.90	5.52–23.45	12.64	3.96
H/m	9.72–19.56	13.33	2.38	8.01–30.37	15.92	4.57
N/(ha^{-1})	525.00–2450.00	1245.00	523.07	725.00–2725.00	1640.83	576.43
G/(m^2·ha^{-1})	17.01–42.78	26.58	5.36	4.70–46.07	20.61	8.68
V/(m^3·ha^{-1})	108.33–355.54	206.27	50.49	22.85–556.20	180.09	118.78

Notes: SD = standard deviation, DBH = diameter at breast height, H = mean height, N = stem density, G = basal area, V = volume.

2.3. Data Acquisition and Pre-Processing

A workflow for predicting diameter distributions by airborne LiDAR data is shown in Figure 2. The workflow includes three main steps: Data pre-processing, modality differentiation and modeling diameter distributions. In step one, a 1 m Digital Terrain Model (DTM) was generated by calculating the average elevation of LiDAR ground returns within a rasterized cell grid. Metrics were then extracted from point clouds normalized by the DTM. In step two, field plots were classified as either unimodal or multimodal using a bimodality coefficient (BC). Then, the capability of BC, along with the Gini coefficient (GC) and Lorenz asymmetry (LA) predicted by LiDAR metrics were assessed to differentiate modality. In step three, the UWM and FMM were generated for diameter distributions of estimated unimodal and multimodal plots. For comparison, a KNN model was also employed for constructing diameter distributions using LiDAR-derived metrics. KNN diameter distribution prediction strategies, including number of neighbors (i.e., k), response configurations and imputation methods (i.e., MSN and RF) were also examined. Finally, the performance of LiDAR predicted diameter distributions was evaluated. These steps are described in further detail below.
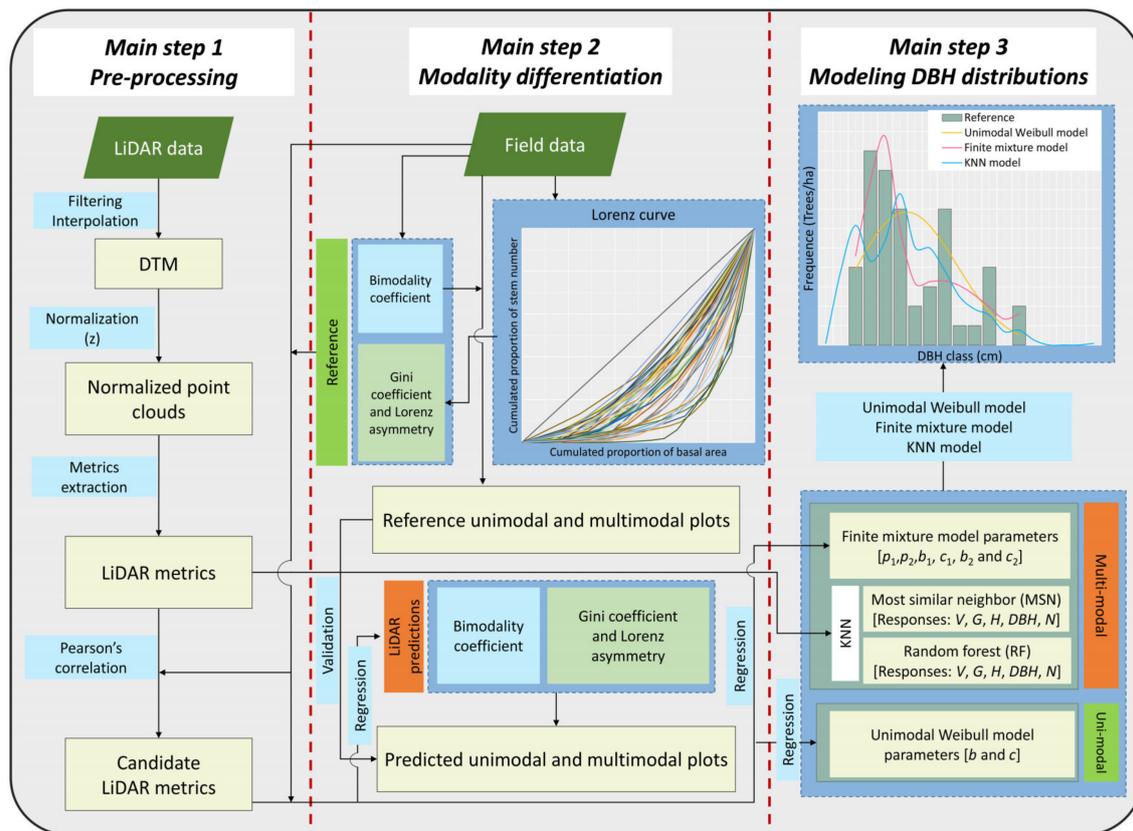
Figure 2. The workflow of estimating diameter distributions in this study. DTM = digital terrain model; V = volume; G = basal area; H = mean height; DBH = diameter at breast; N = stem density; b, c = scale and shape parameters of unimodal Weibull model; p_1, p_2, b_1, c_1, b_2, c_2 = proportion, scale and shape of first component and second of bimodal Weibull in finite mixture model..

2.4. LiDAR Data and Metrics

The small footprint airborne LiDAR data were acquired over the entire study site, initially from 17 to 30 January 2018. The LiDAR data acquisition was operated from a fixed-wing aircraft flown at 750 m above ground level with a flight speed of 180 km/h, using a Riegl LMS-Q680i sensor (Riegl Laser Measurement Systems, GmbH, Horn, Austria) integrated into the LiCHy System [63]. The flight line side-lap was 65% and the sensor recorded a temporal sample spacing of a laser pulse was 1 ns (approximately 15 cm). The emitted laser pulses by LiDAR system were in the near-infrared band (1550 nm) with a 300 kHz pulse repetition frequency and an 80 Hz scanning frequency. The maximum scanning angle and FOV (field of view) were $\pm30°$ from nadir, $60°$, respectively. The average ground point distances of the dataset in flying and scanning direction were both of 0.45 m in a single strip, with a point density of approximately 8.51 pulses/m^2. The average points density on sample plots was 9.58 pulses/m^2 and the average beam footprint size was 0.45 m (nadir) in diameter. Extracted point clouds and associated waveforms were stored in LAS (log ASCII standard) 1.2 format (American Society for Photogrammetry and Remote Sensing, Bethesda, MD, USA).

To obtain tree heights from normalized point clouds, raw point cloud data were first de-noised under the processing of removing outliers. The above-ground return points were filtered using an Improved Progressive TIN Densification (IPTD) filter algorithm adapted from Zhao et al. [64]. After filtering the above-ground points, a 1-m Digital terrain model (DTM) was generated by calculating the average elevation from the ground points within a rasterized cell grid. If there were no returns within cell grids, these cell grids were filled by interpolation using an Inverse Distance Weighted (IDW) algorithm). Then, the point cloud was normalized against the DTM surface height.

Point clouds for all plots (*n* = 49) were extracted using the coordinates of the lower left and upper right corners.

LiDAR metrics were generated from normalized point clouds and were classified into two groups: Standard metrics [65] and canopy metrics [66] (see details in Table A1). Standard metrics mainly involved height-related metrics (e.g., height percentile, mean height, skewness and kurtosis of heights) and canopy return density-related metrics (e.g., density and canopy cover). Canopy metrics included a set of canopy volume (CV) metrics which represent spatial organization of the tree material (e.g., trunk, foliage, etc.) and the total canopy volume within the canopy [67], along with others, such as rumple, coefficient of variation of leaf area density (CvLAD) [68] and vertical complexity index (VCI) [69]. The canopy volume metrics of this study were calculated by a voxel-based canopy volume model (CVM) with a $5 \times 5 \times 1$ m^3 voxel resolution. The metrics used as predictor variables were calculated for the 20×20 m plots using procedures similar to that of Naesset [70] by FUSION [71] and Matlab Software [72].

2.5. Modality Differentiation

Although having relative simple structures compared to other natural forests, subtropical planted forests still show a certain degree of heterogeneity of structural types. To estimate tree diameter distributions of forest stands more accurately than only using a UWM in this heterogeneous forests, prior discrimination of forest structural types (FSTs) and knowledge of appropriate models to pertinently fit diameter distributions is required [48,73,74]. In this study, we attempted to stratify forest stands into two generalized FSTs, i.e., unimodal and multimodal, by examining the modality of diameter distributions for each plot using LiDAR-derived predictions. Referring to Mulverhill et al. [48], we identified modality of sample data distribution each plot using a bimodality coefficient (BC), which reflects a form of the distribution of sample data, to The *BC* is the sum of the squared skewness of a distribution (g) and 1, divided by the sum of the kurtosis (kt) and a correction factor (C), and has the following equation [75]:

$$BC = \frac{g^2 + 1}{kt + C}, \tag{1}$$

$$C = \frac{g^2 + 1}{kt + \frac{3(n-1)^2}{(n-2)(n-3)}}, \tag{2}$$

where *n* the number of samples. *C* varies between 3 and 4, depending on the number of observations *n*, and corrects for the fact that the kurtosis of a normal distribution equals 3. The value of BC for the uniform distribution is 5/9. In this case, values of BC for ground-measured diameter distributions greater than 5/9 will indicate bimodal or multimodal diameter distributions. Thus, total 49 plots were classified into 23 unimodal plots and 26 multimodal plots (Table 2).

Table 2. The results of classified plots using bimodality coefficient for reference measured diameter distributions.

Types	Unimodal Plots	Multimodal Plots
All (*n* = 49)	23	26
Chinese fir (*n* = 20)	6	14
Eucalypt (*n* = 29)	17	12

Previous studies [26,74,76–78] pointed out superiority of Lorenz curves for describing FSTs and suggested their derived indicators as a good alternative for describing tree size inequality. In a forest stand, the Lorenz curve represents dominance relations by comparing relative cumulative proportions of basal area and stem density accounted for each tree [78]. The descriptors of Lorenz curves, the Gini coefficient (GC) and Lorenz asymmetry (LA), were regarded as relevant indicators of forest structure.

These two Lorenz-based indicators have been used as a basis for characterizing tree size inequality to stratify the forest area into homogeneous or heterogeneous stands [77]. The value of GC is equal to half of the relative mean difference in size among all the trees, and it is usually conceptualized as the area comprised between the Lorenz curve and the diagonal line-of absolute-equality [78]. LA describes the skewness of Lorenz curves. Valbuena et al. [74] reported that the values of GC greater than 0.5 and LA less than 0.5 suggest multimodal diameter distributions. Based on this decision rule, GC and LA were as a suite of Lorenz-based indicators for identifying unimodal or multimodal plots in this study. The GC and LA were derived from Lorenz curves generated by 'ineq' in *R* software [79,80].

In an operational context, LiDAR predictions have the capability to extrapolate this modality differentiation from a grid cell to larger spatial extents. Thus, the BC, GC and LA can be predicted by LiDAR metrics and used as estimates of modality differentiation. Finally, LiDAR predictions for modality differentiation were assessed by determining the overall accuracy and Kappa coefficient of the classification. The differentiation accuracy was defined as the percentage of all plots that were correctly classified as either unimodal or multimodal plots. The initial classification in Table 2 was used as a reference.

2.6. Unimodal Weibull Model

As a two-parameter Weibull distribution is greatly adaptive, ranging from an inversed J-shape to unimodal skewed and unimodal symmetrical curve, the Weibull distribution has been shown to be able to characterize a wide range of tree size distributions (e.g., diameter or basal area distribution) in forest stands [40,81]. The two-parameter Weibull distribution is generated by integrating the Weibull density function, and has the following equation:

$$f(x) = \frac{c}{b} \left(\frac{x}{b}\right)^{c-1} exp\left[-\left(\frac{x}{b}\right)^c\right], x \geq 0; b, c \geq 0, \tag{3}$$

where x is the random variable, in this case, the DBH. b and c are the scale and shape parameters of the Weibull density function, respectively. Estimation of Weibull parameters is typically performed by maximum likelihood estimation (MLE), because of its consistently better goodness-of-fit statistics when compared to the previous methods [25,82]. Thus, the MLE method was used for the calculation of the Weibull parameters.

2.7. Finite Mixture Model

A finite mixture model (FMM) is used to describe the multimodal distributions of tree diameter in mixed species or uneven-aged forest stands [46,83]. A mixture distribution can be obtained when analyzing a heterogeneous population [84]. An FMM is essentially an overall distribution made up of two or more smaller distributions, the mixed probability density function p is a weighted sum of k component densities:

$$f(x|\varphi) = \sum_{i=1}^{k} p_i f_i(x|\theta_i), \ 0 \leq p_i \leq 1, \sum_{i=1}^{k} p_i = 1, \tag{4}$$

$$\varphi = \begin{pmatrix} p_1 & \cdots & p_i & \cdots & p_k \\ \theta_1^T & \cdots & \theta_i^T & \cdots & \theta_k^T \end{pmatrix}, \tag{5}$$

where p_i is the relative abundance of the *i*th component as a proportion of the total population and must satisfy the constraints. θ_i denotes the parameters of the distribution, and φ is a complete parameter set for the overall distribution. For this analysis, a bimodal 2-parameter Weibull FMM was evaluated in this study, as follows:

$$f(x|\varphi) = p_1 f_1(x|\theta_1) + p_2 f_2(x|\theta_2). \tag{6}$$

For this case, the FMM is described by six Weibull parameters: Proportion, scale, and shape of each of the two components (i.e., $p_1, p_2, b_1, c_1, b_2, c_2$). The combination of the Expectation Maximization

(EM) algorithm with the Newton-type method was employed for estimating the parameters of the two component mixtures [84]. This procedure started from an initial value in Equation (7):

$$\varphi^0 = \begin{pmatrix} p_1^0 & p_2^0 \\ \theta_1^{0T} & \theta_2^{0T} \end{pmatrix}. \tag{7}$$

The initial values of FMM were set up with the multi-start method (MM) which has the potential to find the global maximum by starting the numerical procedure from a set of grid points covering the data space [16]. The grid points on the data space (10 points were employed) were given by:

$$u_1 = d_{min}, \; u_2 = u_1 + \frac{u_1 - u_{10}}{9}, \cdots, u_9 = u_8 + \frac{u_1 - u_{10}}{9}, u_{10} = d_{max}, \tag{8}$$

where $d_{\min}$, $d_{\max}$ are minimum and maximum values of the DBH in a plot, respectively. Each subset of initial values was given by:

$$\varphi_{c_1,c_2}^0 = \begin{pmatrix} 0.5 & 0.5 \\ \theta_{c_1}^{*0T} & \theta_{c_2}^{*0T} \end{pmatrix}, \tag{9}$$

where $\left(\theta_{c_1}^0 = (u_1,s), \theta_{c_2}^0 = (u_2,s)\right), \cdots, \left(\theta_{c_1}^0 = (u_1,s), \theta_{c_2}^0 = (u_{10},s)\right), \left(\theta_{c_1}^0 = (u_2,s), \theta_{c_2}^0 = (u_3,s)\right),$ $\cdots, \left(\theta_{c_1}^0 = (u_2,s), \theta_{c_2}^0 = (u_{10},s)\right), \cdots, \left(\theta_{c_1}^0 = (u_9,s), \theta_{c_2}^0 = (u_{10},s)\right)$; s is the standard deviation of DBH of all trees in the plot. A complete set of initial values φ^0 consists of 45 φ_{c_1,c_2}^{*0} subsets. The procedure was run 45 times, and then the results were examined to see whether the same solution was produced each run [47]. Once calculated initial values were determined, the FMM was modeled using the 'mixdist' package in the *R* software [79,85].

2.8. KNN Model

2.8.1. The Selection of Number of Neighbors, Response Configurations and Imputation Methods

KNN, a non-parametric approach, assigns the mean of observed response values of the k nearest reference units to the target unit [86]. In this study, diameter distributions could be compiled from tree-level data from the nearest plots [37]. Two nearest neighbor (NN) imputation strategies were tested in this study: Most Similar Neighbor (MSN) and Random Forest (RF). For MSN, a canonical correlation analysis was used to produce a distance metric that determines which k nearest neighbors are most similar to the target observation according to the predictor variables [37,53].

In RF, observations are considered similar if they tend to converge in the same terminal node in a suitably constructed collection of classification and regression trees [54]. The metric used to define statistical distance is calculated as one minus the proportion of trees where a target observation is in the same terminal node as a reference observation [55,87]. Here, the RF imputation was set to generate 500 regression trees ('ntree') in each run, and the number of predictor variables was fixed at the square root of the number of LiDAR metrics (i.e., 'mtry' = $\sqrt{n}$) [58]. In this study, the imputed value was produced as a weighted average over the k neighbors, or as a mode, where the count was the sum of the weights rather than each having a weight of 1, as in the case of a categorical variable [56]. The weight W_{ij} of the reference plot *j* for target plot *i* was determined according to distance D_{ij}:

$$W_{ij} = \frac{\frac{1}{D_{ij}^2}}{\sum_1^k \frac{1}{D_{ij}^2}}, \tag{10}$$

The response variables of imputations used in this study were volume (V), basal area (G), mean height (H), mean diameter at breast height (DBH), and stem density (N), because they have a strong correlation with diameter distributions and can be successfully predicted with LiDAR in previous studies [27,33,37,38]. Combinations of these five responses made up the 15 response sets

listed in Table 3. The effect of the number of neighbors k was investigated at a range of 1, 3, 5, 7, and 9 with the SET1 (V, G, H, DBH, N) response configuration. Once the optimal value of k was determined, a sensitivity analysis of response configurations and imputation methods (i.e., MSN and RF) was carried out. Since RF imputation allows the number of variables to be much larger than the number of training samples [57,88], all LiDAR-derived metrics were used as predictor variables for RF imputation. In order to ensure comparability between the two imputation methods, MSN also employed all metrics. Two imputations were generated by the 'yaImpute' package in *R* [55,79]. In the RF imputation, scaled importance values of Light Detection and Ranging (LiDAR) metrics were calculated by the 'yaiVarImp' function in the 'yaImpute' package [55].

Table 3. Response configurations for k-nearest neighbor (KNN) imputation in this study.

Responses	SET1	SET2	SET3	SET4	SET5	SET6	SET7	SET8	SET9	SET10	SET11	SET12	SET13	SET14	SET15
V	√	√	√	√	√										
G	√	√	√	√		√	√	√	√						
H	√	√	√			√	√	√		√	√	√			
DBH	√	√				√	√			√	√		√	√	
N	√					√				√			√		√

Notes: "√" denotes response variables selected in KNN models. V = volume, G = basal area, H = mean height, DBH = diameter at breast height, N = stem density.

2.8.2. Constructing Diameter Distribution Based on KNN

As nearest neighbors were found by means of KNN imputation, the diameter distributions of reference neighbors can be used to construct a tree list of a target plot. To impute diameter distributions of target plot i, the stem density (N) of reference plot j and weight W_{ij} for the distance between plot i and j were required. In this study, the field measured stem density and KNN imputed N were respectively used for predicting diameter distributions. In addition, the construction of the tree list for the target plot i from the nearest neighbors can be described as follows:

$$T_i = \left\{ W_{i1}t_{j1}, W_{i2}t_{j2}, \cdots, W_{ij}t_{jk} \right\}, \tag{11}$$

where T_i is the KNN imputed diameter distribution of tree list in target i plot, t_{jk} is a list of trees of kth reference plot j. When a tree list of target plot i were constructed based on KNN imputations, the diameter distribution were scaled by N. In this study, diameter distributions were predicted for all tree species (i.e., all plots) by simultaneous KNN imputations and also predicted for different species (i.e., Chinese fir and Eucalypt) by separate KNN imputations.

2.9. Model Creation and Assessment

In this study, we used Pearson's correlation (r) between total LiDAR metrics and parameters of predictive models for selecting optimal metrics. The metrics were chosen as candidate metrics to enter into regression analyses when r is greater than 0.6 [89]. To improve model linearity all of the dependent variables (i.e., BC, GC, LA and parameters of UWM and FMM) and independent variables (i.e., LiDAR-derived metrics) were converted into a form of natural logarithm in the regression analyses, due to the fact that this functional form could provide more accurate estimates of forest structure attributes [70]. In addition, this form model with a positive intercept will lead to non-zero estimates of structural attributes compared with some linear models with negative intercepts when developed models are extrapolated to larger scales. Then all variables were back-transformed from natural logarithm to original arithmetic units using a bias correction factor (BCF) [90].

Regression models were constructed as non-specific models (i.e., all plots), species-specific models (i.e., Chinese fir plots and Eucalypt plots), and structure-specific models (i.e., classified unimodal plots and multimodal plots). Independent variables were left in the model using an *F*-test with a $p < 0.05$ significance level [91]. To avoid a high inter-correlation of independent variables, a Principal Component Analysis (PCA) based on the correlation matrix was employed. To reduce the collinearity

of regressions, the models with low condition factor ($k < 30$) based on PCA were accepted in this study [70]. To avoid overfitting, the maximum number of predictor variables in each regression model was set to four. The best-fitting models were eventually selected according to the lowest Akaike information criterion (AIC) value [92]. The coefficient of determination (R^2), Root-Mean-Square-Error (RMSE) and relative RMSE (rRMSE) were employed to evaluate the performance of predictive models for UWM and FMM parameters. For non-parametric model, all metrics were used as predictors for predicting KNN response sets in MSN and RF imputation. The performance was also assessed using R^2, RMSE, and rRMSE. Due to the limited field data available, predictive accuracies were validated by a leave-one-out cross-validation analysis (LOOCV).

2.10. Evaluating Fit of Predicted Diameter Distributions

The goodness of fit between the reference and predicted diameter distributions on each plot was evaluated by Reynolds error index (e_R) [93] and Packalén error index (e_P) [27]. The calculations of two error indices are as follows:

$$e_R = \sum_{i=1}^{m} 100 \left| \frac{f_{\text{ref } i} - f_{\text{LiDAR } i}}{N_{\text{ref}}} \right|, \tag{12}$$

$$e_P = \sum_{i=1}^{m} 0.5 \left| \frac{f_{\text{ref } i}}{N_{\text{ref}}} - \frac{f_{\text{LiDAR } i}}{N_{\text{LiDAR}}} \right|, \tag{13}$$

where $f_{\text{ref } i}$ is the reference and $f_{\text{LiDAR } i}$ is the LiDAR-estimated stem frequency in DBH class i, m is the total number of classes, N_{ref} is the reference stem density, and N_{LiDAR} is estimated stem density which is equal to the total frequency in each DBH class. Lower values of error indices indicate a better fit of predicted diameter distributions. e_R values range from 0 to 200, and the e_p values range from 0 to 1.

3. Results

3.1. Differentiation of Modality

LiDAR predictions (i.e., LiDAR-predicted BC, GC, and LA) were respectively used as means of differentiating between unimodal and multimodal plots. The results of this differentiation are presented in Table 4. For LiDAR predictions of modality differentiation predictors, models for all plots and species-specific plots were separately developed. In comparison, predictions of the non-specific models (i.e., all plots) of BC, GC and *LA* had relatively lower accuracy than the species-specific models, indicating that the stratification approach according to forest species types may improve the estimates of parameters. GC models had the highest prediction accuracy, followed by LA and BC. For all modality differentiation predictors, both stand metrics (e.g., h$_{50}$, h$_{cv}$) and canopy metrics (e.g., Oligo, CvLAD, VCI) were selected by each predicted models. Notably, all of CG, Oligo and CvLAD were the most frequently selected metrics which separately selected by three out of nine models while most other metrics were selected at least twice in nine models. In addition, vertical distribution dependent metrics, such as L$_{\text{kurt}}$, Oligo and CvLAD contributed to the prediction of distribution-related BC.

In terms of these LiDAR predictions for differentiating modality, the predicted Lorenz-based indicators (i.e., GC and LA) produced higher classification accuracies than predicted BC. In addition, the species-specific predicted GC and LA provided the highest classification accuracy, whereas the predicted BC of the non-specific model produced the lowest classification accuracy in terms of overall accuracy and Kappa coefficient. Thus, final unimodal or multimodal plots were classified by LiDAR-predicted GC and LA from the species-specific model.

Table 4. Performance of modality differentiation using Light Detection and Ranging (LiDAR) predictions.

Classification Methods	Forest Types		Predictors	Overall Accuracy (%)	Kappa Coefficient	Prediction Accuracy			Metrics Used
						R^2	RMSE	rRMSE (%)	
LiDAR predictions	All		BC	48.98	0.45	0.28	0.09	37.85	h_{cv}, L_{kurt}, d_3, Oligo
			GC+LA	59.18	0.59	GC: 0.52	0.11	20.95	h_{max}, h_{CURT}, d_9, CG
						LA: 0.35	0.10	35.80	h_{50}, h_{75}, d_9, Rumple
	Species-specific	Chinese fir	BC	65.31	0.50	0.72	0.06	14.92	h_{95}, h_{var}, Oligo, CG
		Eucalypt				0.32	0.10	32.30	h_{50}, h_{75}, CRR, CvLAD
		Chinese fir	GC + LA	77.55	0.55	GC: 0.90	0.06	10.78	h_{75}, h_{mean}, L_{kurt}, Eu
						LA: 0.45	0.08	34.29	h_{95}, h_{var}, Oligo, CG
		Eucalypt				GC: 0.68	0.09	17.48	h_{cv}, CRR, Close, CvLAD
						LA: 0.56	0.08	28.96	L_{kurt}, CvLAD, Rumple, VCI

Notes: BC = Bimodality coefficient, GC = Gini coefficient, LA = Lorenz asymmetry. See Table A1 for metrics details.

3.2. Unimodal Weibull Model and Finite Mixture Model

3.2.1. Predictive Models

As shown in Table 5, similar to LiDAR-predicted models of BC, GC and LA, the plots with undifferentiated modality were stratified by forest tree species for model development of Weibull parameters (i.e., b and c) derived from a UWM model. As a result, species-specific models yielded better prediction accuracies than the non-specific model (all plots). Two suites of species-specific models performed similarly for predicting b and c. Using predicted GC and LA for identifying modality of plots, 21 unimodal plots and 28 multimodal plots were differentiated (Table 5). The forest type ratio (FTR) of Chinese fir to Eucalypt for 28 unimodal plots was 7/21 and FTR for 21 multimodal plots was 13/8, indicating that most of the Chinese fir plots had uneven tree size distributions whereas most of Eucalypt plots had relative even-sized diameter distributions. UWM and FMM parameters (respectively derived from reference diameter distributions of identified unimodal and multimodal plots) were predicted by LiDAR metrics. The R^2, RMSE and rRMSE of predictive b and c for identified unimodal plots ranged from 0.83 to 0.87, 0.71 to 1.74 and from 12.26 to 15.21%, respectively. Regarding the predictions of bimodal Weibull parameters, predictive accuracies of the other five parameters except for b_2 were lower than unimodal b and c. However, scale parameters (i.e., b, b_1 and b_2) produced notably better performance than shape parameters (i.e., c, c_1 and c_2) regardless of FMM or UWM models. Overall, the predicted second component parameters ($R^2 = 0.61$–0.83, rRMSE = 12.22%–26.66%) of bimodal Weibull model performed better than the first component parameters expect for c_2 obtained a relative lower R^2 value of 0.61 with a larger RMSE value of 47.42.

Each predictive model included standard metrics, and eight of the fourteen models included canopy metrics, except for b and c of UWM for all plots and Eucalypt plots, c of differentiated unimodal plots and b_2 of multimodal plots. Specifically, the c model of all plots used the same set of metrics (i.e., d_1, d_3, d_7, d_9) that were selected for c model of Eucalypt plots. The most frequently selected height-related metric was h_{max}, by six out of 14 models. d_5 and d_7 were the most frequently used density-related metrics, selected by four out of fourteen models. Among all models, the most frequently used canopy metrics were CvLAD and Rumple.

Table 5. A summary statistics for predictive models.

	Forest Types	FTR	Predictors	Metrics Used	R^2	RMSE	rRMSE (%)
Undifferentiated modality	All (*n* = 49)	20/29	b	h_{95}, h_{var}, d_5, CRR	0.80	2.26	14.42
			c	d_1, d_3, d_7, d_9	0.66	0.97	21.99
	Chinese fir (*n* = 20)	/	b	h_{50}, h_{cv}, h_{std}, Oligo	0.94	1.04	5.65
			c	h_{75}, h_{mean}, OG, CvLAD	0.80	0.55	16.39
	Eucalypt (*n* = 29)	/	b	h_{var}, h_{CURT}, L_{cv}, d_5	0.95	0.99	7.17
			c	h_{25}, h_{mean}, h_{max}, d_7	0.75	0.74	14.94
Differentiated modality	Unimodal plots (*n* = 28)	7/21	b	h_{max}, L_{cv}, CvLAD, VCI	0.87	1.74	12.26
			c	d_1, d_3, d_7, d_9	0.83	0.71	15.21
	Multimodal plots (*n* = 21)	13/8	p_1	h_{max}, Close, Rumple, VCI	0.58	0.15	29.21
			p_2	h_{max}, CRR, Eu, Rumple	0.71	0.13	25.89
			b_1	h_{mean}, h_{max}, d_5, Filled	0.77	2.49	18.74
			c_1	h_{95}, h_{max}, h_{kurt}, CvLAD	0.71	18.29	20.53
			b_2	h_{50}, h_{95}, d_5, d_7	0.83	2.44	12.22
			c_2	h_{mean}, L_{cv}, d_9, Close	0.61	47.42	26.66

Notes: FTR = the forest type ratio of Chinese fir plots to Eucalypt plots; b and c = scale and shape of unimodal Weibull distribution; p_1 and p_2 = proportion of first component and second component of finite mixture distribution; b_1 and c_1 = scale and shape of first component of finite mixture distribution; b_2 and c_2 = scale and shape of second component of finite mixture distribution. See Table A1 for metric details.

Accuracy results for LiDAR-predicted UWM parameters in undifferentiated plots using leave-one-out cross-validation are shown in Figure 3. The b and c for all plots had the lowest accuracies,

whereas the b and c of Chinese fir models and that of Eucalypt models produced similar accuracies to one another. Figure 4 shows the cross-validation results of FMM parameters models in differentiated plots. For identified UWM plots, b and c produced same R^2 values of 0.83, with RMSE of 12.73% and 14.76%, respectively. For FMM parameters, b_2 obtained the highest accuracy, followed by b_1, c_1, p_2, c_2, and p_2. Overall, the scale parameter models performed better than shape parameter models, indicating LiDAR metrics had a relatively high correlation with scale parameter.
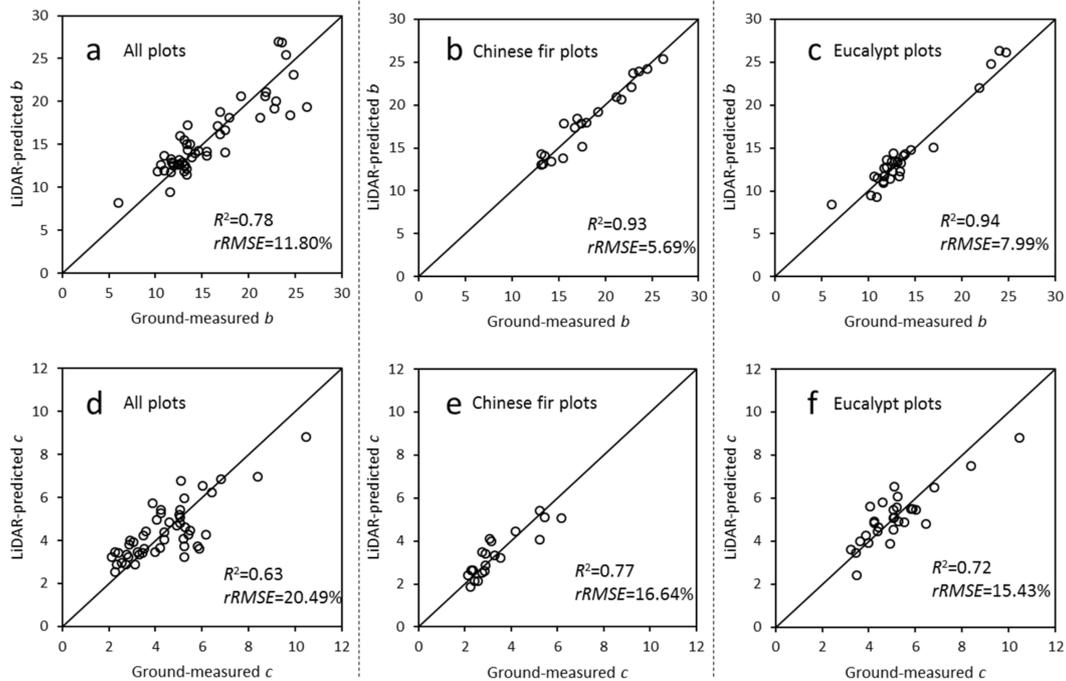
Figure 3. Scatter plot of cross-validations for the ground-measured unimodal Weibull model (UWM) parameters vs. the LiDAR-predicted UWM parameters in undifferentiated plots. (**a**,**e**) all plots; (**b**,**f**) Chinese fir plots; (**c**,**g**) Eucalypt plots.
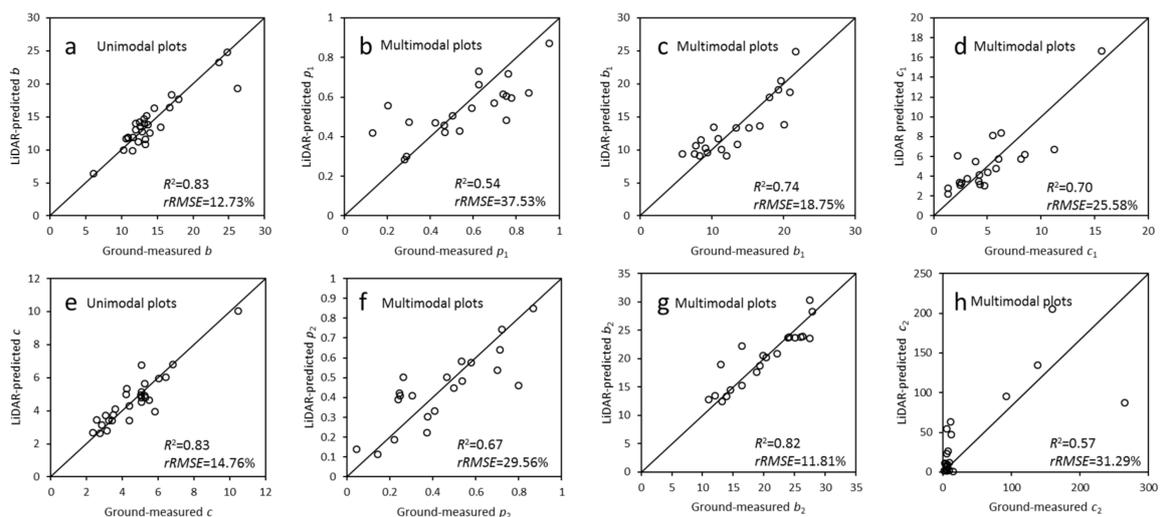
Figure 4. Scatter plot of cross-validations for the ground-measured UWM parameters and finite mixture model (FMM) parameters vs. the LiDAR-predicted UWM parameters and FMM parameters in differentiated unimodal plots and multimodal plots. (**a**,**e**) scale and shape of UWM of differentiated unimodal plots; (**b**,**f**) proportion of differentiated multimodal plots; (**c**,**g**) scale parameters of FMM of differentiated multimodal plots; (**d**,**h**) shape parameters of FMM of differentiated multimodal plots.

3.2.2. Predicted Parametric Diameter Distributions

Based on the procedure of differentiating modal plots and predicting UWM and FMM parameters with LiDAR, we compared ground-estimated diameter distributions to LiDAR-predicted diameter distributions by using mean values of the two error indices for assessment. The results for different plot types are presented in Table 6. Diameter distributions of undifferentiated plots (Model 1–3) generated by UWM from either ground estimations or LiDAR predictions had a poorer accuracy than those of differentiated unimodal and multimodal plots (Model 4–6). For the two tree species, diameter distributions of Eucalypt plots (Model 3) performed better than those of Chinese fir plots (Model 2). After differentiating the modality of plots, models of multimodal plots (Model 6) produced more accurate diameter distributions, whereas identified unimodal plots (Model 5) had slightly less accurate diameter distributions. Overall, LiDAR-predicted diameter distributions had relative higher errors than ground-estimated diameter distributions, but the accuracies of LiDAR-predicted distributions were acceptable. For differentiated plots, FMM (Model 6) produced higher accuracies than UWM.

Table 6. Evaluation results of UWM and FMM models for predicting diameter distributions.

Differentiation	Methods	Models Id	Plot Types	Ground-Estimated Diameter Distributions		LiDAR-Predicted Diameter Distributions	
				Mean e_R	Mean e_P	Mean e_R	Mean e_P
Undifferentiated plots	UWM	Model 1	All plots (n = 49)	33.85	0.17	52.31	0.26
		Model 2	Chinese fir plots (n = 20)	42.15	0.21	44.98	0.22
		Model 3	Eucalypt plots (n = 29)	28.12	0.14	41.36	0.21
Differentiated plots	GM	Model 4	All plots (n = 49)	25.28	0.13	40.85	0.20
	UWM	Model 5	Unimodal plots (n = 28)	26.76	0.13	41.36	0.21
	FMM	Model 6	Multimodal plots (n = 21)	21.65	0.11	39.98	0.20

Notes: UWM = unimodal Weibull model; FMM = finite mixture model; GM = general model for differentiated all plots; e_R = Reynolds error index; e_P = Packalén error index.

3.3. KNN Model

For MSN and RF imputations, a total of 15 response configurations (Refer to Table 3) were originally compared in different forest types (i.e., all plots, Chinese fir plots and Eucalypt plots). The effect of number of neighbors k was evaluated with SET1 and the results are presented in Table 7. It intuitively found that the values of error indexes were the lowest when k = 5 neighbors for both imputations. The accuracy of predicted diameter distributions from 15 response configurations by MSN and RF imputations are presented in Figure 5. Overall, there was a relatively high difference in mean e_R values of MSN-imputed distributions among 15 response configurations compared with RF-imputed distributions. This suggested that the MSN imputation was more sensitive to the selection of response configurations, while RF was much less dependent on the selection of response configurations. As shown in Figure 5 and Table 8, SET14 (i.e., DBH) produced the lowest mean e_R value for all plots by RF imputation, whereas SET14 also produced the lowest mean e_R value for Eucalypt by MSN imputation. The optimal response configuration for Chinese fir was SET7 (i.e., G, H and DBH) when distribution was imputed by the RF imputation. Table 8 summarizes the optimal response configurations for all plots, Chinese fir and Eucalypt plots, the cross-validation results of imputed responses, and the accuracies of predicted diameter distributions. Regarding the responses, DBH of SET14 imputed by RF imputation produced higher accuracy than DBH of SET14 imputed by MSN imputation. In SET7, H had the highest accuracy (R^2 = 0.87, rRMSE = 5.40%), followed by DBH, and G. The LiDAR-predicted distribution of Eucalypt produced the highest accuracy of diameter distribution, followed by Chinese fir, and all plots.

Table 7. The selection results of neighbor number *k* for MSN and RF imputation with SET1 (V, G, H, DBH, N).

k	MSN		RF	
	Mean e_R	**Mean e_P**	**Mean e_R**	**Mean e_P**
1	62.16	0.31	58.08	0.29
3	56.38	0.28	55.61	0.26
5	53.86	0.27	54.13	0.27
7	54.64	0.28	56.28	0.28
9	64.11	0.35	60.10	0.30

MSN = Most Similar Neighbor; RF = Random Forest; e_R = Reynolds error index; e_P = Packalén error index.
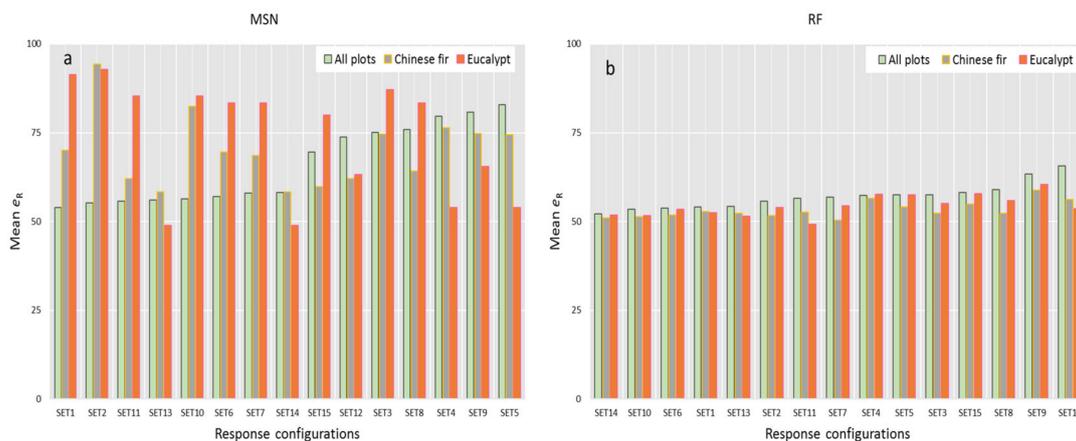
Figure 5. Reynolds error index for diameter distributions of different forest types by MSN (**a**) and RF (**b**) imputations of KNN with 15 response configurations. The response configurations in ascending order of the mean e_R values of all plots; e_R = Reynolds error index.

Table 8. The optimal response configurations and imputation methods for KNN imputed diameter distributions of different forest types and corresponding cross-validation results of the R^2 and rRMSE for selected responses. e_R = Reynolds error index; e_P = Packalén error index.

Method	Model Id	Plot Types	Imputations	Response Configurations	Responses	*CV-R^2*	rRMSE (%)	Mean e_R	Mean e_P
KNN	Model 7	All plots (*n* = 49)	RF	SET14	DBH	0.87	11.41	52.19	0.26
	Model 8	Chinese fir plots (*n* = 20)	RF	SET7	G	0.51	25.09	50.37	0.25
					H	0.87	5.40		
					DBH	0.70	10.25		
	Model 9	Eucalypt plots (*n* = 29)	MSN	SET14	DBH	0.85	10.10	48.93	0.24

Figure 6 displays the scaled importance of top ten predictors that were used for two RF imputation models of Table 8. For the RF-imputed DBH model of all plots (Figure 6a), d_3 was the most important variable in the DBH model, followed by the h_{max}, CvLAD, VCI, Filled, OG, Eu, CC_{2m} and CG. d_1 was the least important variable in the top ten variables. For SET7 responses employed in Chinese fir plots, the h_{75} was the most important metric (Figure 6b), followed by the d_1, OG, VCI, hcv, d_3, h_{50}, Filled, h_{max} and CvLAD. For two RF imputation models, except Eu, CC_{2m} and CG in Figure 6a, the other metrics were ranked among the top ten metrics (Figure 6b).
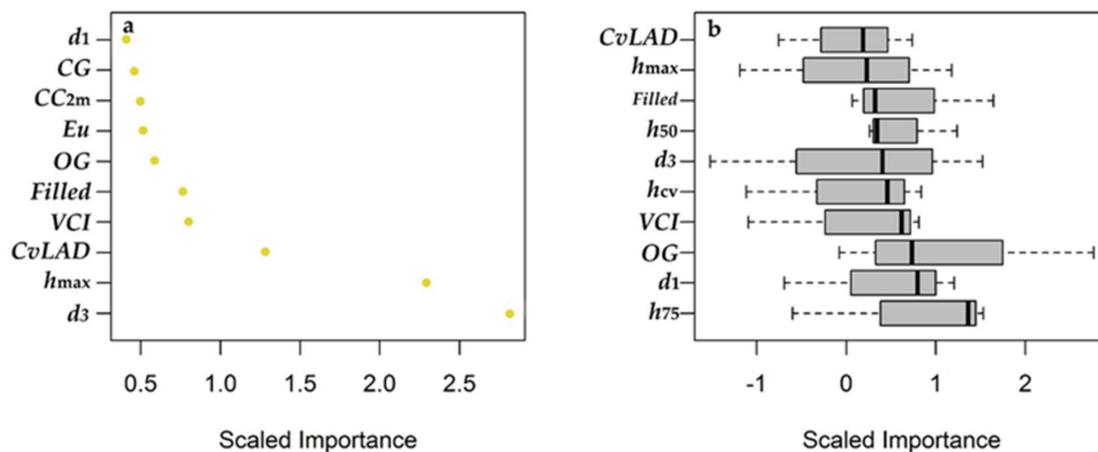
Figure 6. Scaled importance values of Light Detection and Ranging (LiDAR) metrics for RF imputations of all plots and Chinese fir plots. Scaled values were derived by calculating z-scores from the importance values.

3.4. Predicted Diameter Distribution in Different Stem Densities

This study compared three methods for predicting diameter distributions. In terms of the accuracy of diameter distributions, the FMM produced more accurate results than KNN, while UWM produced the lowest accuracies (Tables 6 and 8). To investigate effects of different stem density for predicting distributions, we chose two groups of four sample plots (i.e., Chinese fir plot 1, 2, 3, 4: Cf1, Cf2, Cf3, Cf4; Eucalypt plot 1, 2, 3, 4: Euc1, Euc2, Euc3, Euc4), including two species with relatively low and high stem density to compare the diameter distributions derived from parametric and non-parametric methods. The two groups were separately differentiated as unimodal plots and multimodal plots by LiDAR-predicted Lorenz-based indicators. For four differentiated unimodal sample plots, the diameter distributions were predicted by Model 1 (i.e., UWM of undifferentiated all plots in Table 6), Model 5 (i.e., UWM of differentiated unimodal plots in Table 6) and Model 7 (i.e., KNN model for all plots in Table 8). The Lorenz curves and the predicted diameter distributions results are shown in Figure 7. For the four differentiated multimodal sample plots, the diameter distributions predicted by Model 1, Model 6 (i.e., FMM for differentiated multimodal plots) and Model 7, and their Lorenz curves are shown in Figure 8. From Figure 7a, the predicted Lorenz curves of unimodal plots lied above the line of perfect uniformity (red dotted line), whereas multimodal ones (Figure 8a) were situated below it, indicating larger inequalities among tree sizes in multimodal plots. Overall, the nonparametric predictions (Model 7) were more irregular than the parametric predictions (Model 1, Model 5 and Model 6). In addition, we found that the ground and LiDAR-predicted diameter distributions were skewed to the right and became more positive kurtosis shapes when stem density increased (Figures 7 and 8). Regarding the performance of predicted distributions of these two group plots, the Model 5 and Model 6 performed the best in each group. As shown in Figures 7a and 8a, generally, the error index exhibited a slightly decreasing trend with increasing stem density, indicating diameter distributions with high stem density may be more easily be fitted by three methods. Unlike Model 1 for unimodal distributions, Model 6 and Model 7 seems more suitable for fitting multimodal distributions. Hence, for multimodal plots, the performance of Model 6 and Model 7 was better than Model 1, except for Model 7 for Cf3 and Euc3. The KNN could produce even more multimodal peaks compared with FMM, but the KNN seemed unstable for characterizing bimodal distributions in our study site. The ground distributions of Cf1 and Cf2 were relatively irregular and tended to be multimodal, which may be caused by differentiation errors brought by errors of LiDAR-predicted Lorenz-based indicators (GC and LA). In fact, Cf1 and Cf2 should be classified as multimodal plots, whereas LiDAR predictions identified them as unimodal plots.
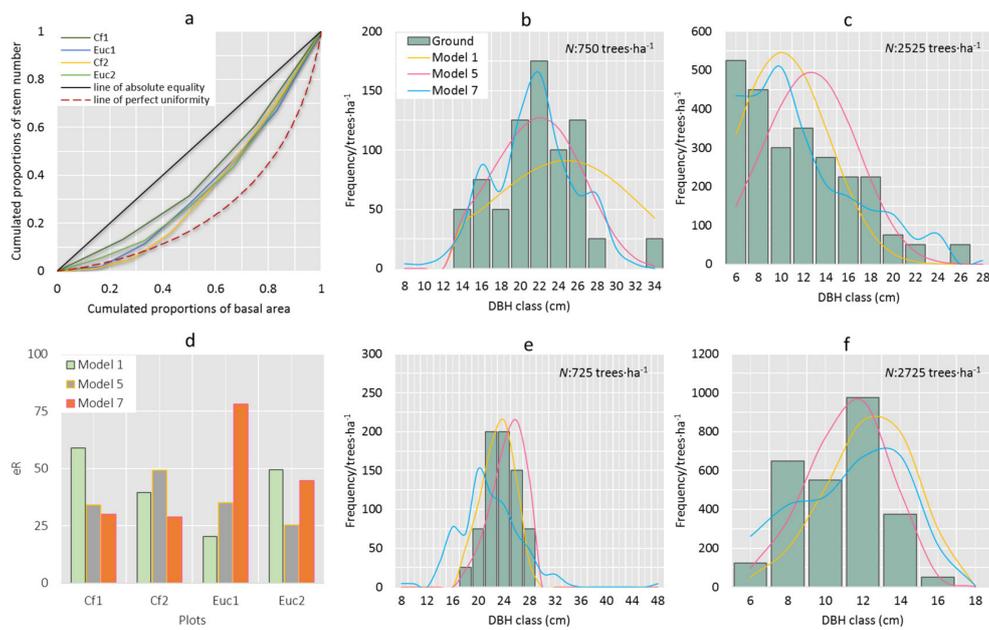
Figure 7. The ground diameter distributions of four differentiated unimodal sample plots with relative low (**b**,**e**) and high stem density (**c**,**f**) predicted by Model 1, Model 5 and Model 7. (**a**) The LiDAR-predicted Lorenz curves of four sample plots; (**b**,**c**) Chinese fir plots; (**e**,**f**) Eucalypt plots; (**d**) The Reynolds error index of predicted diameter distributions of four sample plots. Notes: Cf1 and Cf2 = Chinese fir plot 1 (**b**) and 2 (**c**); Euc1 and Euc2 = Eucalypt plot 1 (**e**) and 2 (**f**); Model 1 = UWM for undifferentiated all plots; Model 5 = UWM for differentiated unimodal plots; Model 7 = KNN model for all plots.
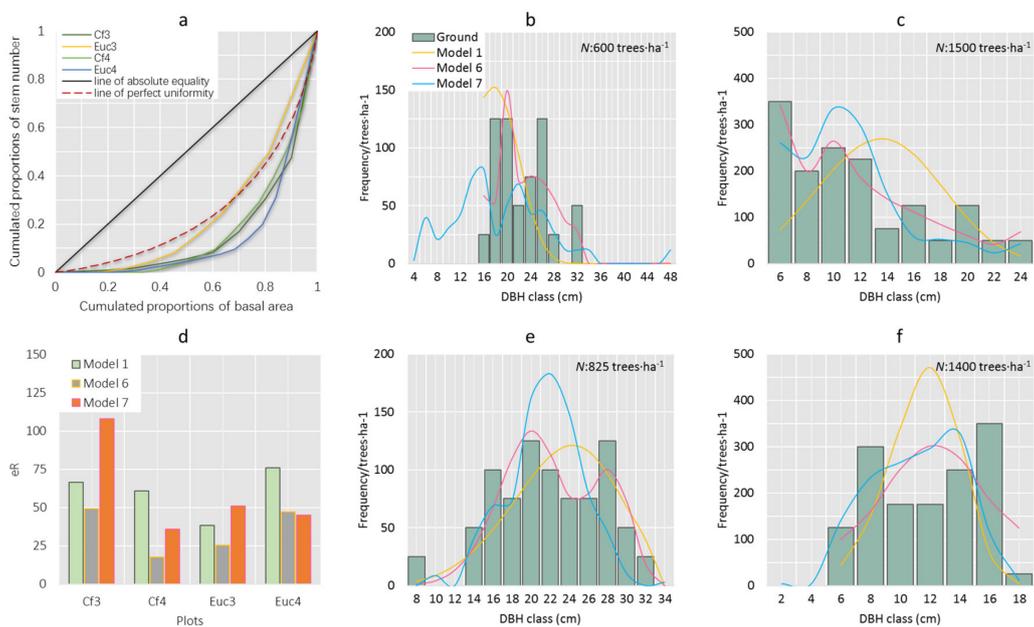
Figure 8. The ground diameter distributions of four differentiated multimodal sample plots with relative low (**b**,**e**) and high stem density (**c**,**f**) predicted by Model 1, Model 6 and Model 7. (**a**) The Lorenz curves of four sample plots; (**b**,**c**) Chinese fir plots; (**e**,**f**) Eucalypt plots; (**d**) The Reynolds error index of predicted diameter distributions of four sample plots. Notes: Cf3 and Cf4 = Chinese fir plot 3 (**b**) and 4 (**c**); Euc3 and Euc4 = Eucalypt plot 3 (**e**) and 4 (**f**); Model 1 = Unimodal Weibull model for undifferentiated all plots; Model 6 = FMM for differentiated multimodal plots; Model 7 = KNN model for all plots.

4. Discussion

4.1. Differentiation of Modality

Assessing the FSTs of forest stands is critical for understanding tree competition and succession [26,94], accessing silvicultural treatments [73,76] and forest management planning [74]. As a key driver of forest ecosystem process [23,95], FSTs are diversified structures induced by either natural forest regeneration or silvicultural practices [96]. Tree diameter inequality, referring to tree diameter diversity interpreted by differently-shaped diameter distributions [9,76], has been shown to be an important feature for characterizing the structural diversity of diameter distributions [76,97]. Prior knowledge of FSTs is a prerequisite to reliably predict tree diameter distributions of forest stands. In this study, we stratified FSTs into unimodal and multimodal stands by means of examining diameter dispersion (i.e., modality of distributions) using concise indicators of describing tree size inequality.

Among a variety of indicators, the GC has been regarded as a consistent descriptor of tree size inequality, outperforming other forest structure indicators characterizing tree size heterogeneity [26]. Lexerød and Eid [76] compared the reliability of a variety of indicators for diameter diversity in boreal forests and they found the GC was the most suitable indicator for assessing stand heterogeneity. Bollandsås and Næsset [14] also used GC to stratify forest areas into different diameter distribution types. However, previous studies [78,98] demonstrated that the same values of GC can be obtained by very different Lorenz curves. For this reason, LA related to the skewness of diameter distributions was developed as a complement to GC in describing tree size inequality [74,78,99]. This study assessed ground-measured and LiDAR-predicted GC and LA derived from Lorenz curves and another statistical indicator, BC, to differentiate unimodal and multimodal stands. The results showed that the LiDAR-predicted Lorenz-based indicators produced higher accuracies (overall accuracy = 59.18%–77.55%) of modality differentiation than BC (overall accuracy = 48.98%–65.31%). In regard to the performance of LiDAR predictions (Table 4), Lorenz-based indicators generally were predicted with relatively high accuracies (R^2 = 0.35–0.90, rRMSE = 10.78%–35.80%) compared to BC (R^2 = 0.28–0.72, rRMSE = 28.96%–35.80%). The reason for the results may be that Lorenz-based indicators simultaneously included more information of both distributions of stems and their basal areas and therefore were more accurately predicted by LiDAR metrics. This may also explain why Lorenz-based indicators were more appropriate for modality discrimination. Furthermore, the results revealed that LiDAR metrics had more explanatory power for predicting GC than LA, which was similar to the results reported by previous studies [9,74,77]. A reason explaining this result may be the fact that GC could bear more variabilities than LA. Valbuena et al. [97] point out the GC was employed to interpret the composition of vertical strata in a forest and LA described the balance between overstory and understory whereas the balance with low variance was relatively difficult to be characterized by LiDAR metrics. In this study, Lorenz-based indicators from the species-specific model (overall accuracy = 77.55%) were as discrimination indices, which may have a disadvantage with regard to limit the application of this method in this study since we need to know the species of the forest for more accurate modality differentiation. The LiCHy System used for remote data acquisition in this study simultaneously integrates LiDAR (RIEGL LMS-Q680i), CCD camera (DigiCAM-60) and hyperspectral (AISA Eagle II) sensors [63], thus accurate classification of tree-species using airborne LiDAR and hyperspectral data in this study should be considered for future work. In addition, more accurate non-specific models (all plots) could be achieved by using either more model forms or using more LiDAR metrics from vertical structure profiles.

Chinese fir plots made up most of the stands differentiated as multimodal (FTR = 13/8 in Table 5), indicating that most Chinese fir plots had multimodal or irregular tree size distributions, whereas most Eucalypt plots had relatively even-sized diameter distributions. This may be caused by different silvicultural treatments in different forest types. As cold-tolerant and commercial fast-growing tree species, Eucalypt played an increasingly important role in plantation forestry in southern China after 1990 [100]. Eucalypt planted forest was highly intensive managed with a short rotation, thereby a clear

cutting was often chosen as the harvest method for Eucalypt plantation and the even-aged seedlings were planted after large area cutting, which resulted in even-sized or even-aged forest structure with regular or unimodal diameter distributions. Compared to Eucalypt, forests planted with Chinese fir have a longer rotation age and succession and regeneration of stands often undergo periodic commercial thinning operations, which increases tree size inequality and results in more uneven-sized stands with multimodal or irregular diameter distributions.

4.2. Predictive Models

Multiple regression analysis using LiDAR-derived metrics as independent variables for establishing predictive models for BC, Lorenz-based indicators and parameters of UWM and FMM. A logarithmic transformation was used, as it was found to be suitable for the estimates of forest structure attributes in previous studies [70,101,102]. However, the regression analysis method may involve overfitting the models for small sample data. Thus, the developed FMM for multimodal plots did not consider tree species separation in this study. As shown in Tables 4 and 5, BC, GC, LA, b and c were predicted by LiDAR-derived metrics in non-specific models (i.e., all plots) and species-specific models. Other studies reported that forest type had a positive impact on predictive models for improving estimates of forest structural attributes [68,89,103,104]. Similarly, the species-specific models obtained higher performance (R^2 = 0.44–0.88, rRMSE = 5.13%–28.42%) than the general models (R^2 = 0.44–0.88, rRMSE = 5.13%–28.42%) in this study. This can likely be explained by the increasing variability in all plots, including multispecies compared with single tree species plots. Notably, the species-specific models for extrapolation to larger scale also need to acquire more accurate tree species classification maps as a prerequisite in future work. For the predictive models of BC, GC, LA, b and c (Tables 4 and 5), Chinese fir models generally performed better than Eucalypt models, although b and c of Chinese fir models obtained similar performance to that of Eucalypt models. This may be likely due to the fact that the relationships between forest structure and estimated structural attributes are usually species dependent, and vertical structures of Chinese fir forests are easily characterized, due to their relatively simple and even conical crown structures. In addition, due to more open overstory in Chinese fir forests, LiDAR-derived metrics could capture more full vertical structure information of understory, which may be a potential complement for performance improvement. More variation was explained with the inclusion of canopy metrics to Chinese fir models. The potential of canopy metrics in enhancing an understanding of the physical characteristics of forest canopy structure and improve estimates of structure attributes has been reported in previous studies [68,89]. As shown in Tables 4 and 5, all Chinese fir models included canopy metrics, which may be another reason explaining slightly better performance in Chinese fir models. Rumple has been demonstrated to be used for characterizing the horizontal complexity of the canopy surface [102,105]. CvLAD metric contains more valuable distribution information on the forest vertical structure [68]. Due to this high correlation with a vertical structure, the CvLAD and Rumple were the most frequently selected in predictive models, indicating their high sensitivity to predicted parameters. Notably, the CvLAD was separately selected by predictive models of BC, GC, LA, b, c and c_1, indicating moderate applicability and universality of this metrics for estimating vertical forest structural attributes.

Similar to the results by other studies in boreal forests [36,48,102], LiDAR-derived metrics had stronger predictive capabilities for scale parameters (i.e., b, b_1 and b_2) (R^2 = 0.77–0.95, rRMSE = 5.65%–18.74%) than shape parameters (i.e., c, c_1 and c_2) (R^2 = 0.61–0.83, rRMSE = 12.22%–26.66%) regardless of UWM or FMM. The shape parameter is responsible for overall distribution form, which could be an exponential, normal, Weibull or inverse-J distribution, while the scale parameter can control the range of overall distribution. However, the shape of the distribution is often more difficulty characterized by LiDAR metrics than the scale parameter. Similar to the results reported by Thomas et al. [36] and Mulverhill et al. [48], the second component parameters of FMM in this study were more accurately predicted than the first component parameters. A reason explaining this result

may be the fact that more mature forest leads to more peaks of the second component of diameter distributions in larger DBH class, which could be easily characterized. Comparing to the performance of LiDAR-predicted parameters of UMM and FMM in Thomas et al. [36] and Mulverhill et al. [48], our results had slightly lower error index accuracy. The cause is likely due to different forest structure in planted forests rather than in unmanaged natural forests.

4.3. The Selection of Number of Neighbors, Response Configurations and Comparisons of MSN and RF Imputations

In this study, a KNN model was employed to predict tree diameter distributions by tree species. To obtain more accurate predictions of diameter distributions and stand attributes, we examined the effects of three imputation strategies, including number of neighbors, response configurations and distance metric (i.e., imputation methods). To shed light on the implication of number neighbors k on fitting diameter distributions, we chose five distribution-related forest structure attributes (i.e., V, G, H, DBH and N) as response variables and carried out a sensitivity analysis of k with a range of 1, 3, 5, 9. After determining the optimal value of k, a total of 15 response sets were examined for their sensitivity to MSN and RF imputations (Table 3). Our results revealed that k = 5 obtained slightly smaller error indices, which was similar to the results reported by Vauhkonen et al. [57]. The optimal k usually depends on the available sample size [57]. Large k shifts the prediction toward the sample mean and may also increase the prediction accuracy [56]. However, for a small sample size, a KNN method with too large or too small k values may lead to less accurate results [49,59]. For KNN imputation by tree species, the total number of plot data (*n* = 49) in this study still seems to be small, thereby this may explain why the values of error indexes showed a trend of first increasing and then decreasing when k increased (Table 7). Figure 5 shows that MSN imputation was more sensitive to the selection of response configurations than RF, which indicates more stability and robustness of RF imputation. With respect to the performance of MSN and RF imputation for diameter distributions, it was found intuitively from Figure 5 that most of RF imputed distributions had higher accuracies than MSN among different response configurations, except that MSN performed slightly better than RF when the response configuration was SET1-2 for all plots, and SET4-5 and SET13-14 for Eucalypt. Similar to Strunk et al. [35] and Shang et al. [59], RF was fairly insensitive to the choice of response configurations and was more robust than MSN. As a robust method against overfitting, the superiority of RF for predicting forest structure attributes had been reported by previous studies [56,57,86,106]. This may be a reason explaining the better performance of RF for diameter distributions with most of the response configurations. Another reason explaining these results may be the fact that RF has an iterative procedure and the average outcome does not describe the worst possible case [57], which may lead to a stable and robust predictive ability of RF for diameter distributions. However, a limitation of RF was that calculating distances is more time-consuming than MSN [35]. As shown in Figure 5, generally, we also found that there was no significant improvement in separate MSN imputations by tree species compared to the imputed diameter distributions for all plots, although there was a significant improvement in RF imputation (Figure 5b). This result was similar to those of Strunk et al. [35] and Räty et al. [37]. This may be explained by the reason that selected neighbors by unstable MSN imputation may include tree species that did not occur in the target plot, due to lower target samples in post-stratification. The sufficient robustness of RF may have only a slightly negative effect on imputed diameter distributions. Our results showed that SET14 (i.e., DBH) was an optimal response configuration for imputing diameter distributions in all plots and Eucalypt plots. Due to meet the needs of predicting diameter distributions in relative complex stands of Chinese fir forests, other responses (i.e., G and H) had to be used in the configurations to increase the capability of searching neighbors for the NN model. Once optimal response configurations and imputation methods were determined for different forest types, the performance of KNN-imputed diameter distributions was evaluated. From Table 8, the diameter distributions of Eucalypt forests derived from MSN imputation performed slightly better than those of Chinese fir forests derived from by RF

imputation with the optimal response configuration. This may be explained by Chinese fir forests having larger tree size inequality compared to Eucalypt forests, although a steadier RF model was chosen by Chinese forests.

4.4. Assessing Diameter Distribution Predictions with UWM, FMM and KNN

In this study, we investigated the capability of UWM, FMM and KNN for predicting tree diameter distributions. Our results revealed that the inclusion of FMM (Mean e_R = 40.85, Mean e_P = 0.20) for all plots (Model 4 in Table 6) produced the highest accuracies of diameter distributions, followed by the KNN (Mean e_R = 52.19, Mean e_P = 0.26) and UWM (Mean e_R = 52.31, Mean e_P = 0.26) methods (Tables 6 and 8). The UWM was expected to have the lowest accuracy, because this model form restricts on a simple unimodal distribution in which UWM has the insufficient capability for fully characterizing multimodal diameter distributions in uneven-sized and multi-species stands. However, this parametric method is still applied in ABA forest inventory, because a relatively few parameters that have to be predicted [58]. Compared to UWM, FMM is more flexible for multimodal empirical distributions. This study discriminated multimodal stands using LiDAR-predicted Lorenz-based indicators. An FMM could be more reliable for capturing a complete representation of the continuous bimodal diameter distributions of differentiated bimodal stands, which has been demonstrated [16,36,47,48]. The FMM for multimodal plots from this study produced higher accuracies (Mean e_R = 39.38, Mean e_P = 0.20) for prediction of diameter distributions than did the results (Mean e_R = 62.21) reported by Mulverhill et al. [48] in boreal natural forests, which may be explained by the effects of higher modality differentiation accuracy (overall accuracy = 77.55%). In non-parametric strategies, KNN has the potential to fit multimodal or irregular distributions rather than unimodal or bimodal distributions. For this reason, KNN could easily produce an overfitting diameter distribution for planted stands with relatively simple structures, such as the imputed distribution in Figure 7e. Although no limit on the number of candidate predictor variables, the KNN model produced unstable distributions. Generally, the KNN model is more sensitive to the quality and amount of field reference data, therefore suggesting that KNN may need a larger number of samples for obtaining more reliable results [18,38]. KNN-imputed diameter distributions in this study obtained relative better performance than did the results (Mean e_R = 80.16) reported by Rana et al. [18], which may be explained by the latter study area was suited in a subtropical natural forest with relative higher tree size heterogeneity.

This study also investigated the effects of stem densities on predicted diameter distributions (Figures 7 and 8). Similar to other studies [39,82], the error index seemly exhibited a slightly decreasing trend with increasing stem density, indicating diameter distributions with high stem density may be more easily be predicted by UWM, FMM and KNN. This may be explained by the fact that the stands with high stem density are often seedlings or young-aged forests and thereby results in relatively low variations.

4.5. Applications and Potential Improvements in Future Work

In this study, we provide a framework for using LiDAR data to discriminate forest structural types and to predict plot-level diameter distributions with unimodal and multimodal models (i.e., UWM, FMM and KNN). This framework provides foresters with a mean to improve differentiation of stand structure types and use differentiated types as prior knowledge to select appropriate models for more accurately estimating diameter distributions. This could contribute to improving the selection of stands for harvest, optimization for stand compositions, silvicultural interventions and enhancing growth and yield projections. However, there is still a lot of potential improvements that need to be done in future work. For instance, plot-level diameter distributions prediction should be used for extrapolation to regional areas. This study only used two error indices to assess goodness of fit. To assess diameter distributions more objectively, more assessment indices, such as the balanced error index (BEI) [59], scaled root mean squared distance (sRMSD) [59], modified error index e_L [107] and bias [45] could be used. In terms of response configuration of MSN and RF imputations, more

structural attributes could be considered as response variables to test, such as basal area weight mean height (H_g) [51], basal area weighted diameter (D_g) [51], Lorey's mean height (H_{lorey}) [38], etc. In addition, more parametric (e.g., mixture Gaussian model, trimodal Weibull models) and non-parametric methods (e.g., Bayesian model and 10-percentiles model) could be attempted in future work. Previous studies [51,108] demonstrated the synergy of LiDAR and other optical data sources (e.g., hyperspectral data [109]) to improve estimates of forest structural attributes and diameter distributions. Thus, future work could consider this synergy for developing links between inherent forest structures and their spectral information to enhance the reliability of predicting diameter distributions. To improve direct interpretability and transferability of LiDAR metrics, multimodal distribution models should be used to characterize more detailed information on vertical forest structure profiles for further extracting more LiDAR metrics.

5. Conclusions

This study compared UWM, bimodal FMM, and multimodal KNN models for predicting tree diameter distributions using airborne LiDAR data over a subtropical planted forest within southern China. To better understand forest structural types and predict diameter distributions using UWM and FMM, prior knowledge of forest structural types is required. For this reason, we evaluated the capability of LiDAR predictions for stratifying forest areas into unimodal and multimodal stands by examining the modality of diameter distributions. When modality differentiation was determined, the UWM and FMM were used to fit diameter distributions in unimodal or multimodal stands and KNN model was employed to predict diameter distributions. In addition, the effective strategies of number neighbors, response configurations and imputation methods in the KNN model with tree species were assessed. The result showed that LiDAR-predicted Lorenz-based indicators, i.e., GC and LA, were the most optimal differentiation index (overall accuracy = 77.55%). Parameters of UWM and FMM were predicted well (R^2 = 0.58–0.95, rRMSE = 5.69%–29.21%) and species-specific models (R^2 = 0.75–0.95, rRMSE = 7.17%–14.94%) had higher accuracies than non-specific models (R^2 = 0.80–0.94, rRMSE = 5.65%–16.39%). The number of neighbors k equal to 5 was been found to be optimal with KNN imputations. Overall, RF imputation generally achieved more stable performance than MSN imputation with respect to changes of responses. The optimal response configurations for all plots, Chinese fir and Eucalypt plots were respectively SET14 (DBH), SET7 (G, H and DBH), and SET14 (DBH). Furthermore, compared to the UWM (Mean e_R = 52.31, Mean e_P = 0.26) method on all plots, FMM and KNN generally produced better performance (Mean e_R = 40.85–52.19, Mean e_P = 0.20–0.26) for predicting diameter distributions. Multimodal KNN model could be expected to provide valuable insights into multimodal diameter distributions in more heterogeneous planted forests with more complex structure. This study demonstrated benefits of multimodal models with LiDAR data to predict diameter distributions and showed that FMM has the great potential to predict diameter distributions and can be a valuable tool for supporting sustainable forest management in subtropical planted forests.

Appendix A

Table A1. Summary of Light Detection and Ranging (LiDAR) metrics.

LiDAR Metrics	Description
Standard metrics	
h_{25}, h_{50}, h_{75} and h_{95} h_{mean}, h_{max}, h_{SQRT}, h_{CURT}, h_{std}, h_{var}, h_{cv}, h_{skew} and h_{kurt}	The percentiles of the canopy height distributions (25th, 50th, 75th and 95th) of first returns. Mean height, max height, quadratic mean height, cubic mean heights, standard deviation, variance, coefficient of variation skewness and kurtosis of the heights above ground of all first returns.
L_{skew}, L_{kurt} and L_{CV}	The ratio of the third (L3) to the second (L2) L-moments; The ratio of the third (L4) to the second (L2) L-moments; The ratio of the second (L2) to the first (L1) L-moments.
d_1, d_3, d_5, d_7 and d_9	The proportion of points above the quantiles (10th, 30th, 50th, 70th and 80th) to total number of points.
CRR	Canopy relief ratio (($mean - min$)/($max - min$)).
Canopy metrics	
Filled, Empty, OG, CG, Eu, Oligo	The voxels filled in point clouds and empty voxels within canopy spaces (Filled, Empty); The empty voxels located above and below the canopy respectively (OG, CG); The voxels located within an uppermost percentile (65%) of filled grid cells of that column, and voxels located below the point in the profile (Eu, Oligo).
Rumple	The ratio of canopy outer surface area to ground surface area.
CvLAD	Variation within the vertical leaf area density profile.
VCI	Vertical complexity index, i.e., Distribution of abundance of returns in specified height bins.

References

1. Food and Agriculture Organization of the United Nations (FAO). *Global Forest Resources Assessment 2015*; FAO: Rome, Italy, 2015.
2. Baral, H.; Guariguata, M.R.; Keenan, R.J. A proposed framework for assessing ecosystem goods and services from planted forests. *Ecosyst. Serv.* **2016**, *22*, 260–268. [CrossRef]
3. Evans, J. *Planted Forests: Uses, Impacts and Sustainability*; Food and Agriculture Organisation of the United Nations (FAO) and CAB International: Rome, Italy, 2009.
4. Bauhus, J.; Schmerbeck, J. Silvicultural options to enhance and use forest plantation biodiversity. In *Ecosystem Goods and Services from Plantation Forests*; Bauhus, J., van der Meer, P.J., Kanninen, M., Eds.; Earthscan: London, UK, 2010; pp. 96–139.
5. Payn, T.; Carnus, J.M.; Freer-Smith, P.; Kimberley, M.; Kollert, W.; Liu, S.; Orazio, C.; Rodriguez, L.; Silva, L.N.; Wingfield, M.J. Changes in planted forests and future global implications. *For. Ecol. Manag.* **2015**, *352*, 57–67. [CrossRef]
6. Messier, C.; Kneeshaw, D. Thinking and acting differently for a sustainable management of the boreal forest. *For. Chron.* **1999**, *75*, 929–938. [CrossRef]
7. Lu, H.; Campbell, E.T.; Campbell, D.E.; Wang, C.; Ren, H. Dynamics of ecosystem services provided by subtropical forests in Southeast China during succession as measured by donor and receiver value. *Ecosyst. Serv.* **2017**, *23*, 248–258. [CrossRef]
8. Cao, L.; Coops, N.C.; Innes, J.; Dai, J.; She, G. Mapping above- and below-ground biomass components in subtropical forests using small-footprint LiDAR. *Forests* **2014**, *5*, 1356–1373. [CrossRef]
9. Valbuena, R.; Vauhkonen, J.; Packalen, P.; Pitkänen, J.; Maltamo, M. Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves. *ISPRS J. Photogramm. Remote Sens.* **2014**, *95*, 23–33. [CrossRef]
10. Kankare, V.; Vauhkonen, J.; Tanhuanpaa, T.; Holopainen, M.; Vastaranta, M.; Joensuu, M.; Krooks, A.; Hyyppa, J.; Hyyppa, H.; Alho, P.; et al. Accuracy in estimation of timber assortments and stem distribution—A comparison of airborne and terrestrial laser scanning techniques. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 89–97. [CrossRef]
11. Silva, C.A.; Klauberg, C.; Hudak, A.T.; Vierling, L.A.; Jaafar, W.S.W.M.; Mohan, M.; Garcia, M.; Ferraz, A.; Cardil, A.; Saatchi, S. Predicting stem total and assortment volumes in an industrial Pinus taeda L. forest plantation using airborne laser scanning data and random forest. *Forests* **2017**, *8*, 254. [CrossRef]

12. Xu, C.; Morgenroth, J. Mapping Net Stocked Plantation Area for Small-Scale Forests in New Zealand Using Integrated RapidEye and LiDAR Sensors. *Forests* **2017**, *8*, 487. [CrossRef]

13. Maltamo, M.; Suvanto, A.; Packalén, P. Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *For. Ecol. Manag.* **2007**, *247*, 26–34. [CrossRef]

14. Bollandsås, O.M.; Næsset, 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]

15. Pukkala, T.; Lähde, E.; Laiho, O. Continuous Cover Forestry in Finland-Recent Research Results. In *Continuous Cover Forestry*; Pukkala, T., von Gadow, K., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2012; p. 4.

16. Jaworski, A.; Podlaski, R. Modelling irregular and multimodal tree diameter distributions by finite mixture models: An approach to stand structure characterisation. *J. Fores* **2012**, *17*, 79–88. [CrossRef]

17. 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]

18. 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]

19. Breidenbach, J.; Gläser, C.; Schmidt, M. Estimation of diameter distributions by means of airborne laser scanner data. *Can. J. For. Res.* **2008**, *38*, 1611–1620. [CrossRef]

20. Kuuluvainen, T.; Tahvonen, O.; Aakala, T. Even-Aged and Uneven-Aged Forest Management in Boreal Fennoscandia: A Review. *Ambio* **2012**, *41*, 720–737. [CrossRef]

21. 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]

22. Næsset, E. Area-Based Inventory in Norway -From Innovation to an Operational Reality. In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 215–240.

23. Burkhart, H.E.; Tomé, M. *Modeling Forest Trees and Stands*; Springer Netherlands: Dordrecht, The Netherlands, 2012.

24. Sironen, S.; Kangas, A.; Maltamo, M.; Kangas, J. Estimating Individual Tree Growth with the k-Nearest Neighbour and k-Most Similar Neighbour Methods Susanna. *Silva Fenn.* **2001**, *35*, 453–467. [CrossRef]

25. Nanos, N.; Montero, G. Spatial prediction of diameter distribution models. *For. Ecol. Manag.* **2002**, *161*, 147–158. [CrossRef]

26. Valbuena, R.; Eerikäinen, K.; Packalen, P.; Maltamo, M. Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. *Ecol. Indic.* **2016**, *60*, 574–585. [CrossRef]

27. Packalén, P.; Maltamo, M. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Can. J. For. Res.* **2008**, *1760*, 1750–1760. [CrossRef]

28. Spriggs, R.A.; Coomes, D.A.; Jones, T.A.; 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]

29. Kotivuori, E.; Maltamo, M.; Korhonen, L.; Packalen, P. Remote Sensing of Environment Calibration of nationwide airborne laser scanning based stem volume models. *Remote Sens. Environ.* **2018**, *210*, 179–192. [CrossRef]

30. 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]

31. White, J.C.; Wulder, M.A; Varhola, A.; Vastaranta, M.; Coops, N.C.; Cook, B.D.; Pitt, D.; Woods, M. *A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach*; Information Report FI-X-10; Canadian Forest Service, Canadian Wood Fibre Centre, Pacific Forestry Centre: Victoria, BC, Canada, 2013; p. 50.

32. Wulder, M.A.; Bater, C.W.; Coops, N.C.; Hilker, T.; White, J.C. The role of LiDAR in sustainable forest management. *For. Chron.* **2008**, *84*, 807–826. [CrossRef]

33. 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]

34. 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.* **2014**, *30*, 186–196. [CrossRef]

35. Strunk, J.L.; Gould, P.J.; Packalen, P.; Poudel, K.P.; Andersen, H.; Temesgen, H. An Examination of Diameter Density Prediction with k-NN and Airborne Lidar. *Forests* **2017**, *8*, 444. [CrossRef]

36. 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]

37. Räty, J.; Packalen, P.; Maltamo, M. Comparing nearest neighbor configurations in the prediction of species-specific diameter distributions. *Ann. For. Sci.* **2018**, *75*, 26. [CrossRef]

38. Maltamo, M.; Næsset, E.; Bollandsås, O.M.; Gobakken, T.; Gobakken, T.; Maltamo, M.; Næsset, E.; Bollandsa, O.L.E.M. Non-parametric prediction of diameter distributions using airborne laser scanner data. *Scand. J. For. Res. ISSN* **2009**, *24*, 541–553. [CrossRef]

39. 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]

40. Bailey, R.L.; Dell, T.R. Quantifying diameter distributions with the Weibull function. *For. Sci.* **1973**, *19*, 97–104.

41. Siipilehto, J.; Mehtätalo, L. Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland. *Silva Fenn.* **2013**, *47*, 1–22. [CrossRef]

42. Siipilehto, J.; Lindeman, H.; Vastaranta, M.; Yu, X.; Uusitalo, J. Reliability of the predicted stand structure for clear-cut stands using optional methods: Airborne laser scanning-based methods, smartphone-based forest inventory application Trestima and pre-harvest measurement tool EMO. *Silva Fenn.* **2016**, *50*, 1–24. [CrossRef]

43. Siipilehto, J. Modelling stand structure in young Scots pine dominated stands. *For. Ecol. Manag.* **2009**, *257*, 223–232. [CrossRef]

44. Mehtätalo, L.; Maltamo, M.; Packalen, P. Recovering plot-specific diameter distribution and height-diameter curve using ALS based stand characteristics. In Proceedings of the ISPRS Working Group, Laser Scanning 2007 and Silvilaser 2007, Espoo, Finland, 12–14 September 2007; Volume XXXVI-3/W52, pp. 288–293.

45. 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]

46. 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]

47. Podlaski, R.; Roesch, F.A. Mathematical Biosciences Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species: A two-component mixture model approach. *Math. Biosci.* **2014**, *249*, 60–74. [CrossRef]

48. 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]

49. Eskelson, B.N.I.; Temesgen, H.; Lemay, V.; Barrett, T.M.; Crookston, N.L.; Hudak, A.T. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. *Scand. J. For. Res.* **2009**, *24*, 235–246. [CrossRef]

50. Chirici, G.; Mura, M.; Mcinerney, D.; Py, N.; Tomppo, E.O.; Waser, L.T.; Travaglini, D.; Mcroberts, R.E. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sens. Environ.* **2016**, *176*, 282–294. [CrossRef]

51. Peuhkurinen, J.; Tokola, T.; Plevak, K.; Sirparanta, S.; Kedrov, A. 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]

52. Maltamo, M.; Malinen, J.; Kangas, A.S.; Härkönen, S.; Pasanen, A.M. A-Pasanen Most similar neighbour-based stand variable estimation for use in inventory by compartments in Finland Most similar neighbour-based stand variable estimation for use in inventory by compartments in. *Forestry* **2003**, *76*, 2003. [CrossRef]

53. Moeur, M.; Stage, A.R. Most similar neighbor: An improved sampling inference procedure for natural resource planning. *For. Sci.* **1995**, *41*, 337–359.

54. Breiman, L. Breiman Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

55. Crookston, N.L.; Finley, A.O. yaImpute: An R Package for k NN Imputation. *J. Stat.* **2008**. [CrossRef]

56. Hudak, A.T.; Crookston, N.L.; Evans, J.S.; Hall, D.E.; Falkowski, M.J. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens. Environ.* **2008**, *112*, 2232–2245. [CrossRef]

57. Vauhkonen, J.; Korpela, I.; Maltamo, M.; Tokola, T. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ.* **2010**, *114*, 1263–1276. [CrossRef]

58. Penner, M.; Woods, M.; Pitt, D.G. A comparison of airborne laser scanning and image point cloud derived tree size class distribution models in Boreal Ontario. *Forests* **2015**, *6*, 4034–4054. [CrossRef]

59. Shang, C.; Treitz, P.; Caspersen, J.; Jones, T.; Shang, C.; Treitz, P.; Caspersen, J.; Estimating, T.J.; Shang, C.; Treitz, P.; et al. Estimating Stem Diameter Distributions in a Management Context for a Tolerant Hardwood Forest Using ALS Height and Intensity Data 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.

60. 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]

61. Maltamo1, M.; Mehtätalo, L.; Valbuena, R.; Vauhkonen, J.; Packalen1, P. Airborne laser scanning for tree diameter distribution modelling: A comparison of different modelling alternatives in a tropical single-species plantation. *Forestry* **2017**. [CrossRef]

62. Song, Y.B.; Ding, Y.P. The development and latest progress of JSCORS. *Bull. Surv. Mapp.* **2009**, *2*, 73–74.

63. Pang, Y.; Ju, H.; Lu, H.; Jia, W.; Si, L.; Guo, Y.; Liu, Q.; Li, S.; Liu, L.; Xie, B.; et al. LiCHy: The CAF's LiDAR, CCD and Hyperspectral Integrated Airborne Observation System. *Remote Sens.* **2016**, *8*, 398. [CrossRef]

64. Zhao, X.; Guo, Q.; Su, Y.; Xue, B. Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 79–91. [CrossRef]

65. White, J.C.; Stepper, C.; Tompalski, P.; Coops, N.C.; Wulder, M.A. Comparing ALS and image-based point cloud metrics and modelled forest inventory attributes in a complex coastal forest environment. *Forests* **2015**, *6*, 3704–3732. [CrossRef]

66. Hilker, T.; Van Leeuwen, M.; Coops, N.C. Comparing canopy metrics derived from terrestrial and airborne laser scanning in a Douglas-fir dominated forest stand. *Trees* **2010**, *24*, 819–832. [CrossRef]

67. Lefsky, M.A.; Cohen, W.B.; Acker, S.A.; Parker, G.G. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-Fir western hemlock forests. *Remote Sens. Environ.* **1999**, *361*, 339–361. [CrossRef]

68. 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. [CrossRef]

69. 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*, 261–270. [CrossRef]

70. 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]

71. McGaughey, R.J. *FUSION/LDV: Software for LiDAR Data Analysis and Visualization*; US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA, 2014.

72. MATLAB. Version 8.4.0.150421 (R2014b). Natick, M.T.M.I., 2013. Available online: https://www.mathworks.com (accessed on 12 December 2018).

73. Pach, M.; Podlaski, R. Tree diameter structural diversity in Central European forests with Abies alba and Fagus sylvatica: Managed versus unmanaged forest stands. *Ecol. Res.* **2015**, *30*, 367–384. [CrossRef]

74. Valbuena, R.; Packalen, P.; Mehtätalo, L.; García-Abril, A.; Maltamo, M. Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. *Can. J. For. Res.* **2013**, *43*, 1063–1074. [CrossRef]

75. Hosenfeld, B.; Bos, E.H.; Wardenaar, K.J.; Conradi, H.J.; Van Der Maas, H.L.J.; Visser, I.; Jonge, P. De Major depressive disorder as a nonlinear dynamic system: Bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry* **2015**, *15*, 222. [CrossRef]

76. Lexerød, N.L.; Eid, T. An evaluation of different diameter diversity indices based on criteria related to forest management planning. *For. Ecol. Manag.* **2006**, *222*, 17–28. [CrossRef]

77. Valbuena, R.; Maltamo, M.; Martı, S.; Packalen, P.; Pascual, C. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. *Can. J. Remote Sens.* **2013**, *39*, S18–S31. [CrossRef]

78. Valbuena, R.; Packalén, P.; Martı´n-Fernández, S.; Maltamo, M. Diversity and equitability ordering profiles applied to study forest structure. *For. Ecol. Manag. Ecol. Manag.* **2012**, *276*, 185–195. [CrossRef]

79. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.

80. Cowell, F. Measurement of Inequality. In *Handbook of Income Distribution*; Atkinson, A.B., Bourguignon, F., Eds.; STICERD: London, UK, 2000; pp. 87–166.

81. Liu, C.; Beaulieu, J.; Prégent, 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]

82. Gorgoso, J.J.; Álvarez González, J.G.; Rojo, A.; Grandas Arias, J.A. Modelling diameter distributions of Betula alba L. stands in northwest Spain with the two-paramenter Weibull function. *Investig. Agrar. Sist. Recur. For.* **2007**, *16*, 113–123. [CrossRef]

83. Liu, C.; Zhang, L.; Davis, C.J.; Solomon, D.S.; Gove, J.H. A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *For. Sci.* **2002**, *48*, 653–661.

84. Zasada, M.; Cieszewski, C.J. A finite mixture distribution approach for characterizing tree diameter distributions by natural social class in pure even-aged Scots pine stands in Poland. *For. Ecol. Manag.* **2005**, *204*, 145–158. [CrossRef]

85. Peter Macdonald, with Contributions from Juan Du. Mixdist: Mixture Distribution Models. R Package Version 0.5-5. 2012. Available online: https://cran.r-project.org/web/packages/mixdist/index.html (accessed on 12 December 2018).

86. Latifi, H.; Nothdurft, A.; Koch, B. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry* **2010**, *83*, 395–407. [CrossRef]

87. Dash, J.P.; Marshall, H.M.; Rawley, B. Methods for estimating multivariate stand yields and errors using k -NN and aerial laser scanning. *Forestry* **2015**, *88*, 237–247. [CrossRef]

88. Belgiu, M.; Drǎguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

89. Zhang, Z.; Cao, L.; She, G. Estimating forest structural parameters using canopy metrics derived from airborne LiDAR data in subtropical forests. *Remote Sens.* **2017**, *9*, 940. [CrossRef]

90. Sprugel, D. Correcting for Bias in Log-Transformed Allometric Equations. *Ecology* **1983**, *64*, 209–210. [CrossRef]

91. Næsset, E.; Økland, T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115. [CrossRef]

92. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Control* **1974**, *19*, 716–723. [CrossRef]

93. Reynolds, M.; Burk, T.; Huang, W. Goodness-of-fit tests and model selection procedures for diameter distribution models. *For. Sci.* **1988**, *34*, 373–399.

94. Kato, J.; Hayashi, I. Quantitative analysis of a stand of Pinus densiflora undergoing succession to Quercus mongolica ssp. crispula: II. Growth and population dynamics of Q. mongolica ssp. crispula under the *P. densiflora* canopy. *Ecol. Res.* **2007**, *22*, 527–533. [CrossRef]

95. Hui, G.; Pommerening, A. Analysing tree species and size diversity patterns in multi-species uneven-aged forests of Northern China. *For. Ecol. Manag.* **2014**, *316*, 125–138. [CrossRef]

96. 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]

97. Valbuena, R. *Forest Structure Indicators Based on Tree Size Inequality and Their Relationships to Airborne Laser Scanning*; University of Eastern Finland: Joensuu, Finland, 2015.

98. Damgaard, C.; Weiner, J. Describing inequality in plant size or fecundity. *Ecology* **2000**, *81*, 1139–1142. [CrossRef]

99. Knox, R.; Peet, R.; Christensen, N. Population dynamics in Loblolly Pine stands: Changes in skewness and size inequality. *Ecology* **1989**, *70*, 1153–1166. [CrossRef]

100. Turnbull, J.W. Development of sustainable forestry plantations in China: A review. *Int. For. Rev.* **2008**, *1*, 95–96.

101. Næsset, E.; Gobakken, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090. [CrossRef]

102. 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]

103. Cao, L.; Coops, N.C.; Hermosilla, T.; Innes, J.; Dai, J. Using Small-Footprint Discrete and Full-Waveform Airborne LiDAR Metrics to Estimate Total Biomass and Biomass Components in Subtropical Forests. *Remote Sens.* **2014**, *6*, 7110–7135. [CrossRef]

104. Tonolli, S.; Dalponte, M.; Neteler, M.; Rodeghiero, M.; Vescovo, L.; Gianelle, D. Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps. *Remote Sens. Environ.* **2011**, *115*, 2486–2498. [CrossRef]

105. Kane, V.R.; Bakker, J.D.; Lutz, J.A.; Franklin, J.F.; Kane, V.R.; Mcgaughey, R.J.; Bakker, J.D.; Gersonde, R.F.; Lutz, J.A.; Franklin, J.F. Comparisons between field- and LiDAR-based measures of stand structural complexity. *Can. J. For. Res.* **2010**, *40*, 761–773. [CrossRef]

106. Gagliasso, D.; Hummel, S.; Temesgen, H. A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area. *Open J. For.* **2014**, *4*, 42–48. [CrossRef]

107. Lindberg, E.; Holmgren, J.; Olofsson, K.; Wallerman, J.; Olsson, H. Estimation of tree lists from airborne laser scanning by combining single- tree and area-based methods. *Int. J. Remote Sens.* **2010**, *31*, 1175–1192. [CrossRef]

108. Zald, H.S.J.; Wulder, M.A.; White, J.C.; Hilker, T.; Hermosilla, T.; Hobart, G.W.; Coops, N.C. Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sens. Environ.* **2016**, *176*, 188–201. [CrossRef]

109. Shen, X.; Cao, L.; Chen, D.; Sun, Y.; Wang, G.; Ruan, H. Prediction of Forest Structural Parameters Using Airborne Full-Waveform LiDAR and Hyperspectral Data in Subtropical Forests. *Remote Sens.* **2018**, *10*, 1729. [CrossRef]