LiDAR and Weibull modeling of diameter and basal area

by V. Thomas[1], R.D. Oliver[2], K. Lim[3] and M. Woods[4]

ABSTRACT

This study investigates the ability to predict forest diameter distributions from light detection and ranging (LiDAR) data using Weibull modelling for forest stands in central Ontario. Results suggest that the unimodal 2-parameter Weibull model is a promising technique for the prediction of diameter class distributions, with strong relationships evident for several subgroups (at 95% confidence, r^2_{adj}=0.83, 0.78, 0.88, 0.80, 0.83, and 0.65, with validation RMSE of 4.09 m^2/ha, 0.61 stems/ha, 6.05, 0.64, 4.73, and 0.09 for basal area, stem density, and the Weibull a and b parameters for basal area and stem density, respectively). The unimodal models were found to be least effective for the irregularly shaped diameter distributions, particularly for low-density coniferous plots that have undergone shelterwood treatment. A significant improvement in results for these irregular plots was found with a finite mixture modelling approach, suggesting that finite mixture models may extend our ability to predict diameter distributions over large portions of the landscape.

Key words: LiDAR, Weibull, finite mixture modeling, diameter class distributions, multiple linear regression

RÉSUMÉ

Cette étude porte sur la capacité de prédiction de la distribution des diamètres d'arbres à partir de données de détection de la lumière et de calcul de la distance (LiDAR) utilisées dans un modèle Weibull pour des peuplements forestiers du centre de l'Ontario. Les résultats laissent entendre que le modèle Weibull unimodal à 2 paramètres constitue une technique prometteuse de la prédiction de la distribution des classes de diamètre comportant de fortes corrélations évidentes pour certains sous-groupes (à un niveau de confiance 95%, r^2_{adj}=0.83, 0.78, 0.88, 0.80, 0.83 et 0.65, une variance de 4,09 m^2/ha, 0,61 tiges/ha, 6,05, 0,64, 4,73 et de 0,09 respectivement pour la surface terrière, le nombre de tige par hectare et pour les paramètres a et b de Weibull portant sur la surface terrière et le nombre de tige). Les modèles unimodaux se sont avérés être moins efficaces dans le cas des distributions de forme irrégulière des diamètres, particulièrement dans le cas de parcelles de faible densité de résineux ayant subi des traitements de coupe progressive. Une amélioration significative des résultats de ces parcelles irrégulières est apparue à la suite d'une approche de modélisation pour peuplement mélangé spécifique, laissant entendre que les modèles pour peuplement mélangé spécifique pourrait accroître notre capacité de prédire la distribution des diamètres pour de grandes portions de l'écosystème.

Mots clés : LiDAR, Weibull, modélisation de peuplement mélangé spécifique, distributions des classes de diamètre, régression linéaire multiple

V. Thomas

R.D. Oliver

K. Lim

M. Woods

[1]Department of Forestry, Virginia Tech, Blacksburg, VA, USA 24061. Corresponding author. E-mail: thomasv@vt.edu
[2]Department of Geography, Queen's University, Kingston, Ontario K7L 1J9. E-mail: 8rdo@qlink.queensu.ca
[3]Lim Geomatics Inc., 264 Catamount Court, Ottawa, Ontario K2M 0A9. E-mail: kevin@limgeomatics.com
[4]Ontario Ministry of Natural Resources, 3301 Trout Lake Road, North Bay, Ontario P1A 4L7. E-mail: murray.woods@Ontario.ca

Introduction

Weibull modeling has been shown to be an effective technique to quantify the diameter frequency distributions of various types of forest stands (Bailey and Dell 1973, Maltamo *et al.* 1995, Nanang 1998, Zhang and Liu 2006). Recently, attempts have been made to relate the 2-parameter Weibull density and cumulative distribution functions of forest diameter and basal area distributions to airborne discrete light detection and ranging (LiDAR) metrics (Gobakken and Næsset 2004, Maltamo *et al.* 2007). This would enable the characterization of diameter and basal area distributions using remotely sensed data.

Thus far, the attempt to relate Weibull models of forest stands to LiDAR metrics has been limited to predominantly coniferous stands stratified by age and site quality (Gobakken and Næsset 2004). This stratification of forest types normally results in regular and uniform diameter distributions, which can be well modeled by the 2-parameter Weibull function (Zhang and Liu 2006). Alternatively, uneven-aged stands, mixedwood stands, or old-growth stands are often characterized by multimodal or irregular diameter distributions, for which the 2-parameter Weibull distribution may be an over simplification (Zhang *et al.* 2001, Liu *et al.* 2002, Zasada and Cieszewski 2005, Zhang and Liu 2006) and age may be an inappropriate stratifier (Peng 2000). For these stands, a number of alternate distributions have been proposed, the most flexible being finite mixture models (Zhang and Liu 2006). A finite mixture model has an overall distribution which consists of 2 or more smaller distributions, combined in an additive fashion. For the bimodal case, the 2-parameter finite mixture model is described by 6 variables (4 are the original parameters from each of the 2 unimodal Weibull models that make up the overall mixture model), with 2 parameters describing the proportion of the overall distribution comprised by each mode. Finite mixture modeling of forest diameter distributions is a relatively new approach and, to the authors' knowledge, there has been no attempt to predict finite mixture Weibull models of irregular diameter class distributions from LiDAR data.

The objective of this study is to develop predictive models of diameter and basal area distributions for forest stands in Central Ontario, Canada. To achieve this objective, airborne discrete LiDAR metrics are regressed against Weibull parameters fitted to field-measured diameter and basal area distributions. Stands are stratified according to common structural types, based on their average stem density and canopy top height. Weibull parameters are predicted from LiDAR metrics using multiple linear regression analysis and are tested for robustness using cross validation statistics. For the structural group characterized by irregular multimodal diameter distributions, finite mixture Weibull models are also predicted from LiDAR to evaluate the utility of this approach for Ontario forests.

Methods

The study sites consist of 115 large-tree plots located within a 100 km × 130 km region of the Great Lakes – St. Lawrence forest ecosystem of Ontario (Fig. 1). The plots cover a wide range of forest types and species, including natural hardwood stands (sugar maple [*Acer saccharum* Marsh.], red oak [*Quercus rubra* L.], white birch [*Betula papyrifera* Marsh.], yellow birch [*Betula alleghaniensis* Britt.], beech [*Fagus grandifolia*

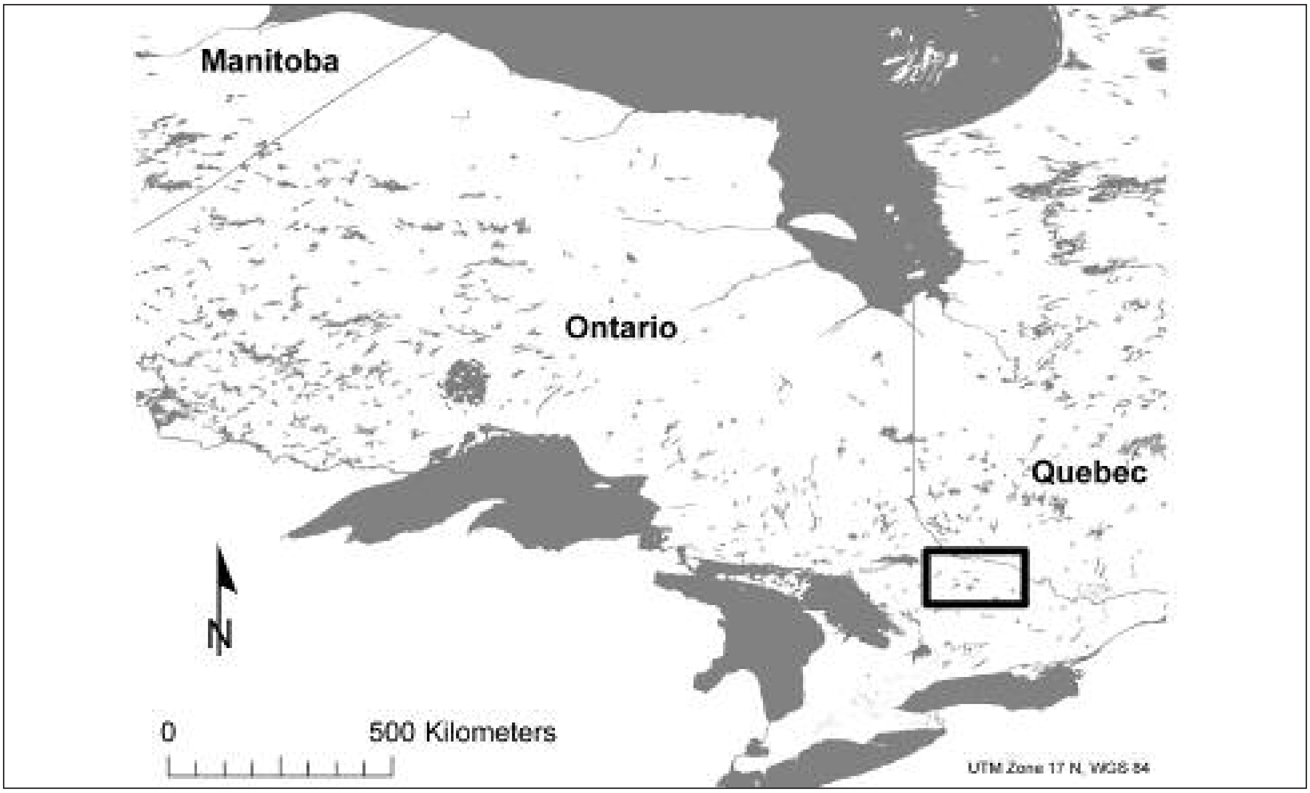

Fig. 1. Bounding area of study sites within the Great Lakes – St. Lawrence Forest Region of Ontario.

Ehrh.], and trembling aspen [*Populus tremuloides* Michx.]), natural conifer stands (northern white cedar [*Thuja occidentalis* (L.)], white spruce [*Picea glauca* (Moench) Voss], black spruce [*Picea mariana* (Mill.) BSP), red pine [*Pinus resinosa* Ait.], white pine [*Pinus strobes* L.], jack pine [*Pinus banksiana* Lamb.], and balsam fir [*Abies balsamea* (L.)]), conifer plantations, and natural mixedwoods.

Large-tree plot sizes range from 400 m^2 to 2500 m^2, with the exception of one hemlock plot at 100 m^2. At all plots, all live trees with a diameter at breast height (DBH) greater than 9 cm were tallied for average height, top height, basal area, quadratic mean DBH (DBHq), species, and stem density. Stem and basal area distributions were generated with 2-cm and 5-cm diameter classes for each plot. Hardwood plots were also examined at the much coarser operational scale, which consists of 4 classes (poles, small, medium, and large trees).

Given the wide range of canopy structures present within the entire study area, hardwood plots were treated separately for the development of predictive models. Coniferous and mixedwood plots (C&MW) were stratified according to their ratio of stem density by canopy top height (hereafter, this ratio is referred to as D/H). This was based upon previous work by numerous authors that reported strong relationships between LiDAR returns and functions of canopy height and density (e.g., Magnussen and Boudewyn 1998, Lim and Treitz 2004, Thomas *et al.* 2006b). Four C&MW groups were made, according to D/H ≤ 15, 15 < D/H ≤ 25, 25 < D/H ≤ 50, D/H > 50. These groupings corresponded to distinct structure differences resulting in unique diameter distributions for each group. To consider the possibility that the standard 0.04-ha plots may be too small to adequately characterize the species in this study and to account for the structural differences, small hardwood plots (~0.04 ha) and large plots (>0.1ha) were considered separately. Further stratification according to the D/H ratio was unnecessary for hardwoods, because most of the plots fell within a relatively narrow range of this metric (11 ≤ D/H ≤ 33). The range of plot averages of various field measurements according to groupings is shown in Table 1.

Weibull modeling of stem and basal area distributions

The simplest Weibull model approach is the 2-parameter model, which is characterized by a curve that can range from an inverse-J shape that asymptotically approaches the x and y axis to a unimodal curve that is either symmetrical or skewed. This flexibility in shape enables the model to characterize many types of forest stands, resulting in its wide use for diameter distribution modelling (Bailey and Dell 1973, Maltamo *et al.* 1995, Nanang 1998, Zhang and Liu 2006). In this study, diameter distributions were modeled for each plot by estimating the "a" and "b" parameters of the 2-parameter Weibull probability density function (pdf), shown as follows:

$$[1] \quad f(x \mid a,b) = b a^{-b} x^{b-1} e^{-\left(\frac{x}{a}\right)^{b}} \qquad 0 \le x < \infty, a > 0, b > 0$$

where a is the scale parameter and b is the shape parameter.

The scale and shape parameters are determined by fitting the Weibull probability density function to the field measures of diameter and basal area distributions in 2-cm classes. The "a" and "b" parameters are used as the dependent variables within the multiple regression analysis.

In some circumstances, a unimodal or inverse-J shape may not adequately describe the shape of a diameter distribution. This is particularly true in cases where the diameter distribution has significant gaps, which may occur in very low density plots containing a few large old trees and a number of smaller trees. A unimodal curve cannot account for a gap in the mid-sized trees, and will greatly overestimate the true distribution. Weibull finite mixture models may provide some benefit when attempting to reproduce highly irregular diameter class distributions, such as those found within very low density plots, or those that have undergone some type of silviculture harvest. Finite mixture models have 2 or more modes and have been used to model diameter class distributions of mixed species or uneven-aged forest stands (Liu *et al.* 2002, Zhang and Liu 2006).

A finite mixture model is essentially an overall distribution made up of 2 or more smaller distributions. For this analysis, a 2-mode 2-parameter Weibull finite mixture model was evaluated. The pdf is a linear combination of the simple 2-parameter Weibull model described above, and was described by Zhang and Liu (2006) as:

$$[2] \quad f(x) = \sum_{i=1}^{k} \rho_i f_i(x) = \rho_1 f_1(x) + \ldots + \rho_k f_k(x)$$

For the bimodal case, the finite mixture model is described by 6 parameters (a_1, b_1, a_2, b_2, ρ_1, ρ_2) where the ρ_1 and ρ_2

Table 1. Range of mensuration values for the study plots according to D/H grouping

Group	N	DBHq (cm)	Basal area (m^2/ha)	Density (#/ha)	Average height (m)	Top height (m)
All C&MW: 1 < D/H ≤ 146	68	11–59	5–70	20–2700	9–32	10–38
C&MW: D/H > 50	19	11–24	12–67	775–2700	9–26	10–27
C&MW: 25 < D/H ≤ 50	18	16–33	7–70	320–1225	10–28	12–29
C&MW: 15 < D/H ≤ 25	13	22–34	17–60	340–676	14–23	16–27
C&MW: D/H ≤ 15	18	22–59	5–48	20–468	19–32	22–38
All hardwoods	47	14–40	6–47	75–1475	12–26	16–28
Hardwoods, large plots >1000m^2	19	17–33	9–35	289–630	12–23	17–27
Hardwoods, small plots ~400m^2	28	14–40	6–47	75–1475	15–26	16–28

Note: C&MW refers to coniferous and mixedwood plots. D/H refers to the density divided by canopy top height metric used to group the C&MW plots. DBHq is the quadratic average diameter at breast height.

parameters describe the proportion of the overall distribution comprised by each mode. Based on the a and b parameters of the Weibull model, the mean and standard deviation can also be derived for each mode using general statistical relationships characteristic of Weibull distributions. For example, when the scale parameter is equal to 1, the mean of the Weibull distribution can be determined as:

[3] $$\mu_w = \Gamma(\frac{b+1}{b})$$ (NIST 2006)

where Γ is the gamma function, defined as:

[4] $$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$ (NIST 2006)

The standard deviation of the Weibull function can similarly be calculated as:

[5] $$\sigma_w = \sqrt{\Gamma\frac{(b+2)}{b} - \left(\Gamma(\frac{b+1}{b})^2\right)}$$ (NIST 2006)

For irregularly shaped plots, a multi-modal finite mixture model may be required to adequately describe the diameter distribution. However, each mode added to a 2-parameter finite mixture model requires 3 additional parameters. To predict the entire model from LiDAR would then require the accurate prediction of each parameter in a finite mixture model. As the number of models increases, the experimentwise error rate (EER, or the probability that one of the variables has a Type I error) will also increase (Lane 2007), suggesting that limiting the models to 2 or 3 modes (i.e., 6 or 9 MLR equations) is highly desirable if prediction from LiDAR is the goal. This is of particular concern for the scale parameter, which specifies the location of peak of the individual modes.

In some cases, it may be more appropriate to view the diameter distribution at a coarser scale, which would have the effect of smoothing the histogram. For all stands, models were compared for 2-cm and 5-cm diameter classes. Hardwoods were also examined at the operational scale, which represents a significant simplification of the diameter distribution. Operational classes are defined as poles (diameter class [DC] 10–24 cm), small (DC 26–36 cm), medium (DC 38–48 cm), and large (DC 50+ cm).

LiDAR analysis

Several authors have demonstrated the relationship between LiDAR percentiles and other statistics and field-measured forest structural characteristics (Lim and Treitz 2004; Næsset 2004; Thomas *et al.* 2006a, b). For this analysis, a number of LiDAR metrics were generated to compare with the "a" and "b" parameters of the 2 types of Weibull models. These include the mean, median, standard deviation, and coefficient of variation of the first returns, every 10th percentile of the LiDAR first returns, and ten canopy density metrics (based on Gobakken and Næsset 2004, where the range of heights greater than 2 m is divided into 10 equal fractions and for each fraction the proportion of 1st returns in the fraction relative to the total number 1st returns is calculated).

Multiple regression analysis

Multiple linear regression models were generated using the best subsets approach (Hudak *et al.* 2006). This approach has an advantage over forward and backward stepwise regression techniques in that it compares all models created from all possible combinations of predictor variables, based on user-specified criteria. To determine the optimal model after the best subsets analysis, 5 main criteria were considered. First, variance inflation factors (VIF) were used to eliminate models with multiple collinearity. A low VIF for all predictor variables was desirable and in all cases the maximum VIF was less than 10 (Hyde *et al.* 2007). Second, all predictor variables should be statistically significant (preferably at 95% confidence or better). Third, simple models were preferred over more complex models, with no more than 6 predictor variables in any model. Fourth, the entire model must be statistically significant (preferably at 95% confidence or better). Finally, the amount of variance explained by the model (r^2 and r^2_{adj}) should be high. Of all the models generated by the best subsets procedure, the one that best fit the above 5 criteria was chosen as the optimal model.

To help ensure the assumptions of multiple regression analysis were met, all distributions were tested for normality using the Shapiro Wilks test. The Weibull "b" parameters (BAb and Nb) and average stem density were found to be positively skewed and were normalized with a ln(x) transformation (e.g., Gobakken and Næsset 2004). Pairwise correlation analysis was used prior to the multiple regression analysis to identify and eliminate predictor variables that were highly correlated.

Multiple linear regression (MLR) models had the following form:

[6] $$\hat{y} = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where $x_1 - x_n$ are the significant independent variables (i.e., the relevant LiDAR metrics). Multiple regression models were generated to relate LiDAR metrics to the dependent variables listed in Table 2.

Given the relatively low number of plots within each structural class, cross validation was performed as a way of assessing the error of the MLR predictions (e.g., Gobakken and Næsset 2004). For each structural group, one of the plots was removed from the dataset and the MLR models were fitted to the remaining plot data. This procedure was repeated for each of the plots within the group to generate an adjusted Root Mean Square Error for the cross validation ($RMSE_{xval}$). This could be compared to the overall model RMSE to determine the stability of the model and the influence of individual plots.

Table 2. Dependent variables used in multiple regression models relating LiDAR metrics to Weibull parameters

Variable	Definition
Na	Weibull "a" parameter for diameter distribution
Nb	Weibull "b" parameter for diameter distribution
BAa	Weibull "a" parameter for basal area distribution
BAb	Weibull "a" parameter for basal area distribution
BA	Average basal area (m^2/ha)
N	Average stem density (#/ha)

Note that where necessary, the output parameters were then back-transformed to their original units when reporting RMSE. This was accomplished by multiplying by a correction factor (CF) for each model based on the relationship described in Sprugel (1983):

[7] $$CF = \exp(SEE^2 / 2)$$

where SEE is the standard error of the estimate for the regression model.

Table 3. Definition of LiDAR metrics selected by the best subsets regression (for all models)

LiDAR metric in predictive model	Definition of LiDAR Metric
mean	Mean of first returns
sd	Standard deviation of first returns
cv	Coefficient of variation of first returns
p10	Mean of the 10^{th} percentile of first returns
p20	Mean of the 20^{th} percentile of first returns
p90	Mean of the 90^{th} percentile of first returns
d1 d2 d3 d4 d5 d6 d7 d8 d9 d10	Canopy density metrics, where the range of heights greater than 2 m is divided into 10 equal fractions and for each fraction the proportion of 1^{st} returns in the fraction relative to the total number 1^{st} returns is calculated (Gobakken and Naesset 2004). The value of D1 represents the lowest height fraction and D10 represents the highest height fraction.

Results

The LiDAR metrics selected as predictive variables in all of the best subsets regression analysis are defined in Table 3.

Results for the coniferous and mixedwood groups are shown in Tables 4 and 5 and Fig. 2. The first C&MW group (D/H > 50) consists of conifer plantation plots that are densely spaced with homogeneous species (black spruce, red spruce, white pine, jack pine, and red pine). These characteristics create narrow unimodal diameter distributions, with most trees falling into diameter classes less than 30 cm. The unimodal shape is well characterized by the 2-parameter Weibull model (Fig. 2A, B). MLR analysis demonstrates strong predictive capabilities for the Weibull "a" parameters using LiDAR metrics, with relatively low RMSE values (Table 4). The Weibull "b" parameter (which indicates the shape of the distribution) is less well predicted than Weibull "a" for the first C&MW group (Table 4). This suggests that the largest DBH class (i.e., the peak of the diameter distribution curve) is more easily predicted from LiDAR than the very narrow shape of the distribution. Average stem density and basal area are also well predicted from LiDAR metrics for this group (r^2_{adj} = 0.79, 0.76 for average basal area and stem density, respectively, Table 5). By calibrating with the LiDAR-predicted average basal area for any plot, the Weibull model predicted by LiDAR can be used to recreate a diameter distribution per hectare for 2-cm or 5-cm diameter classes (Fig. 2A, B).

Table 4. Density and basal area distribution MLR models and statistics derived from LiDAR data for the coniferous and mixedwood groups

Dependent variable	Linear predictive model	r^2_{adj}	F	p	RMSE	RMSE cross validation
C&MW 1 Na	-25.3 + 47.9cv + 1.03p90 – 128d1 + 48.5d2 + 66.7d5 + 60.4d8	0.88	23.9	0.00	1.25	2.02
C&MW 1 Nb	3.62 + 0.126sd - 1.62cv - 0.0521p10 + 4.21d3 - 6.32d4 - 3.96d5 - 4.20d9	0.50	3.7	0.02	1.18	1.39
C&MW 1 BAa	27.1 + 51.3cv+1.12p90-142d1+47.2d2+74.5d5+61.4d8	0.86	20.5	0.00	1.4	2.30
C&MW 1 BAb	3.93 + 0.134sd - 1.39cv - 0.0519p10 - 5.48d4 - 4.23d5 - 5.39d9	0.53	4.52	0.01	1.17	1.31
C&MW 2 Na	50.5 - 23.9cv - 91.4d4 - 31.4d7 - 43.4d9	0.82	21.1	0.00	1.90	2.60
C&MW 2 Nb	3.20 - 0.0901mean + 0.132p10 - 13.5d1 - 5.98d5 - 2.65d8	0.80	15.8	0.00	1.23	1.39
C&MW 2 BAa	-17.8 + 297d1 + 221d2 - 92.4d4 + 116d5 + 53.4d6 + 103d8	0.83	15.6	0.00	2.65	4.80
C&MW 2 BAb	2.05 - 0.0830mean + 1.08cv + 0.146p10 - 9.38d1 - 4.87d5	0.83	15.2	0.00	1.18	1.28
C&MW 3 Na	41.3 - 4.07sd - 1.37p10 + 155d8 - 120d1[b]	0.83	15.6	0.00	3.74	5.80
C&MW 3 Nb	2.30 - 0.251sd - 0.0685p10 + 4.91d1 + 3.57d8	0.88	26.2	0.00	1.14	1.22
C&MW 3 BAa	39.7 - 1.52mean - 213d1 + 157d8 + 63.0d10	0.79	12.1	0.00	4.28	6.79
C&MW 3 BAb	2.47 - 0.199sd - 0.0602p10 + 3.39d8[b]	0.47	4.6	0.03	1.35	1.57
C&MW 4 Na[a]	-0.1 + 2.87p10 – 1.93p20 + 2.50p90 - 56.9d8 - 99.1d10	0.80	15.0	0.00	3.97	6.74
C&MW 4 Nb	1.94 - 1.09cv + 4.65d1 - 5.39d4 + 6.13d6 – 0.0359p20[b]	0.46	3.9	0.03	1.40	0.63
C&MW 4 BAa[a]	6.7 + 1.80p90 + 65.2d3 + 83.9d4 - 135d10	0.69	10.4	0.00	5.32	7.00
C&MW 4 BAb[c]	0.035 - 0.0294p20 + 6.17d3 + 4.92d1 + 7.11d6 + 2.68d8 + 3.15d9	0.40	2.9	0.06	1.03	0.62

Note: variables are defined in Tables 2 and 3.
[a]Intercept not significant
[b]Predictor significant at 90% confidence
[c]Model significant at 90% confidence

Table 5. Plot average basal area and stem density MLR models and statistics derived from LiDAR data for the coniferous and mixedwood groups

Dependent variable	Linear predictive model	r^2_{adj}	F	p	RMSE	RMSE cross validation
C&MW 1 BA	16.2 + 32.1cv + 2.45p10 - 150d3 - 79.7d10	0.79	18.34	0.00	5.19 m^2/ha	7.37 m^2/ha
C&MW 1 N	6.18 + 0.160p10 - 0.158p90 + 13.1d1 + 4.13d4 + 4.20d7 + 5.74d9	0.76	11.0	0.00	1.18 #/ha	1.35 #/ha
C&MW 2 BA	131 - 95.3cv - 272d4 - 150d7 - 133d9	0.87	32.3	0.00	4.92 m^2/ha	7.40 m^2/ha
C&MW 2 N	5.19 + 0.103mean - 0.0606p10 - 5.91d2 + 3.29d4 + 2.00d6 + 2.99d10	0.79	12.4	0.00	1.12 #/ha	1.18 #/ha
C&MW 3 BA[a]	-3.47 + 4.34mean - 162d8 - 96.4d10[b]	0.77	14.5	0.00	6.27 m^2/ha	8.35 m^2/ha
C&MW 3 N	5.55 + 0.209mean - 17.0d8	0.73	17.5	0.00	1.85 #/ha	2.61 #/ha
C&MW 4 BA	22.4 + 2.15mean - 1.53p20 - 71.8d1 + 59.0d2 - 125d3 - 70.1d10	0.83	15.2	0.00	2.94 m^2/ha	4.51 m^2/ha
C&MW 4 N	6.00 + 0.543sd + 0.156p20 - 0.293p90 + 2.20d2 + 4.39d8 + 0.970cv[b]	0.74	9.0	0.00	1.36 #/ha	1.75 #/ha

Note: variables are defined in Tables 2 and 3.
[a]Intercept not significant
[b]Predictor significant at 90% confidence

The second C&MW group, 25 < D/H ≤ 50, consists mostly of mixedwoods and natural conifers, along with 2 jack pine plantation plots. Although there is more variability in the shape of the diameter distributions for these plots, particularly at the 2-cm class scale, the histograms can generally be described as unimodal, but covering a wider range of diameter classes as compared to the previous group. The jack pine plantation plots that are included in this group have a histogram peak at 30 cm to 35cm, with virtually no overlap in the distributions between these 2 plots and the plantation plots in the previous group. MLR analysis shows strong Weibull predictive capabilities for this group for both Weibull "a" and "b" parameters (Table 4). Average basal area and stem density are also well predicted (r^2_{adj} = 0.87 and 0.79, respectively, Table 5).

The third C&MW group, 15 < D/H ≤ 25, is mixture of natural conifers, mixedwoods, and a few plantation plots. The diameter distributions within this group are more variable than the previous groups but can be generally described as unimodal (Fig. 2E, F). MLR statistics are also strong for this group, although the cross-validation statistics show the models to be less robust (Table 4). This is also related to the small number of plots for the group (n = 13).

The final C&MW group (D/H ≤ 15) is comprised mainly of natural pine stands, some of which had been thinned under a full and half-crown spacing shelterwood treatment. These plots are characterized by fewer but larger trees. The resulting stem and basal area distributions are left skewed with gaps (Fig. 2). Of all the groups, this histogram shape is most poorly represented by the 2-parameter Weibull model, which has a unimodal shape. As a result, MLR results show the weakest predictive abilities between the Weibull parameters and LiDAR metrics for this group (r^2_{adj} = 0.80, 0.46, 0.69, and 0.40 for Na, Nb, BAa, and BAb respectively. The RMSE is also notably higher for this group, including for cross validation, suggesting the largest error (Table 4).

Overall, the r^2_{adj} values shown in Table 4 illustrate the ability of LiDAR to predict Weibull parameters that best characterize the diameter distribution. However, this does not necessarily mean that a Weibull model is an accurate representation of the diameter distribution, regardless of whether the Weibull parameters are well predicted by LiDAR. In fact, results demonstrate that even with strong relationships between LiDAR and Weibull parameters, the method is not suitable for cases where the diameter distribution has significant gaps. This tends to occur in low stem density plots, which tend to have gaps in the mid-sized trees, creating a multi-modal diameter distribution.

Hardwood results are compared for small-area versus large-area plots in Fig. 3. The 2 plots highlighted in Fig. 3 show the typical differences between the small- versus large-area plots. Fig. 3A to 3C show the 2-cm, 5-cm, and operational classes for a small-area hardwood plot. Gaps in the histogram are evident in the 2-cm and 5-cm classes compared to the relatively full distribution shown for the large-area hardwood plot (3D to 3F). Best subsets regression models are summarized in Table 6. Regression analysis demonstrated that for the small plots, the 5 previously mentioned model selection criteria could not be adequately satisfied and, thus, optimal models were only selected for the large plots. For the large plots, regression results are similar to the C&MW groups (Table 6). At the operational scale, the gaps noticeable in the small plots are not apparent and the distributions appear similar for both the large and small plots (Fig. 3C and 3F).

Mixture model regression statistics for the final C&MW, which had the most irregularly shaped diameter distributions, are shown in Table 7. Results suggest a significant improvement in the ability of LiDAR to predict the MLR parameters for these irregular plots. One of the useful aspects of the multi-modal Weibull mixture models is the amount of information readily available from the model parameters. For example, the ρ parameter can be thought of as the proportion of the stem distribution for a particular mode, while the μ_w, σ_w parameters (derived from the "a" and "b" parameters as described above) represent the mode's mean and standard deviation, respectively.

One plot with an irregularly shaped (bi-modal) histogram was selected to demonstrate the utility of the mixture model approach, as shown in Fig. 4. For this plot, the following parameters were derived from the mixture model analysis:

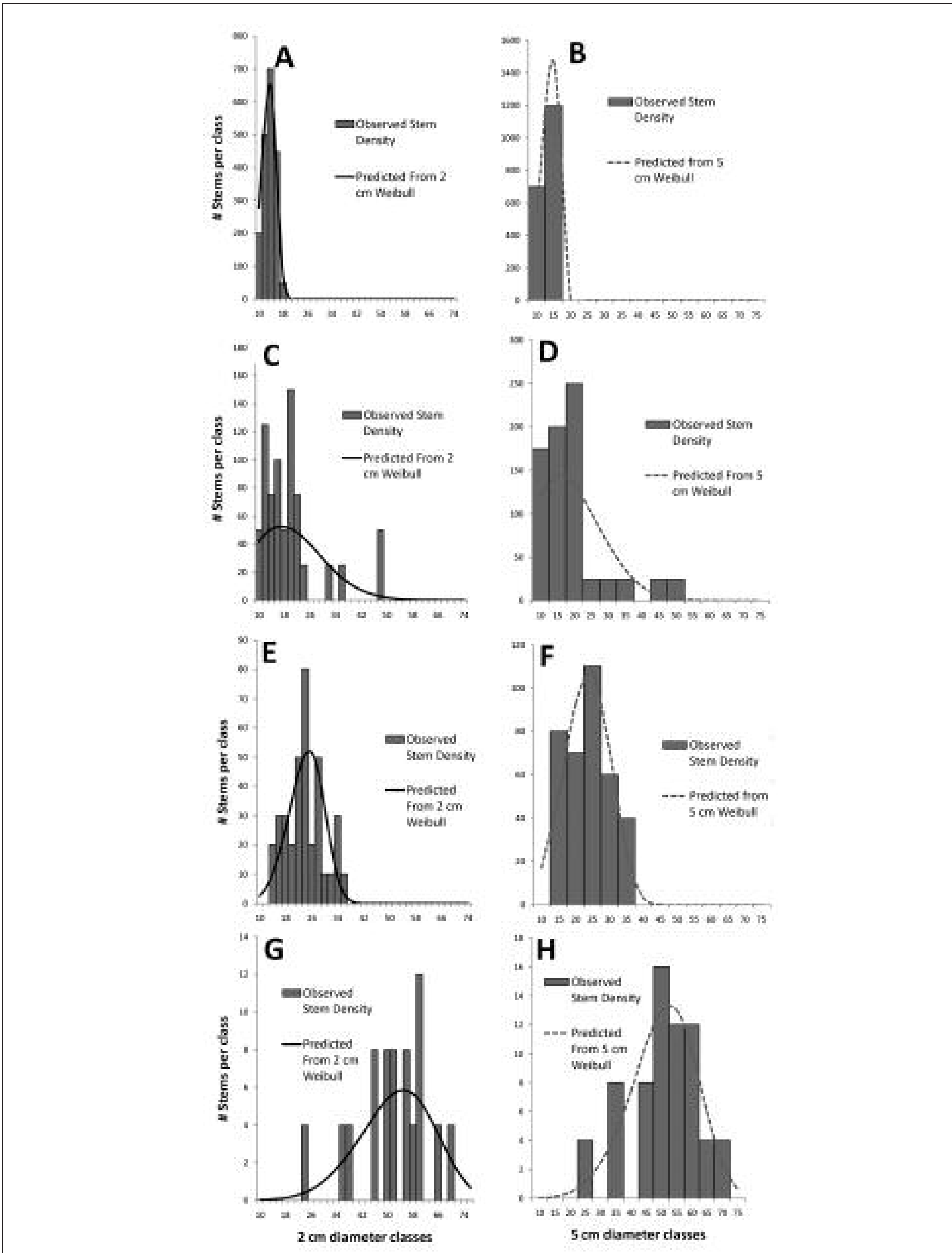

Fig. 2. Weibull estimates for 2-cm and 5-cm density distributions for the coniferous and mixedwood plots (C&MW) grouped according to the D/H metric as follows: A and B) The first C&MW group (D/H > 50) at 2-cm and 5-cm increments, respectively; C and D) the second C&MW group (25 < D/H ≤ 50); E and F) the third C&MW group (15 < D/H ≤ 25); G and H) the fourth C&MW group (D/H ≤ 15).

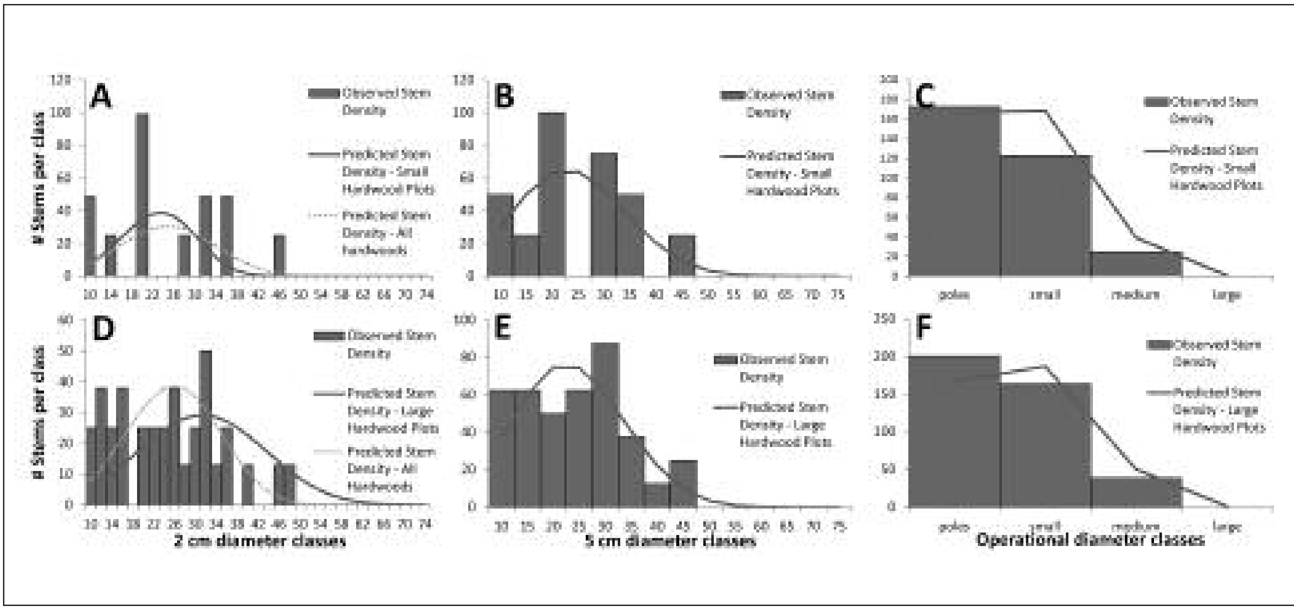

Fig. 3. Diameter distributions and LiDAR-predicted Weibull models for a small-area hardwood plot (A–C, not significant) and a large-area hardwood plot (D–F).

Table 6. Weibull MLR models and statistics for large hardwood plots derived from LiDAR data

D.V.[a]	Linear predictive model	r^2_{adj}	F	p	RMSE	RMSE cross validation
BA	14.8 - 3.06sd + 0.896p90 + 204d2 - 227d3	0.83	22.30	0.00	2.80 m^2/ha	4.09 m^2/ha
N	7.44 + 22.8d2 - 28.4d3 - 9.92d4 - 2.86d7 - 7.24d10	0.78	13.71	0.00	0.67 #/ha	0.61 #/ha
Na	-21.4 - 265d2 + 387d3 + 276d4 + 103d7 + 37.8d9 + 146d10	0.88	23.62	0.00	3.00	4.73
Nb	1.56 + 0.0241p20 - 0.0358p90 - 23.5d1 - 11.8d2 + 15.8d3	0.80	15.23	0.00	1.38	0.09
BAa	-22.5 + 1.44p90 + 253d4 + 62.8d7 + 127d10	0.83	23.55	0.00	4.11	6.05
BAb	1.99 - 0.0922sd - 16.8d1 + 10.5d3 - 3.57d6 + 1.86d7 - 2.75d10	0.65	6.21	0.00	0.95	0.64

[a]D.V. = dependent variable

Table 7. Finite mixture model MLR regression statistics derived from LiDAR data

D.V.[a]	Linear predictive model	r^2_{adj}	F	p	RMSE	RMSE cross validation
ρ_1	1.91 - 0.104sd + 2.81d1 - 1.76d2 - 3.19d6 - 2.11d8 - 2.22d10[c]	0.55	4.4	0.02	0.16	0.35
a_1	-331[c] + 11.2p90[c] + 1603d3 - 1542d5 + 2154d6	0.76	20.01	0.00	65.6	114.2
b_1	41.8 - 2.32sd + 161d1 - 127d10	0.77	20.39	0.00	5.34	7.08
ρ_2	-0.912[d] + 0.104sd - 2.81d1 + 1.76d2 + 3.19d6 + 2.11d8 + 2.22d10[c]	0.55	4.4	0.02	0.16	0.35
a_2[b]	-0.283 + 0.0191p90 - 0.996d1 - 0.438d4	0.77	20.42	0.00	24.4	29.4
b_2	25.2[c] - 3.05mean + 1.23p10 + 2.85 p90 + 88.4d4 - 122d10	0.77	12.38	0.00	5.28	7.81

[a]D.V. = dependent variable;
[b]1/x transformation applied for normality. RMSE values are back-transformed.
[c]term significant at 80% confidence
[d]term significant at 90% confidence.
Note: RMSE ranges from 12%–20% for the a1 term.

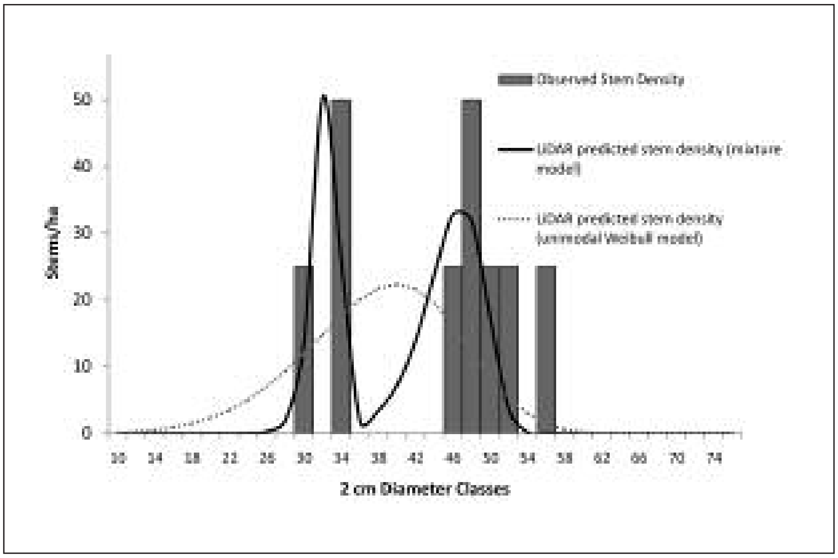

Fig. 4. Weibull finite mixture models for a bi-modal distribution derived from LiDAR data.

$\mu_{w1} = 23$, $\sigma_{w1} = 2$, $\rho_1 = 0.33$ and $\mu_{w2} = 58$, $\sigma_{w2} = 3.6$, $\rho_2 = 0.67$. Based on the mixture model parameters, one would expect a clump of trees with a mean DBH of around 23 cm, with a standard deviation of 2 cm. The ρ parameter suggests that approximately 33% of the trees are within this mode. The second mode occurs at a mean of approximately 58 cm, with a standard deviation of 3.6 cm. Approximately 67% of the trees are within this second clump. Without even examining the graph, one can tell that the 2 clumps are separated by a noticeable gap (given the lack of overlap in the DBH ranges for each clump). If this information is combined with the average plot stem density predicted from LiDAR, the number of trees in each clump can be estimated. As is clearly highlighted in Fig. 4, in the case where significant gaps in the diameter distribution are evident, mixture modeling will reduce the overestimation in stem density that will always be present with the unimodal 2-parameter Weibull distribution.

Discussion

Use of a D/H metric to stratify field plots is a quick, systematic and repeatable approach that may increase operational efficiency of the Weibull modeling approach. Prior knowledge of individual species is not needed, although separation of hardwoods from softwoods was necessary for the success of this study. Examination of the histograms of the hardwood plots demonstrated that plot size is an important consideration for this group. Slightly more than half of the plots had a plot size of approximately 0.04 hectares, which is consistent with the National Forest Inventory (NFI) standards for the large-tree plot size (CCFM 2004). However, the expected reverse "J-Shape" distribution is not apparent in the histograms of many of the smaller plots (e.g., Fig. 3A). This shape is often thought to be characteristic of the diameter class distributions of mature stands that have reached an equilibrium state (Westphal *et al.* 2006). If diameter distributions are an operational goal for hardwood stands, it may be necessary to either combine multiple small field plots to capture the complete distribution or to collect field data from larger plots. In either case, the total area required for field sampling will be larger for hardwood stands if the J-shaped diameter distribution is to be captured (Peng 2000).

For the coniferous and mixedwood stands, use of the D/H metric captured the characteristics of even- and uneven-aged stands. For example, the first C&MW group (D/H > 50) has a similar appearance to young, even-aged pure stands, such as the 20-year-old oak stand described in Husch *et al.* (2003). This is a logical comparison, since the D/H stratification resulted in a group of homogenous plots with densely-spaced small trees. As the D/H metric decreases, the diameter distributions of the field plots appear increasingly irregular, which is typical of uneven-aged stands, particularly those for small areas (Peng 2000). This can be seen in Fig. 2C, E, G and it was most apparent for the smallest D/H metric (Fig. 2G). Unfortunately, irregular uneven-aged stands are commonly seen in the forested landscape and difficulty modeling these distributions with the classic unimodal Weibull approach has been encountered by other authors (Zhang *et al.* 2001, Liu *et al.* 2002, Zasada and Cieszewski 2005, Zhang and Liu 2006, Maltamo *et al.* 2007). The complexity of uneven-aged stands also makes it difficult to model their growth and yield, given that stand age and site index are commonly used to assess site quality for many growth and yield models (Peng 2000). This has resulted in the development of growth and yield techniques that do not rely on age as a predictor variable (Monserud and Sterba 1996, O'Hara 1996, Peng 2000). The D/H metric is consistent with this approach, as it relies only on the number of trees and their height.

There are 2 approaches that could be used to obtain more reliable models of diameter distributions of uneven-aged stands. Peng (2000) suggests that increasing the sampled area will eventually remove the irregularities in the diameter distribution, which would therefore make it possible to model stands using the unimodal Weibull distribution. However, this is not a practical operational solution because it increases the expense of field sampling and does not produce reliable results for small areas. On the other hand, the relatively understudied mixture model approach appears to be a flexible solution where diameter distributions for small uneven-aged stands are required (Zhang and Liu 2006) (e.g., Fig. 4). Mixture modelling was untested in this study for the hardwood plots, but may reduce the need for larger plot sizes in hardwood stands. As previously mentioned, the drawback to the mixture modelling approach is the increase in number of parameters required to define the model. While it may be possible to accurately model the diameter distribution with multiple Weibull modes, as the number of parameters increase it becomes increasingly unlikely that the entire model can be accurately predicted from LiDAR point clouds. Although this study suggests that LiDAR predications of diameter distributions are possible for a 2-mode mixture model, more studies are required before this approach can be considered operationally reliable.

Conclusions

The ability to predict Weibull parameters from LiDAR data using multiple regression analysis has been demonstrated. This has important implications for our ability to map diameter distribution information over very large areas. The classic, unimodal Weibull shape has been shown to accurately characterize young, homogenous stands. Stratification based on number of stems and tree height demonstrates an increased irregularity in diameter distributions for heterogeneous, uneven-aged stands. Without stratification, 2-parameter Weibull models do not correlate well to LiDAR metrics. However, initial stratification into similar structural groups suggest the Weibull model is a promising technique, with strong predictive capabilities evident for several subgroups ($r^2_{adj} > 0.8$ for many of the parameters reported when predicting the Weibull parameters from LiDAR data). The necessity of stratification has implications for operational management using the unimodal technique, particularly for areas that contain both managed and natural stands, with very different diameter distributions. At the very least, enough field plots must be measured to enable the development of valid statistical models for each major stratification.

The findings suggest that a better approach to model irregular distributions would be to predict a 2-mode Weibull mixture model from LiDAR data. If this technique were adopted universally, homogenous even-aged stands could still be represented—the scale parameter would be the same for both modes. However, although the initial attempts at finite mixture modelling on a stratified subgroup appear promising, further work is required to determine if mixture models can be reliably predicted from LiDAR data for different forest types. If so, prediction of mixture models from LiDAR data may prove to be an effective operational tool for modeling diameter distributions and growth and yield of homogenous and heterogeneous stands.

Acknowledgements

The authors would like to thank the Associate Editor for this special issue, Dr. Ron Hall, and the anonymous reviewers for their efforts and suggestions regarding this manuscript. We also gratefully acknowledge the contributions of the Forest Research Partnership (comprised of the Canadian Ecology Centre, the Canadian Forest Service, the Ontario Ministry of Natural Resources, and Tembec Inc.) and the Enhanced Forest Productivity Science Fund.

References

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

[CCFM]. Canadian Council of Forest Ministers. 2004. Canada's National Forest Inventory Ground Sampling Guidelines Version 4.1 [online]. Available at https://nfi.nfis.org/documentation/ground_plot/Gp_guidelines_v4.1.pdf [accessed March 3, 2008].

Gobakken, T. and E. Næsset. 2004. Estimation of Diameter and Basal Area Distributions in Coniferous Forest by Means of Airborne Laser Scanner Data. Scand. J. For. Res. 19: 529–542.

Hudak, A.T., N.L Crookston, J.S. Evans, M.J. Falkowski, A.M.S. Smith, P.E. Gessler and P. Morgan. 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return LiDAR and multispectral satellite data. Can. J. Remote Sensing 32 (2): 126–138.

Husch, B., T.W. Beers and J.A. Kershaw Jr. 2003. Forest Mensuration, 4th Edition. John Wiley and Sons, New York.

Hyde, P., R. Nelson, D. Kimes and E. Levine. 2007. Exploring LiDAR–RaDAR synergy—predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR. Rem. Sens. Environ. 106(1): 28–38.

Lane, D.M. 2007. HyperStats Online Statistics Textbook [online]. Available at http://davidmlane.com/hyperstat/A43646.html [accessed March 3, 2008].

Lim, K.S. and P.M. Treitz. 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand. J. For. Res. 19(6): 558–570.

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

Magnussen, S. and P. Boudewyn. 1998. Derivation of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can. J. For. Res. 28: 1016–1031.

Maltamo, M., J. Puumalainen and R. Paivinen. 1995. Comparison of beta and Weibull functions for modeling basal area diameter distribution in stands of *Pinus sylvestris* and *Picea abies*. Scand. J. For. Res. 10: 284–295.

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

Monserud, R.A. and H. Sterba. 1996. A basal area increment model for individual trees growing in even-and uneven-aged forest stands in Austria. For. Ecol. Manag. 80: 57–80.

Næsset, E. 2004. Practical large-scale forest stand inventory using small-footprint airborne scanning laser. Scand. J. For. Res. 19: 164–179.

Nanang, D.M. 1998. Suitability of the normal, log-normal and Weibull distributions for fitting diameter distributions of neem plantations in Northern Ghana. For. Ecol. Manag. 103: 1–7.

[NIST] National Institute of Standards and Technology. 2006. NIST/SEMATECH e-handbook of statistical methods. [online]. Available at http://www.itl.nist.gov/div898/handbook/ [accessed March 4, 2008].

O'Hara, K.L. 1996. Dynamics and stocking-level relationships of multi-aged Ponderosa pine stands. For. Sci. 42: 1–34.

Peng, C. 2000. Growth and yield models for uneven-aged stands: past, present, and future. For. Ecol. Manag. 132: 259–279.

Sprugel, D.G. 1983. Correcting for bias in log-transformed allometric equations. Ecology 64(1): 209–210.

Thomas, V., D.A. Finch, J.H. Mccaughey, T. Noland, L. Rich and P. Treitz. 2006a. Spatial modelling of the fraction of photosynthetically active radiation absorbed by a boreal mixedwood forest using a LiDAR-hyperspectral approach. Agric. For. Met. 140: 287–307.

Thomas, V., P. Treitz, J.H. McCaughey and I. Morrison. 2006b. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using LiDAR: an examination of scanning density. Can. J. For. Res. 36, 34–47.

Westphal, C., N. Tremer, G. von Oheimb, J. Hansen, K. von Gadow and W. Hardtle. 2006. Is the reverse J-shaped diameter distribution universally applicable in European virgin beech forests? For. Ecol. Manag. 223: 75–83.

Zasada, M. and C.J. Cieszewski. 2005. 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. 204: 145–158.

Zhang, L., J.H. Gove, C. Liu and W.B. Leak. 2001. A finite mixture distribution for modeling the diameter distribution of rotated-sigmoid, uneven-aged stands. Can. J. For. Res. 31: 1654–1659.

Zhang, L. and C. Liu. 2006. Fitting irregular diameter distributions of forest stands by Weibull, modified Weibull, and mixture Weibull models. J. For. Res. 11: 369–372.