Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry, Vol. 84, No. 4, 2011. doi:10.1093/forestry/cpr033
Advance Access publication date: 24 August 2011

Predicting the number of trees in small diameter classes using predictions from a two-parameter Weibull distribution

ELIZABETH MCGARRIGLE[1,*], JOHN A. KERSHAW, Jr[1], MICHAEL B. LAVIGNE[2], AARON R. WEISKITTEL[3] AND MARK DUCEY[4]

[1] Faculty of Forestry and Environmental Management, University of New Brunswick, 28 Dineen Drive, Fredericton, NB E3B 5A3 Canada
[2] Natural Resources Canada, Canadian Forest Service – Atlantic Forestry Centre, P.O. Box 4000, Fredericton, NB E3B 5P7 Canada
[3] School of Forest Resources, 5755 Nutting Hall, The University of Maine, Orono, ME 04469, USA
[4] Department of Natural Resources and the Environment, University of New Hampshire, 114 James Hall, Durham, NH 03824, USA
*Corresponding author. E-mail: emcgarrigle@gmail.com

Summary

Knowing the abundance of small trees is necessary for accurate calculation of gross production, total carbon and/or biomass of forest stands. The abundance of small trees can also be used to predict ingrowth into larger tree diameter classes. We present a method of predicting numbers of trees in small diameter classes using diameter distributions of larger trees in stands. A truncated two-parameter Weibull distribution was fit to large tree diameters (diameter at breast height (d.b.h.) ≥ 9.0 cm). These parameters were then used to predict the number of small trees in d.b.h. classes smaller than 9.0 cm. Three methods of predicting densities of small trees were used: (1) an extrapolation of the truncated Weibull to a full two-parameter Weibull distribution; (2) a modification of the Weibull using an empirical estimate and (3) a combined approach. While the full two-parameter Weibull distribution generally fitted the distribution of small trees, densities were typically under-predicted. The empirical method (i.e. method 2) produced the best predictions of small tree densities, with a root mean square error of 132 trees h^{-1} (28 per cent of mean small tree density). Overall, predicting the distribution of small trees using the distribution of large trees worked very well in this study.

Introduction

Collection of inventory data above a minimum diameter threshold is common practice (Husch *et al.*, 2003; Curtis and Marshall, 2005). Trees above a predefined minimum diameter are measured, while those below this threshold are not. Thresholds are often selected based on merchantability limits for products in a region. Measurement of trees below the minimum threshold adds significant field time, with relatively small contributions to stand-level merchantable totals. In addition, these trees often have high turnover rates making them difficult to track over time on remeasured sample plots, resulting in a significant increase in effort searching for missing trees and tagging new cohorts at each periodic measurement. Nevertheless, estimates of small trees are important for many applications. For example, calculations of gross production, total biomass or carbon of stands require small tree information.

Growth and yield models need to be able to predict ingrowth – the recruitment of stems into diameter classes above a diameter threshold – to accurately forecast long-term stand dynamics in naturally regenerated forests (Garcia, 1988; Shifley *et al.*, 1993). Most ingrowth models account for very little of the variation in observed ingrowth from remeasured sample plots (e.g. Adame *et al.*, 2010). Predictions of ingrowth are improved if estimates of the number of trees below the threshold diameter (small trees) are available (Lundqvist, 1995). A method to predict diameter distributions of small trees may provide useful information to extrapolate tree lists of the smaller trees or at least provide an additional independent variable to improve prediction of either small tree distributions or ingrowth.

The Weibull distribution, first introduced to forestry by Bailey and Dell (1973), is one of the more common distributions used to predict diameter distributions (e.g. Bailey and Dell, 1973; Little, 1983; Zutter *et al.*, 1986; Zhang and Liu, 2006). The Weibull distribution can represent a wide range of unimodal distributions, including skewed and mound-shaped curves (Little, 1983). Also, the Weibull distribution can be modified to predict distributions for truncated data. In cases when inventory data are collected only for trees above a minimum threshold diameter, the left-truncated Weibull is effectively used for modelling such diameter distributions using the threshold diameter as the truncation point (Zutter *et al.*, 1986).

Predicting diameter distributions using a truncated Weibull distribution produces estimates of distribution of trees among diameter classes above the minimum diameter or truncation point. Merganič and Sterba (2006) showed that estimated Weibull parameters from a truncated dataset deviated very little from parameters estimated using a complete dataset, except when the diameter distributions were bimodal. This implies that distributions of diameters below the truncation point could be predicted using the estimated parameters for scale and shape from the truncated distribution. More flexibility in the prediction of small trees can be attained by applying an empirical adjustment to the theoretical distribution predicted by the Weibull distribution. An empirical model can be used to produce an estimate of small trees based on the Weibull-based prediction as well as variables typically found significant when modelling ingrowth: stand density, site index, species composition, average tree size or quadratic mean diameter (Shifley *et al.*, 1993; Lexerod and Eid, 2005; Adame *et al.*, 2010).

In this study, we assess the applicability of using left-truncated Weibull parameter estimates based on data collected above a threshold diameter to predict distributions of trees below the threshold diameter in a range of stand structures in New Brunswick, Canada. Initial goodness of fits are assessed, and a system of equations to modify the Weibull predictions is developed. Our objective is to assess alternative methods of using the Weibull distribution to estimate small tree diameter distributions.

Methods

Permanent sample plot data

Data for this analysis come from a network of permanent sample plots in New Brunswick, Canada. These plots span the province of New Brunswick and consist of a mix of multi-species multi-cohort stands typical of the Acadian forest region. Each plot is 400 m^2 in size, and diameter at breast height (d.b.h.) and height data are collected for every tree above a minimum d.b.h. of either 1.0 or 5.0 cm. Data collection began in 1987 with plot remeasurements completed every 3–5 years. For this study, only the first plot measurements were used.

A total of 1621 plots were used in this analysis. On 266 plots, all trees with d.b.h. greater than or equal to 1.0 cm were measured. On the remaining 1355 plots, all trees with d.b.h. greater than or equal to 5.0 cm were measured. The majority, ~56 per cent, of plots were dominated by spruce (primarily *Picea rubens* Sarg., and *Picea mariana* (Mill.) B.S.P.) and 30 per cent were dominated by balsam fir (*Abies balsamea* (L.) Mill.). The remaining plots were dominated by pine (primarily *Pinus banksiana* Lamb.), intolerant hardwood, tolerant hardwood or other softwood species.

Truncated distribution modelling

A common truncation point was set at 9.0 cm as this is a typical truncation point used in the region. Trees on each plot were divided into two groups: (1) those greater than or equal to 9.0 cm d.b.h. and (2) those less than 9.0 cm d.b.h.. The goal of this study was to predict number and diameter distribution of trees in category 2 based on these characteristics for trees in category 1.

The two-parameter truncated Weibull distribution was fit to the diameter distribution for trees ≥9.0 cm d.b.h.. The probability density function for the two-parameter truncated Weibull is specified by:

$$\mathrm{ft}(x) = \left(\frac{c}{b}\right)\left(\frac{x}{b}\right)^{c-1} \mathrm{e}^{\left(\left(\frac{t}{b}\right)^{c} - \left(\frac{x}{b}\right)^{c}\right)} \tag{1}$$

and the cumulative density function is specified by:

$$\mathrm{Ft}(x) = 1 - \mathrm{e}^{\left(\left(\frac{t}{b}\right)^{c} - \left(\frac{x}{b}\right)^{c}\right)}, \tag{2}$$

where x is a Weibull variate (d.b.h., in this case; $x \geq t$), b is the scale parameter ($b > 0$), c is the shape parameter ($c > 0$) and t is the truncation point ($t \geq 0$). Maximum likelihood estimates of the parameters were obtained for each plot using a modification of an algorithm from Robinson (2004) for the reverse Weibull distribution (Kershaw, 2011) with a minimum bound of 1.6 imposed on the shape parameter. Goodness of fit was assessed using the Kolmogorov–Smirnov (K-S) test (Zar, 2009). Since very few (22 plots) of the observed d.b.h. classes had distributions approaching a negative exponential distribution (shape = 1.0), the lower bound for the shape parameter was set at 1.6 rather than 1.0. For most of the highly right-skewed distributed plots, this choice also produced better fits as measured by the K-S test. An example of a truncated Weibull fit is shown in Figure 1.

Small tree prediction

Three approaches are compared for predicting small tree numbers: (1) full Weibull distribution, (2) empirical estimation and (3) a combination approach.

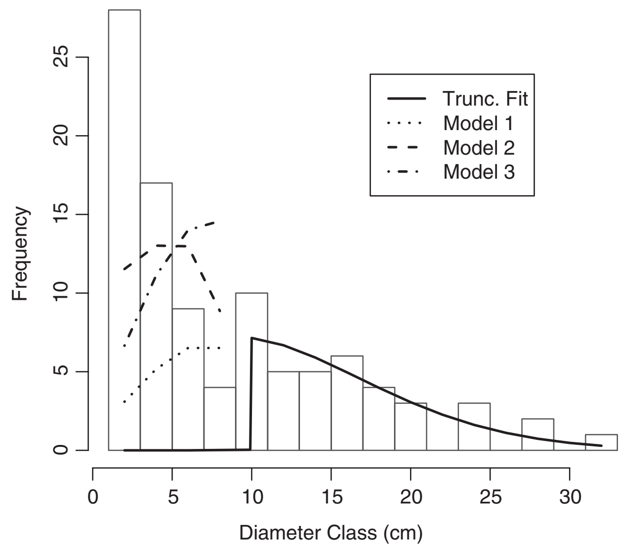

Figure 1. Example histogram of observed diameters *vs* the predicted truncated Weibull distribution for large trees (shape parameter equal to 1.8, scale parameter equal to 13.8) and the predicted densities of small trees by diameter class and estimation method.

Full Weibull distribution approach (model 1)

The scale and shape parameters estimated for the left truncated distributions described above were used to predict the small tree distributions. The truncated scale and shape parameters were used in the two-parameter Weibull distribution to calculate the numbers of small trees (Figure 1). The cumulative two-parameter Weibull distribution is specified by:

$$F(x) = 1 - e^{\left(-\left(\frac{x}{b}\right)^c\right)}. \tag{3}$$

Using equation (3), the following steps were used to estimate small tree distributions:

1. Total tree density was estimated by dividing observed density for trees ≥ 9.0 cm d.b.h., N_9, by the probability of being larger than 9.0 cm based on the estimated scale and shape parameters:

$$\hat{N}_{\mathrm{Tot}} = \frac{N_9}{1 - P(x \geq 9.0)} = \frac{N_9}{1 - \left(1 - e^{\left(-\left(\frac{9}{b}\right)^c\right)}\right)} = \frac{N_9}{e^{\left(-\left(\frac{9}{b}\right)^c\right)}}. \tag{4}$$

2. The number of small trees in cumulative subsets (7.0–9.0, 5.0–9.0, 3.0–9.0 and 1.0–9.0) was estimated using $\hat{N}_{\mathrm{Tot}}$ and the scale and shape parameters:

$$\hat{N}_{W(S)} = \hat{N}_{\mathrm{Tot}}\left(P(x \geq 9.0) - P(x < S)\right) = \hat{N}_{\mathrm{Tot}}\left(e^{\left(-\left(\frac{9}{b}\right)^c\right)} - e^{\left(-\left(\frac{S}{b}\right)^c\right)}\right), \tag{5}$$

where S = the lower bound of the small tree distribution (1.0, 3.0, 5.0 or 7.0 cm)

3. Numbers of trees for each 2.0-cm midpoint d.b.h. class (2.0, 4.0, 6.0 and 8.0 cm) were obtained by subtraction:

$$\hat{N}_{\mathrm{W}(D)} = \hat{N}_{\mathrm{W(S+2)}} - \hat{N}_{\mathrm{W(S)}}. \tag{6}$$

To test the sensitivity of this approach to threshold diameter, an additional analysis for the full Weibull approach using a 7-cm truncation point was also conducted. Weibull parameters for the left truncated distribution were determined using 7 cm as the truncation point and the same method described above was used to determine the number of small trees in each 2-cm diameter class less than 7 cm.

Empirical estimation approach (model 2)

A nested additive model was developed to predict number of trees in the cumulative subsets, $\hat{N}_{E(S)}$. Using the predicted numbers of small trees based on the full Weibull distribution, $\hat{N}_{W(S)}$, the following non-linear mixed effects regression equation was fitted:

$$\begin{aligned}\hat{N}_{E(s)} &= \left(b_{0j} + b_{S,0j} \cdot S\right) \\ &\quad + (\beta_0 + \beta_{S,0} \cdot S)\left(\hat{N}_{W(S)}{}^{(\beta_1 + \beta_{S,2} \cdot S)}\right) e^{\left((\beta_2 + \beta_{S,2} \cdot S)\mathrm{SI}_{\mathrm{Clim}} + -\left(\beta_3 + \beta_{S,3} \cdot S\right)\hat{\mathrm{dbh}}_Q\right)},\end{aligned} \tag{7}$$

where $\hat{N}_{E(S)}$ is the observed number of trees between S and 9.0 cm, S is the lower bound of the cumulative small tree subset; $\hat{N}_{W(S)}$ is the predicted number of small trees based on the full Weibull distribution; $\mathrm{SI}_{\mathrm{Clim}}$ is a climate-based prediction of site index (Weiskittel *et al*., in press); $\hat{\mathrm{dbh}}_Q$ is the quadratic mean diameter of trees greater than or equal to 9.0 cm diameter; the β_i and $\beta_{S,i}$ are fixed effect parameter estimates and b_{ij} are estimated plot-level random effects.

The climate-based site index is the work of Weiskittel *et al*. (2010). It was developed using height-age data collected on permanent sample plots and climate information for this region produced by the US Forest Rocky Mountain Research Station Moscow Laboratory (2010). A total of 35 climate normal (1961–1990) variables were initially input into random forests, a non-parametric regression model (Breiman, 2001; Weiskittel *et al*., 2010). The model was run iteratively and the most parsimonious model included seven climate variables and explained 73.6 per cent of the original variation in site index and had a root mean square error of 1.04 m (Weiskittel *et al*., 2010).

To account for autocorrelation within plots, a continuous autocorrelation structure was modelled; however, the autocorrelation structure was found to not be significant (P > 0.05) and was not included in the final model. Similarly, a weighted model using a power function of N_9 was also fit but did not improve predictions and was not used in the final model. As with the full Weibull approach, the numbers of trees in each small tree d.b.h. class were obtained by subtraction.

Combined approach (model 3)

For many plots, the scale and shape parameters appeared to produce the 'right' distribution form to fit the small trees; however, the density estimated from the full Weibull distribution, $\hat{N}_{W(S)}$, appeared incorrect. So the third approach to predict small tree densities combined the full Weibull distribution approach with the empirical approach. Numbers of small trees between 1.0 and 9.0 cm d.b.h. were estimated using the non-linear mixed effects model (equation 7) with S equal to 1.0, and the trees were distributed using the relative probabilities predicted from the full Weibull distributions and the scale and shape parameter estimates (Figure 1). Relative probability was calculated as the ratio of probability within a d.b.h. class to the cumulative probability of being less than 9.0 cm d.b.h.:

$$\hat{N}_{C(S)} = \hat{N}_{W(S)}\left(\hat{N}_{E(S)} / \hat{N}_{\mathrm{Tot}}\right), \tag{8}$$

Model validation

Two approaches were used to validate empirical model predictions. In the first approach, the second measurements of permanent sample plots were used to make small tree density predictions. The Weibull parameters were fit for each plot using equation (1). The parameters were then used to predict small trees using the empirical approach (equation 7). Parameters and random effects for each plot remained the same as those calculated using the first plot measurements. Error, bias and deviance were calculated for the model predictions using the full model including the random effects and fixed effects portion only.

The second approach to model validation was to use 75 per cent of the plots for fitting the empirical model (equation 7) and 25 per cent for validating the model. This approach examined the fixed effects portion of the model only. Error, bias and deviance were calculated for both the fitted and validation plots.

Model evaluation

Goodness of fit for each cumulative subset was assessed using the Kolmogorov–Smirnov test, and three measures of prediction accuracy were used to assess predicted numbers in each small tree d.b.h. class for each method:

$$\text{Error} = \sqrt{\frac{\sum_{i=1}^{n}\left(N_{Di} - \hat{N}_{Di}\right)^2}{n}}, \tag{9}$$

where N_{Di} and $\hat{N}_{Di}$ are, respectively, observed and predicted number of trees in the Dth class in plot i.

$$\text{Bias} = \frac{\sum_{i=1}^{n}\left(N_{Di} - \hat{N}_{Di}\right)}{n} \tag{10}$$

and

$$\text{Deviance} = \frac{\sum_{i=1}^{n}\left|N_{\mathrm{Di}} - \hat{N}_{\mathrm{Di}}\right|}{n} \tag{11}$$

Results

Weibull distribution fits (model 1)

The two-parameter left-truncated Weibull distribution fit the observed distributions of diameters above the 9-cm truncation point very well (Table 1). P-values were classified into four broad categories: very good, VG ($P \geq 0.50$); good, G ($0.25 \leq P < 0.50$); adequate, A ($0.05 \leq P < 0.25$) or bad, B ($P < 0.05$). Ninety-two per cent of the plots had what we classified as good or very good fits. Plots classified as bad fits had distributions that were either bimodal or had very long tails with gaps in the distribution.

Results of applying the parameters estimated from the truncated data to the small trees were more variable and depended upon which small trees were being estimated (Table 1). Approximately 65 per cent of the plots had good or very good fits when considering only the 7.0–9.0 cm d.b.h. small trees; however, this decreased to 20 per cent when considering all small trees between 1.0 and 9.0 cm d.b.h.. Almost 66 per cent of the plots had distributions that were significantly different from the distribution predicted from the left-truncated Weibull distribution (Table 1).

In terms of predicted densities by diameter class using a 9-cm truncation (Table 2), the 7.0- to 9.0-cm diameter

Table 1: Number of plots by goodness of fit (K-S test) and diameter subset based on parameter estimates for a two-parameter left-truncated Weibull applied to trees ≥9.0 cm d.b.h.

Plot and distribution subset	Goodness of fit: Very good ($P > 0.50$), n (%)	Good ($0.25 \leq P < 0.50$), n (%)	Adequate ($0.05 \leq P < 0.25$), n (%)	Bad ($P < 0.05$), n (%)
All plots				
≥ 9.0 cm	1269 (78%)	228 (14%)	93 (6%)	31 (2%)
Plots with trees 1.0–9.0 cm				
≥1.0 < 9.0 cm	40 (15%)	14 (5%)	37 (14%)	175 (66%)
≥3.0 < 9.0 cm	55 (21%)	35 (14%)	56 (21%)	116 (44%)
≥5.0 < 9.0 cm	81 (31%)	48 (19%)	58 (22%)	72 (28%)
≥7.0 < 9.0 cm	119 (48%)	52 (21%)	55 (22%)	24 (9%)
Plots with trees 5.0–9.0 cm				
≥5.0 < 9.0 cm	463 (36%)	282 (21%)	320 (24%)	256 (19%)
≥7.0 < 9.0 cm	556 (42%)	314 (24%)	307 (24%)	132 (10%)

class had an average error of just less than 200 trees ha^{-1}, while the 1.0- to 3.0-cm diameter class had an average error of almost 650 trees ha^{-1}. Since the plots used in this study were 0.04 ha in size, this translates into ~8 and 26 plot trees, respectively. Bias and deviance were also highest in the smallest diameter class, 328 and 375, respectively, compared with 88 and 131 in the 7.0- to 9.0-cm diameter class (Table 2). Bias indicates a consistent underestimate of small trees in all diameter classes (Table 2). Trends over diameter classes were the same when the 7-cm truncation was used, with improvements in all goodness-of-fit measures in all diameter classes (Table 2).

Empirical estimation (model 2)

Parameter estimates and associated SEs for equation (7) are shown in Table 3. As climate-based site index increases or quadratic mean diameter decreases, the predicted number of small trees increases (Figure 2). Though limited in non-linear regression, an overall sense of performance of the model is given by the fit statistics. The non-linear pseudo-R^2 for the fixed effects was 0.51 and for the combined fixed and random effects was 0.96. Root mean squared error (rMSE) associated with the fixed effects was 369, and the overall residual rMSE, including the random effects, was 132. With a tree factor of 25 trees ha^{-1}, a residual SE of 132 represents an error of ±6 small trees per plot or ~28 per cent of the mean number of small trees per plot.

Table 2: Error, bias and deviance in predicted densities for the full Weibull approach (model 1) by truncation point and small tree diameter class

Truncation	d.b.h. class	Goodness-of-fit measure		
		Error	Bias	Deviance
7 cm	1–3	636	322	364
	3–5	261	118	169
	5–7	202	98	136
9 cm	1–3	648	328	375
	3–5	288	125	188
	5–7	258	117	170
	7–9	194	88	131

Table 3: Model 2 parameter estimates and associated SE and *P*-value for equation (7), estimating numbers of small trees by diameter subset

Parameter	Estimate	SE	*P*-value
β_0	470.7893	120.5035	0.0001
$\beta_{S,0}$	−39.9632	11.2967	0.0004
β_1	0.2751	0.0213	0.0000
$\beta_{S,1}$	0.0224	0.00218	0.0000
β_2	0.0724	0.0147	0.0000
$\beta_{S,2}$	−0.0094	0.0020	0.0000
β_3	0.0728	0.0089	0.0000
$\beta_{S,3}$	0.0062	0.0009	0.0000

Predicted densities by diameter class using the empirical model (model 2) are presented for both the fixed effects alone (Table 4, Figure 3D) and including random effects (Table 4, Figure 3E). Values for the fixed effects only are averages for all plots in the study, while averages for fixed and random effects excluded 92 plots with negative density predictions. Negative predictions were a result of the additive structure, particularly the random effects associated with the diameter limit, *S*. The average error using fixed effects only was 139 in the 7.0–9.0 diameter class and increases to 578 in the 1.0–3.0 diameter class, translating roughly to 6 and 23 trees, respectively. Deviance also decreases as diameter class increases, ranging from 391 in the 1.0–3.0 diameter class to 100 in the 7.0–9.0 diameter class. Bias was within ±10 for all diameter classes, except the 3.0–5.0 class which was overestimated by 92 trees ha^{-1} (Table 4).

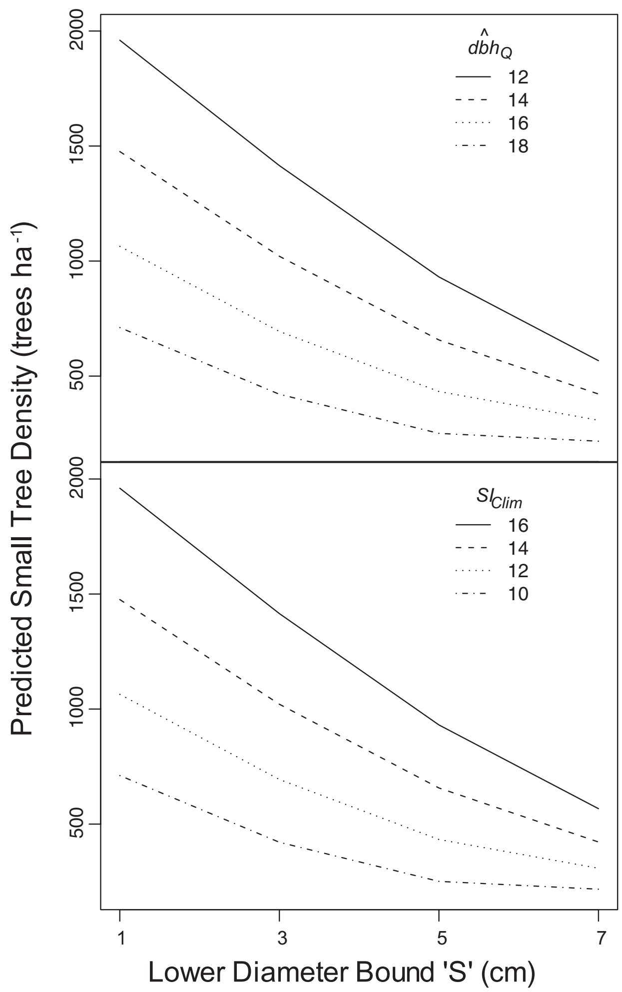

Figure 2. Predicted small tree densities from the empirical model *vs* quadratic mean diameter ($\hat{\mathrm{dbh}}_Q$) and climate-based site index ($\mathrm{SI}_{\mathrm{Clim}}$). Constants were taken from a plot in the database – model 1 small tree density predictions ($\hat{N}_{W(S)}$) were held constant for all variations (density by lower bound 'S': 1-1320, 3-1130, 5-811, 7-411), $\mathrm{SI}_{\mathrm{Clim}}$ held constant at 13.5 for varying levels of $\hat{\mathrm{dbh}}_Q$, and $\hat{\mathrm{dbh}}_Q$ held constant at 16.4 for varying levels of $\mathrm{SI}_{\mathrm{Clim}}$.

Table 4: Error, bias and deviance in predicted densities by estimation method and small tree d.b.h. classes for models 2 and 3

Method	d.b.h. class	Goodness-of-fit measure Error	Bias	Deviance
Empirical fixed effects	1–3	578	7	392
	3–5	270	−92	214
	5–7	215	−6	152
	7–9	139	7	100
Empirical fixed and random effects	1–3	433	99	274
	3–5	103	−35	77
	5–7	91	−9	63
	7–9	74	8	54
Combination fixed effects	1–3	629	271	365
	3–5	260	−7	182
	5–7	228	12	158
	7–9	194	−68	143
Combination fixed and random effects	1–3	656	359	392
	3–5	152	51	104
	5–7	129	16	84
	7–9	170	−65	89

Plots with negative predictions using the empirical model with fixed and random effects are removed.

Average errors associated with model 2 including the random effects (Table 4) are lower than with fixed effects alone for all diameter classes (Table 4). The 1.0–3.0 cm d.b.h. class had an average error of 433 or ~18 trees per plot, while the 7.0–9.0 cm d.b.h. class had an average error of 74 – about three trees per plot. Deviance decreased with the use of random effects for all diameter classes (Table 4). Empirical estimation underestimates the 3.0–5.0 cm and the 5.0–7.0 cm d.b.h. classes and slightly overestimates the 1.0–3.0 and 7.0–9.0 cm d.b.h. classes (Table 4). The goodness-of-fit measures associated with the empirical approach (model 2) were consistently better than those from the full Weibull (model 1, Table 2).

Combined estimation (model 3)

The combined prediction method (model 3) with fixed effects and all plots results in an average error of 629 or ~26 trees in the 1.0–3.0 cm d.b.h. class (Table 4). The 7.0–9.0 cm d.b.h. class has an average error of just under 200 or about eight trees. A slight increase in average error up to 656 (27 trees) was seen using fixed and random effects in the 1–3 cm d.b.h. class, while the 7–9 cm d.b.h. class saw a decrease to 170 or about seven trees per plot (Table 4). The average error for the 3.0–5.0 and 5.0–7.0 cm d.b.h. classes also decreased using the random effects model (Table 4). Deviance decreased as diameter class increased for both fixed effects alone and the random effects model (Table 4).

In both the fixed effect and random effect models bias indicates that the 7.0–9.0 cm d.b.h. class is over-predicted by about three trees per plot (Table 4). The fixed effect model over-predicts the 3.0–5.0 cm d.b.h. class by one tree per plot, but all other diameter classes are under-predicted for both models (Table 4). The 1.0–3.0 cm d.b.h. class has the largest under-prediction, ~270 or 11 trees when all plots are included (Table 4) and 15 trees per plot for positive random effect plots only (Table 4). Overall, the goodness-of-fit measures for the combined approach (model 3) generally fell between the full Weibull (model 1) and the empirical approaches (model 2).

Validation

Empirical model predictions (model 2) calculated using the second permanent sample plot measurements show lower average errors for all diameter classes when random effects were included (Table 5). With fixed effects only, the 1.0–3.0 and 3.0–5.0 cm d.b.h. classes were over-predicted and including random effects results in under-prediction in all diameter classes (Table 5). In general, the goodness-of-fit measures for the second measurement were comparable to those of the first measurement. The alternative approach to validation by splitting the data is not as powerful for fitting equations, but this approach does provide a more independent set of data for validation. Using a 75 per cent subsample for fitting and a 25 per cent subsample for validation, errors based on predictions without random effects were similar to second measurement errors with no random effects (rMSE for 25 per cent validation sample by diameter class: 1–3 cm: 500 trees ha^{-1}, 3–5 cm: 244 trees ha^{-1}, 5–7 cm: 232 trees ha^{-1}, 7–9 cm: 132 trees ha^{-1}).

Discussion

Predicting the distribution of small trees using the distribution of large trees worked very well in this study. In their study, which utilized a total of 86 plots in managed and natural stands of Norway spruce (*Picea abies* L. Karst), Scots pine (*Pinus sylvestris* L.) and some silver fir (*Abies alba* Mill.), Merganič and Sterba (2006) reported that the parameters determined using truncated data and a truncated two-parameter Weibull were relatively stable in comparison to parameters determined using a complete dataset and a two-parameter Weibull. The stability of the parameter estimates between a truncated and complete diameter distribution was the basis of this analysis. Good or very good truncated Weibull fits were calculated for 92 per cent of plots (Table 1). When the truncated Weibull distribution was extended to include all small tree diameter classes, only 20 per cent of plots had a good or very good fit over the full range of small tree diameter classes (Table 1); however, the number of plots with good or very good fits increased as the range of diameters decreased (Table 1).

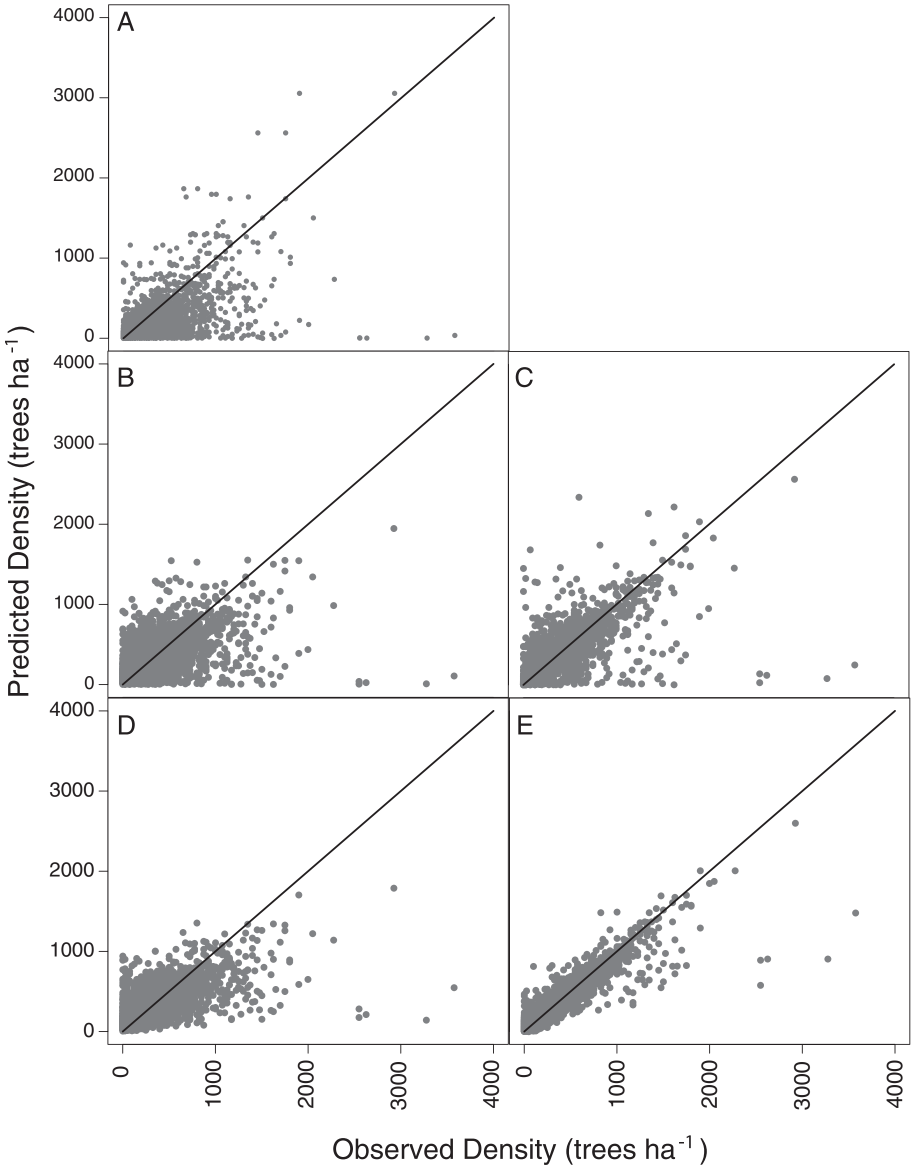

Figure 3. Small tree density predictions for the fitted model *vs* observed densities for each prediction method [A: full Weibull (model 1), B: combination/fixed (model 3), C: combination/random (model 3), D: empirical/fixed (model 2), E: empirical/random (model 2)].

Table 5: Error, bias, and deviance for predicted small tree densities with and without random effects using second plot measurements for the empirical approach (model 2)

Diameter class (cm)	Predictions with random effects			Predictions without random effects		
	Error	Bias	Deviance	Error	Bias	Deviance
1.0–3.0	322	23	204	505	−36	356
3.0–5.0	102	7	70	290	−50	217
5.0–7.0	136	8	89	219	8	151
7.0–9.0	90	25	63	146	26	104

The full Weibull distribution method (model 1) consistently underestimated the number of small trees in all diameter classes (Table 2). The full Weibull approach (model 1) does have the benefit of not requiring a large number of plots to fit an empirical model and could thus be applied to small samples or individual plots.

Reducing the truncation point to 7 cm from 9 cm for the full Weibull method (model 1) showed improvements in all goodness-of-fit measures for all diameter classes (Table 2). These improvements indicate that supplying extra density to the Weibull, probability density function increases the predictive ability of the method. Another strategy might be to fit the observed truncated basal area distribution to the area-biased Weibull (Gove 2003a, b). Intuitively, because the truncated small trees contribute relatively little basal area, the fit of the basal area distribution might be less sensitive to truncation. Initial tests using the basal area distribution did not lead to marked improvements using our data, but we speculate that this approach is worth further exploration, especially if horizontal point samples are used rather than fixed area plots, as in this study.

Improvements were obtained using the combined (model 3) and empirical (model 2) approaches, which used densities predicted from the full Weibull (model 1), climate-based site index and quadratic mean diameter to predict small tree densities. Because the combined approach (model 3) is limited by the shape of the Weibull in the small diameter classes, the improvements in goodness-of-fit measures are not as great as those seen in the empirical approach (model 2).

Given the better distribution fit seen in the larger small tree diameter classes (Table 1), it is not surprising that the larger small tree diameter classes (i.e. 7.0–9.0 cm) also have more accurate and precise density predictions in comparison to the smaller diameter classes (Tables 2 and 4). Average error, bias and deviance decreased as diameter class size increased. In general, the results indicate that the smaller d.b.h. classes are less well predicted by the larger trees. The poor fit in the smallest diameter classes may be a result of several factors. In stands where mortality has begun and stand break up is occurring, the distribution of diameters above the truncation point is skewed to the right and often is discontinuous. Given the skewed distribution of large trees, very few small trees are predicted by the Weibull distribution. An abundance of small trees can also be a result of minor stand disturbances that lead to small clumps of shade-tolerant species and localized understorey reinitiation or clumps of shade-tolerant species can be left over from patches of favourable seedbed during the original stand initiation. Either of these situations would lead to an abundance of small trees that may not be connected to the current large tree distribution and results in situations where predicting small tree abundances based only on current large tree abundance and size distribution, without information on stand history, is difficult. Since the analyses presented here are based on the initial permanent sample plot (PSP) measurements, no disturbance history was available from the individual tree records; however, when the PSPs were established, they were generally placed in undisturbed locations in the selected stands with the overstorey canopy intact (Dunlap, 1988). Even with no history of disturbance, stands in this area are composed of different number of cohorts and can be on different successional trajectories.

The empirical approach (model 2, equation 7) consistently had better measures of goodness of fit across all diameter classes than the other tested alternatives. The inclusion of random effects into the predictions significantly improved the fit using the empirical model (Table 4). In less than 6 per cent of plots, negative predictions occurred. One option for dealing with these negative predictions is to set negative predictions equal to zero. Alternatively, in the case of negative predictions from the empirical model, one of the other two methods of small tree prediction is recommended.

While not truly an independent dataset, the second plot measurement dataset shows that the method is extendable over time (Table 5). The similarities in goodness-of-fit measures between the second plot measurement dataset and the split dataset (Table 5) are expected as validation by data splitting has been shown to provide little additional information in evaluating regression models (Kozak and Kozak, 2003).

The empirical model (model 2, equation 7) developed for small tree prediction in this analysis uses stand-level variables commonly used when modelling ingrowth. Shifley *et al.* (1993) report both stand density and size structure as common variables used in ingrowth models. In their paper, crown competition factor was used to set a maximum level of ingrowth possible for a given diameter threshold (Shifley *et al.*, 1993). A two-stage ingrowth model produced by Adame *et al.* (2010) used quadratic mean diameter and average height to model occurrence of ingrowth, while the amount of ingrowth was modelled using total stand density and average diameter. Lexerod and Eid (2005) modelled recruitment in mixed species stands in Norway using a two-stage model that included site index, age, total stand density and species composition. The empirical model used in this analysis uses small tree density predicted by the Weibull distribution, quadratic mean diameter and climate-based site index as variables. Given tree density, quadratic mean diameter and site index have all previously been influential in modelling ingrowth, a possible application of our small tree prediction model is the prediction of ingrowth. An example of applying an initial density distribution to predict ingrowth is given by Lundqvist (1995) and Lundqvist and Nilson (2007). While these studies used height distributions, the same process should be applicable to modelling small tree dynamics and ingrowth prediction based on d.b.h..

Conclusions

Model 2, the empirical approach, performed better than the alternative models. The empirical approach is however limited to inventories where at least one measurement includes small tree diameters and the best fits are possible only when plot-level random effects are known or estimated. The model as shown is applicable to the Acadian forest region where the climate-based site index is available. The full Weibull approach (model 1) is not as limited as the empirical, and it can be applied when no measurement of small trees is available and is not limited geographically, however, increased prediction errors result.

Future work should focus on using these small tree predictions to predict ingrowth above the threshold diameters. Additionally, this study only attempted to predict total density and not species-specific densities. A two-stage approach using these predictions of total density and species probabilities might be used to predict the species composition and size distribution of the small tree component.

Funding

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, Collaborative Research Development Grant (349436-06).

Conflict of interest statement

None declared.

References

Adame, P., del Rìo, M. and Cañellas, I. 2010 Ingrowth model for pyrenean oak stands in north-western Spain using continuous forest inventory data. *Eur. J. For. Res.* **129**, 669–678.

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

Breiman, L. 2001 Random forests. *Machine Learning*. **45**, 5–32.

Curtis, R.O. and Marshall, D. 2005 *Permanent-Plot Procedures for Silvicultural and Yield Research*. United States Department of Agricultural. Forest Service. Pacific Northwest Research Station, Portland, OR. General Technical Report PNW-GTR-634: 92.

Dunlap, J.A. 1988 *1987 Permanent Sampling Program Annual Report*. New Brunswick Department of Natural Resources and Energy, Timber Management Branch, Fredericton, New Brunswick, 27 p.

Garcia, O. 1988 Growth modelling – a (Re)view. *N. Z. J. For.* **33**, 14–17.

Gove, J.H. 2003a Estimation and applications of size-biased distributions in forestry. In *Modeling Forest Systems*. A. Amaro, D. Reed and P. Soares (eds). CABI Publishing, Wallingford, UK, pp. 201–212.

Gove, J.H. 2003b Moment and maximum likelihood estimators for Weibull distributions under length- and area-biased sampling. *Environ. Ecol. Stat.* **10**, 455–467.

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

Kershaw, J.A., Jr. 2011. Weibull Library in R. http://ifmlab.for.unb.ca/people/kershaw/index.php/r-applications/ (Accessed July 18, 2011).

Kozak, A. and Kozak, R. 2003 Does cross validation provide additional information in the evaluation of regression models? *Can. J. For. Res.* **33**, 976–987.

Lexerod, N. and Eid, T. 2005 Recruitment models for Norway spruce, Scots pine, birch and other broadleaves in young growth forests in Norway. *Silva Fennica*. **39**, 391–406.

Little, S.N. 1983 Weibull diameter distributions for mixed stands of western conifers. *Can. J. For. Res.* **13**, 85–88.

Lundqvist, L. 1995 Simulation of sapling population dynamics in uneven-aged *Picea abies* forests. *Ann. Bot.* **76**, 371–380.

Lundqvist, L. and Nilson, K. 2007 Regeneration dynamics in an uneven-aged virgin Norway spruce forest in northern Sweden. *Scand. J. For. Res.* **22**, 304–309.

Merganič, J. and Sterba, H. 2006 Characterisation of diameter distribution using the Weibull function: method of moments. *Eur. J. For. Res.* **125**, 427–439.

Robinson, A. 2004 Preserving correlation while modelling diameter distributions. *Can. J. For. Res.* **34**, 221–232.

Shifley, S.R., Ek, A.R. and Burk, T.E. 1993 A generalized methodology for estimating forest ingrowth at multiple threshold diameters. *For. Sci.* **39**, 776–798.

US Forest Rocky Mountain Research Station Moscow Laboratory. 2010 *Research on Forest Climate Change: Potential Effects of Global Warming on Forests and Plant Climate Relationships in Western North America and Mexico*. http://forest.moscowfsl.wsu.edu/climate/ (accessed on November 16, 2010).

Weiskittel, A.R., Crookston, N. and Radtke, P. Linking climate, gross primary productivity, and site index across forests of the western United States. *Can. J. For. Res.* (in press).

Weiskittel, A.R., Wagner, R.G. and Seymour, R.S. 2010 *Refinement of the Forest Vegetation Simulator Northeastern Variant Growth and Yield Model: Phase 2. Cooperative Forestry Research Unit Annual Report*. University of Maine, Orono, ME.

Zar, J.H. 2009 *Biostatistical Analysis*. 5th edn. Prentice Hall, Upper Saddle River, NJ.

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

Zutter, B.R., Oderwald, R.G., Murphy, P.A. and Farrar, R.M.J. 1986 Characterizing diameter distributions with modified data types and forms of the Weibull distribution. *For. Sci.* **32**, 37–48.

Received 9 December 2010