Forest Ecology and Management 400 (2017) 332–338

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Predicting tree-seedling height distributions using subcontinental-scale forest inventory data

James A. Westfall *, William H. McWilliams

U.S. Forest Service, Northern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073, USA

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 6 June 2017
Accepted 10 June 2017
Available online 20 June 2017

Keywords:
Tree regeneration
Weibull distribution
Stand composition
Forest management

ABSTRACT

The importance of tree seedlings in determining future stand composition and structure is well-documented in forestry literature. When planned or unanticipated overstory removal events occur, subsequent regeneration success is often linked to the number of seedlings and their height distribution. Yet, in most forest inventories, only counts of seedlings are obtained as it is too time-consuming to measure individual seedlings. To better understand the expected height distribution, models were developed to predict Weibull distribution parameters based on seedling abundance information and stand/site characteristics. A number of these characteristics were found to be statistically significant predictors of the distribution parameters; however, a more parsimonious model using stand basal area, stand age, number of seedlings, and latitude provided essentially the same fit statistics. Models were fitted for all species and for selected species subgroups, but there was generally insufficient data at this time to develop species-level analyses.

Published by Elsevier B.V.

1. Introduction

Planning for post-harvest natural regeneration success is a key component of pre-harvest planning and assessment. One of the primary indicators of post-harvest regeneration outcomes is the species, number, and size of seedlings present prior to harvest (Loftis, 1990; Marquis et al., 1992). Also to be considered is whether seedlings germinate from seed or sprout from stumps or roots (Decocq et al., 2004; Del Tredici, 2001). In the early stages of stand initiation and development, both sources contribute to the seedling population that will largely determine the composition of the future stand. Under typical stand growth trajectories and increased canopy closure, the number, size, and species of seedlings at any given time depends on numerous factors, but primarily by amount of seed production (Zaczek, 2002; Standovár and Kenderes, 2003), available light in the understory (Lieffers et al., 1999; Stancioiu and O'Hara, 2006), herbivory (Brose et al., 2008), and more generally by climate (Rochefort et al., 1994; Bazzaz et al., 1990). It is important for forest managers to track the seedling component in all phases of stand development. In young stands, seedlings are indicators of future stand composition and structure. These are key factors in projections of stand development and are used for planning timing/intensity of silvicultural activities, such as prescribed fire and thinning (Albrecht and McCarthy, 2006). The seedling component of older stands is also important as either planned or unplanned stand-replacement disturbances may occur (Swanson et al., 2011).

Quantifying the seedling component requires consideration of seedling vigor and height as these often indicate the likelihood of survival and resultant future canopy characteristics. Root-collar diameter (rcd) is used evaluate vigor. Often, seedlings are only measured if they are considered 'established' based on a minimum rcd threshold. Also, rcd thresholds for large-seeded taxa are used to classify seedlings as competitive and indicate the probability of developmental success (Brose, 2008). Seedling height is useful as an indicator of freedom from competition with other tree reproduction, understory vegetation (e.g., ferns and grasses; McWilliams et al., 1995), and ungulate browsing of the upper stem (Horsley et al., 2003; Jobidon et al., 2003). Seedling heights are largely driven by light availability and age, which is corroborated in modeling strategies that use age and site index as covariates (Puhlick et al., 2013); consider canopy openness as a surrogate for light availability and age since harvest (Millington et al., 2011); and employ various overstory density measures and understory diffuse light measurements to develop seedling height models (Lochhead and Comeau, 2012).

To facilitate research on seedling dynamics and inform forest managers on the condition and health of this vital component,

* Corresponding author.
E-mail addresses: jameswestfall@fs.fed.us (J.A. Westfall), wmcwilliams@fs.fed.us (W.H. McWilliams).

http://dx.doi.org/10.1016/j.foreco.2017.06.025
0378-1127/Published by Elsevier B.V.

the Forest Inventory and Analysis program of the U.S. Forest Service, Northern Research Station (NRS-FIA) recently implemented Regeneration Indicator (RI) measurements on a subset of inventory plots. These data are critically important because the region's forests are aging and face numerous, inter-related stressors that challenge forest regeneration managers (McWilliams et al., 2015). Of particular importance to this study is the addition of height class to the seedling data collection protocol. This affords the opportunity to develop relationships between tree seedling height distributions and typical forest inventory variables to help foresters understand factors affecting seedling size dynamics across a range of stand and site conditions. To this end, specific objectives were to (1) use seedling height class information to generate a continuous seedling height distribution, (2) model the tree seedling height distribution where the parameters may be a function of stand and location attributes, (3) relate the modeling outcomes to stand development patterns in the context of expected biological relationships, and (4) describe how this knowledge assists in making informed forest management decisions.

2. Methods

2.1. Data

The data used for this study were collected by NRS-FIA from 2012 to 2015 across the states of Pennsylvania, New Jersey, New York, Massachusetts, Vermont, New Hampshire, and Maine. The FIA Phase 2 (P2) quasi-systematic sample has an intensity of approximately 1 plot per 2428 ha. Each sample plot consists of four 7.32 m radius subplots, and within each subplot is a 2.07 m radius microplot (Bechtold and Scott, 2005). All trees with a diameter breast height (dbh) of 12.70 cm and larger are measured on the subplot, while information on saplings (2.54 cm $\leq$ dbh < 12.70 cm) and seedlings (dbh < 2.54 cm and height > 0.05 m) are collected on the microplot. Additional data were collected on a 1/8 subset of the FIA P2 sample (hereafter denoted as P2+ dataset) chosen for additional measurements associated with regeneration attributes (McWilliams et al., 2015). Seedling counts by species and height class were taken on these plots, where the height classes corresponded to (1) 0.05–0.15 m, (2) 0.16–0.30 m, (3) 0.31–0.90 m, (4) 0.91–1.51 m, and (5) 1.52–3.05 m, and (6) >3.05 m.

Within each plot, areas having different forest conditions are mapped and data associated with each of the distinct areas are collected. Specifically, different areas within the plot are delineated when there are differences in reserved status, owner group, forest type, stand size class, regeneration status, or tree density (U.S. Forest Service, 2013). For this study, the data were summarized at the condition-level as the aforementioned attributes and other factors may influence the seedling component. Additional data collected at the condition level and relevant to this study include basal area per hectare (dbh $\geq$ 2.54 cm), stand age, site productivity, slope, aspect, and physiographic class (Woudenberg et al., 2010). To include recently harvested conditions and other new forests, stand age of 0 was set at 0.5. Plot-level data used included latitude, longitude, and elevation. Data from 2012 to 2014 were used for analysis; with the 2015 data serving as validation data. Table 1 provides summary statistics for various data attributes.

2.2. Analysis

Actual seedling heights are a continuous variable, but in these data the seedlings are counted by height classes to identify early developmental traits by strata. Nonetheless, the overall pattern can be deduced by using the height-class midpoints. As expected, this trend shows rapid decreases in seedling density as size increases, with more seedlings near the lower threshold than near the upper threshold within a height class. To provide a basis for creation of a continuous seedling height distribution, models that describe the height trend were sought. Initial analyses consisted of evaluating several distributional forms (e.g., exponential, gamma, Weibull, beta), which indicated a 3-parameter Weibull function provided the most flexibility. This distribution has a cumulative distribution function given by Teimouri and Gupta (2013):

$$F(y) = 1 - \exp\left(-\frac{y-\mu}{\beta}\right)^{\alpha} \tag{1}$$

where μ (location), α (shape), and β (scale) are parameters.As the distribution of seedling heights may be influenced by various factors associated with the local environment, relationships between distribution parameters and forest type, site productivity, stand basal area, stand age, numbers of seedlings, slope, aspect, physiographic class, elevation, latitude, and longitude were assessed. This was accomplished by creating categories for continuous variables and then fitting (1) to each category within each variable of interest. Graphical analyses of the relationship between the categories and the distribution parameters revealed whether relationships were present and if so, their likely form, e.g., linear or nonlinear. While these analyses were useful for examining potential underlying relationships, the information does not imply a final form of the model because correlations among the environmental variables, as well as correlations among model parameters, were not accounted for. Still, the basic form of the model considered has a fixed value for μ (the smallest height in the data = 0.0508 m) with the scale and shape parameters being functions of certain environment variables (E_g), i.e., $\beta = f(E_1, E_2, \ldots, E_g)$ and $\alpha = f(E_{g+1}, E_{g+2}, \ldots)$.

Goodness-of-fit statistics for the candidate models were assessed via the root mean squared error (RMSE) and the concordance correlation (Rc; Vonesh et al., 1996):

$$\mathrm{RMSE} = \sqrt{\frac{\sum (y-\hat{y})^2}{n}} \tag{2}$$

$$\mathrm{Rc} = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2 + \sum(\hat{y}-\bar{\hat{y}})(y-\bar{\hat{y}}) + n(\bar{y}-\bar{\hat{y}})} \tag{3}$$

where $\hat{y}$ is the model prediction, $\bar{\hat{y}}$ is the mean model prediction, y is the observed value,$\bar{y}$ is the mean observed value, and n is the number of observations. The Rc statistic spans the interval between −1 and +1, with r_c = 1 indicating a perfect fit to the data.

For each candidate model, pseudo-heights were predicted for each seedling based on its observed height class. A complication was the lack of an upper threshold for the largest height class. To determine an approximate upper-limit for seedling heights, the distribution of sapling (2.54 cm $\leq$ dbh < 12.70 cm) heights was examined. The distribution of sapling heights is conditional on the sapling having attained a minimum dbh of 2.54 cm. Based on a visual inspection of the data, a minimum height for saplings of 4.58 m was established; therefore the maximum seedling height was assumed to be 4.57 m. Due to the right-tailed nature of the Weibull distribution, there were some cases where the predicted heights exceeded the 4.57 m threshold. When this occurred, a new Weibull random variate was selected until the predicted height was $\leq$4.57 m. The performance of the Weibull distribution implementation of pseudo-heights was compared to the height-class means distribution through visual comparisons of histograms for practical differences. Statistical comparisons between the original distribution and pseudo-heights aggregated back to height class means were performed using the Kolmogorov-Smirnov (K-S) test for equality of distributions (Conover, 1999). While the

Table 1
P2+ data summary statistics for 1153 conditions by major forest type.

Forest type	n	Attribute	Min.	Mean	Max.	Std. dev.
Deciduous	895	Basal area (m^2/ha)	0.0	25.9	96.8	13.6
		Stand age (yrs)	3	67	142	24
		Seedlings (count/ha)	339	4805	73,673	3922
		Slope (%)	0	15	92	15
		Latitude (deg.)	39.30	42.47	47.40	1.84
		Longitude (deg.)	−80.48	−75.12	−67.39	3.48
		Elevation (m)	0	409	1011	169
Conifer	258	Basal area (m^2/ha)	0.0	29.7	144.3	18.9
		Stand age (yrs)	1	60	162	34
		Seedlings (count/ha)	371	4941	20,748	3763
		Slope (%)	0	10	80	13
		Latitude	39.53	44.16	47.29	1.90
		Longitude	−79.96	−71.77	−67.10	3.40
		Elevation (m)	30	379	963	194

K-S test provides useful information, it should be noted that test results are based on only six observations and thus the power to detect differences is low.

To evaluate the performance of the analytical results, pseudo-heights for each seedling in the 2015 validation data were generated as described above using the height class information and the final form of model (1). As with the model fit data, results were compared graphically and through K-S tests to assess the model performance as applied to independent data.

While the above analyses pertain to all species, there are likely various reasons why the height distribution for certain species or species groups might be of interest. Such information may be beneficial for determining the likelihood of certain desirable species to replace previous canopy dominants for stands regenerating naturally following stand-initiation disturbance. As an example, all seedling species were assigned to severe (22 species) or low/moderate (68 species) browse intensity classes based on species palatability to white-tailed deer. These ratings are somewhat subjective, but for this study the assignments were based primarily on Benner (2007) and Petrides (1941); although in some cases other sources were consulted. Model (1) was fitted separately to each subgroup to evaluate how seedling height distributions might be affected by selected browsing.

3. Results

After considering and testing various formulations of model (1), including several interactions among predictor variables, the best model fit statistics were obtained when the scale parameter was a function of stand age and physiographic class. The shape parameter was influenced by number of seedlings per ha, basal area per ha, latitude, terrain slope, and broad forest type classification. The final specification was determined to be:

$$F(y) = 1 - \exp\left(-\frac{y - 0.0508}{\beta_0 A^{\beta_1} + \beta_2 P_X + \beta_3 P_M}\right)^{\alpha} + e \quad (4)$$

$$\alpha = \alpha_0 + \alpha_1 S + \alpha_2 B + \alpha_3 L + \alpha_4 T + \alpha_5 F$$

where A = stand age (yrs); $P_X = 1$ if physiographic class is xeric, 0 otherwise; $P_M = 1$ if physiographic class is mesic, 0 otherwise; S = number of seedlings ha^{-1}; B = basal area ($m^2\ ha^{-1}$, dbh $\geq$ 2.54 cm); L = latitude (degrees); T = terrain slope (percent); F = 1 if the forest type is conifer dominant, 0 otherwise; β_0-β_3 and α_0-α_5 are estimated parameters; *e* is random residual error.

Further analyses indicated a more parsimonious formulation provided essentially the same fit statistics:

$$F(y) = 1 - \exp\left(-\frac{y - 0.0508}{\beta_0 A^{\beta_1}}\right)^{\alpha_0 + \alpha_1 S + \alpha_2 B + \alpha_3 L} + e \quad (5)$$

All estimated parameters for models (4) and (5) were statistically different from zero at the 95% confidence level (Table 2). Both models had concordance correlation (Rc) values near 0.73 and root mean squared error (RMSE) of approximately 0.21. In the context presented here, the units for RMSE are distribution percentile points where, for example, the 50th percentile would be expressed as 0.50.

Predicted values from (5) were used to generate a continuous height distribution having similar shape to that inferred by the proportion of data in each height class and the height-class means. Fig. 1a depicts the original and predicted distributions, where the original data are now distributed in a monotonically-decreasing nonlinear trend within and among height classes. To better illustrate model performance, Fig. 1b shows the predictions aggregated back to the height class means; where it is shown that slight over- or under-prediction is present for each height class, although the discrepancies are relatively minor. Descriptive statistics for each distribution suggest application of model (5) produces slightly smaller measures of central tendency and variation in comparison to the original data (Table 3). Potential users of the model results should consider whether these differences are of practical importance to the intended application. Results of the K-S test based on these comparative distributions produced a D statistic of 0.33 ($p = 0.89$), suggesting there are no statistically-significant differences between the two distributions at the 95% confidence level ($\alpha = 0.05$).

Application of model (5) to the 2015 validation data showed close agreement between the original and pseudo-height distributions (Fig. 2). Comparisons between descriptive statistics and the K-S test results were essentially the same as those reported above, suggesting the model performance is similar applied to an independent data set.

Model (5) was also fitted to species subdomains based on likely browse intensity as indicated by deer palatability, i.e., severe or low-moderate. The results for the low-moderate browse group were generally similar to the all species group, with all estimated parameters statistically significant and model fit statistics essentially unchanged (Table 4). The model fit to species susceptible to severe browsing also included all estimated parameters being statistically significant; albeit with larger standard errors due to the smaller sample size. The severe group also exhibited a degradation in fit statistics, with the model now having an Rc statistic of nearly 0.43 and RMSE of 0.27. These results indicate that severe browsing produces substantial increases in variability of seedling height dis-

Table 2
Estimated parameters and fit statistics for models (4) and (5).

Parameter	Model (4)		Model (5)	
	Estimate	Std. err.	Estimate	Std. err.
β_0	2.16281	0.14710	2.07548	0.17000
β_1	−0.22697	0.02010	−0.28579	0.01990
β_2	−0.24593	0.06380	–	–
β_3	−0.21209	0.05060	–	–
α_0	0.90320	0.12630	1.13897	0.11350
α_1	4.12E−05	2.05E−06	4.22E−05	2.05E−06
α_2	−0.00138	0.00035	−0.00132	0.00036
α_3	−0.01203	0.00299	−0.01753	0.00268
α_4	0.00113	0.00038	–	–
α_5	−0.03578	0.01300	–	–
Rc	0.7286		0.7276	
RMSE	0.2066		0.2069	

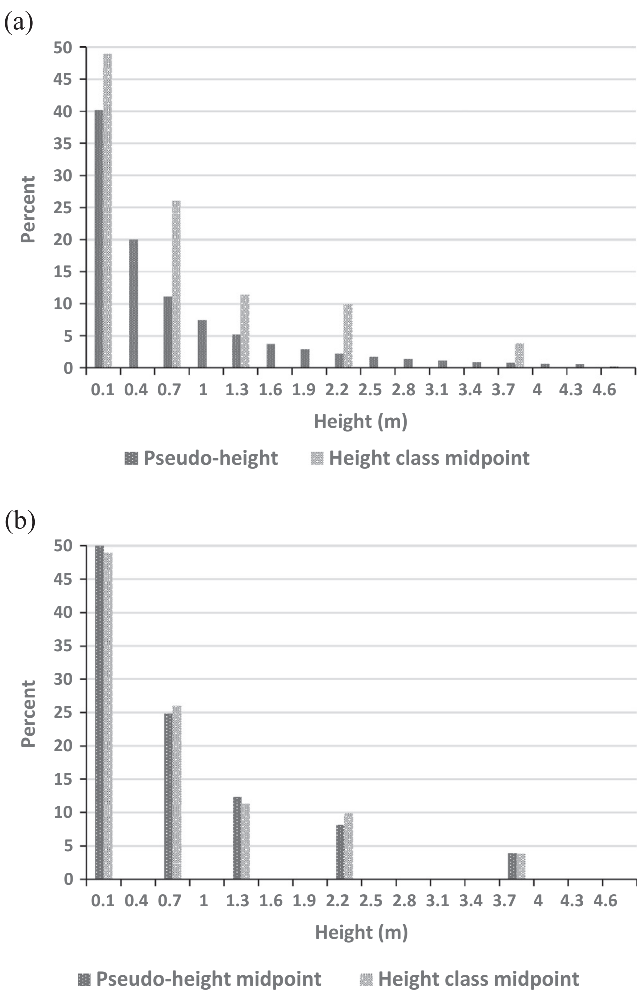

Fig. 1. Observed height-class midpoints distribution compared to (a) the continuous distribution of seedling heights from model (5), and (b) the continuous height distribution reduced to height-class midpoints.

Table 3
Statistics for original height-class midpoint distribution and pseudo-height midpoint distribution arising from model (5).

Statistic	Original	Predicted
Mean	1.38	1.31
Std. dev.	1.44	1.39
Median	0.91	0.87
IQR	2.06	1.90

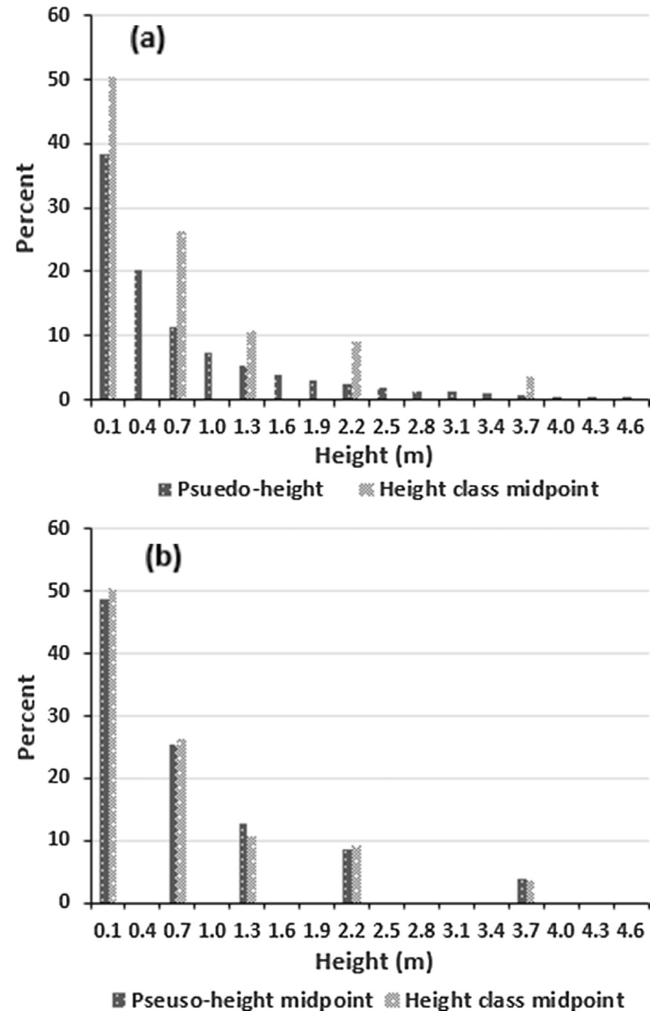

Fig. 2. Validation data: Observed height-class midpoints distribution compared to (a) the continuous distribution of seedling heights from model (5), and (b) the continuous height distribution reduced to height-class midpoints.

tributions. Comparisons between the predicted height distributions of the two groups suggested that the severely-browsed species tend to have more seedlings in the very small height range, e.g., less than 0.4 m; whereas there appears to be more low-moderate species persisting in the height classes ranging from approximately 0.7 to 2.2 m in height (Fig. 3). Thereafter, the distributions are quite similar, which is not surprising as seedlings having height of 2.5+ m have likely escaped browsing of the terminal leader.

4. Discussion

As the reduced model (5) provided nearly identical fit statistics to those from the full model (4), the reduced model should be preferred by most users as the required predictor variables are commonly available in most forest inventory data, e.g., the entire NRS-FIA P2 sample. The resultant parameter estimates show that as stand age increases, the distribution shifts towards having a larger proportion of smaller seedlings. The shape parameter increases as number of seedlings increases, and decreases with larger values for basal area and latitude. Fig. 4 depicts seedling height distributions at various stand ages based on an approximate average stand trajectory developed from the data. Typical stand dynamics suggest that higher basal area would be associated with greater crown

Table 4
Estimated parameters and fit statistics for model (5) fitted to species subdomains categorized as severe and low-moderate browse intensity.

Parameter	Severe		Low-Moderate	
	Estimate	Std. err.	Estimate	Std. err.
β_0	2.25453	0.4519	1.95076	0.1701
β_1	−0.31041	0.0482	−0.26505	0.0212
α_0	1.50285	0.1985	0.73206	0.1119
α_1	2.60E−05	3.12E−06	4.18E−05	2.03E−06
α_2	−0.00277	0.0006	−0.00111	0.0004
α_3	−0.02654	0.0048	−0.00865	0.0027
Rc	0.4262		0.7132	
RMSE	0.2702		0.2112	

closure and decreased light availability for seedling survival and growth (Vickers et al., 2017). Lower basal area would permit more light in the understory, and seedlings receiving the additional light through canopy gaps will tend to grow faster and thus create more differentiation in heights across the seedling population (Coates, 2002; York et al., 2003). Both relatively fewer seedlings and higher basal area conspire to produce smaller values of the shape parameter, which favor a higher proportion of seedlings with small heights. Other influences on number of seedlings include deer herbivory and competing vegetation such as ferns and grasses (Horsley and Marquis, 1983). High levels of either factor would suggest fewer, shorter seedlings, corresponding with the resultant smaller shape parameter due to a lower seedling density. Typical stand development suggests the number of seedlings and their height variability is lessened as the stand matures.

The models should be used with caution when stand growth trajectories have been altered by disturbance or treatment. At a minimum, enough time should be allowed for the stand response to have occurred, e.g., increased light availability due to thinning may affect number of seedlings. As an example, Fig. 5 depicts a scenario where 40% of the basal area was removed at age 70 and subsequent post-thinning seedling distributions at age 90 and 120. The resultant decrease in overstory basal area was presumed to create $S = 10{,}000$ and $B = 30$ as the 20 year post-thin response. The seedling distribution is shown to have shifted towards a broader range of heights than were present prior to thinning. Fur-

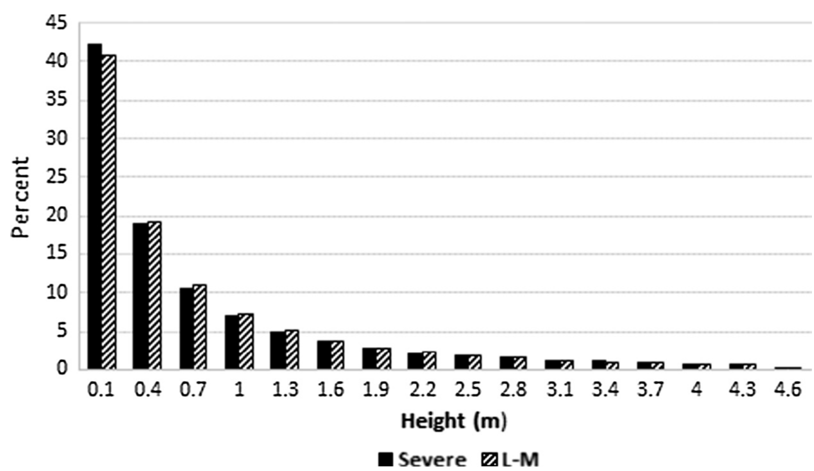

Fig. 3. Predicted seedling height distributions for severe and low-moderate browse intensity species subdomains.

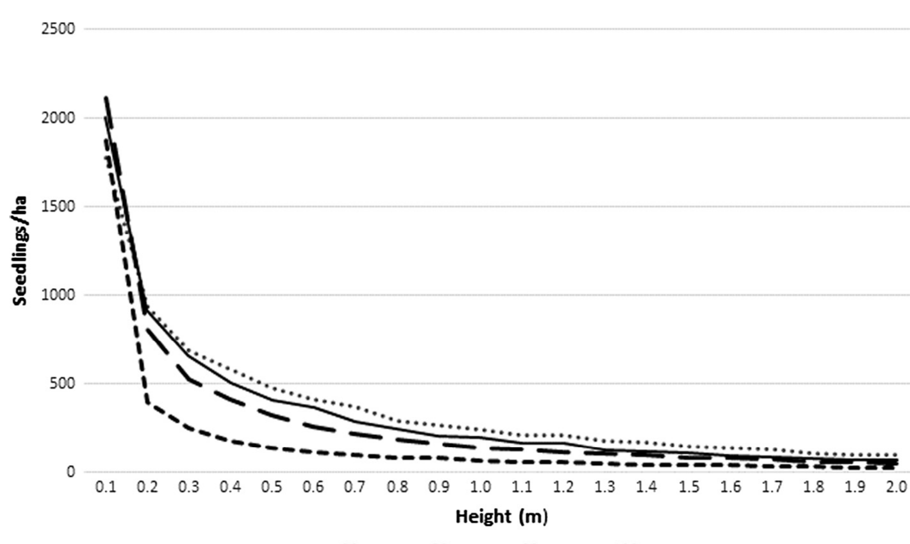

Fig. 4. Seedling height frequency distributions 0.05–2.0 m for various ages in a typical stand development pattern. Scenarios include (1) $A = 20$ years, $B = 2\ m^2$/ha, $S = 8500$/ha; (2) $A = 50$ years, $B = 10\ m^2$/ha, $S = 7500$/ha; (3) $A = 80$ years, $B = 40\ m^2$/ha, $S = 6500$/ha; (4) $A = 120$ years, $B = 100\ m^2$/ha, $S = 4000$/ha. $L = 42$ degrees for all scenarios.

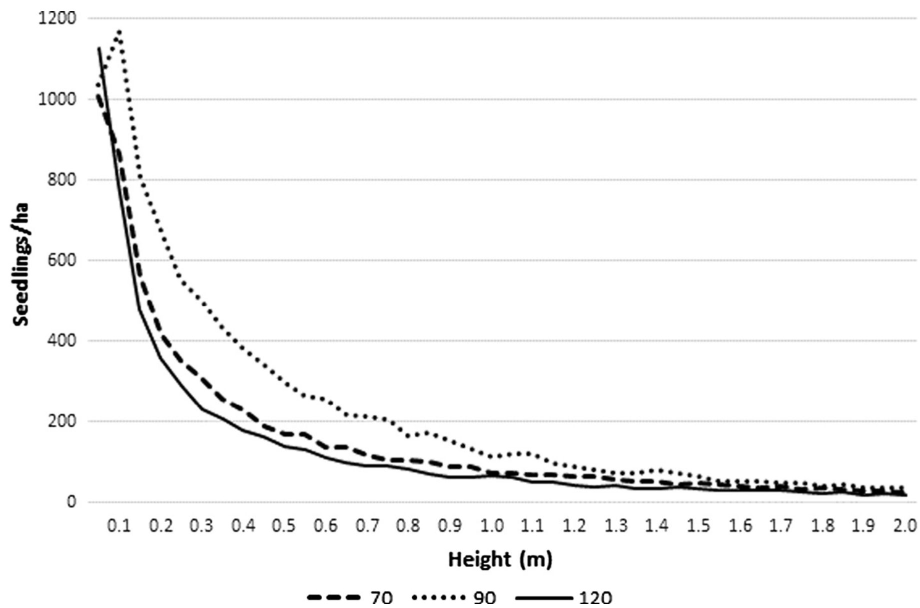

Fig. 5. Seedling height frequency distributions 0.05–2.0 m for an example thinned stand with 40% basal area removed. Stand conditions just prior to thinning were A = 70 years, B = 35 m^2/ha, S = 7000/ha. 20 years of post-thinning growth produced conditions of A = 90 years, B = 30 m^2/ha, S = 10,000/ha; whereas conditions 50 years post-thinning were A = 120 years, B = 50 m^2/ha, S = 6000/ha; L = 42 degrees for all scenarios.

Table 5
Estimates of seedlings/ha, standard errors based on the current sample, and standard errors revised to reflect the use of the entire P2 sample via application of model (5) for Maryland, U.S. in 2015.

	Height class						Total
	1	2	3	4	5	6	
Seedlings/ha	9102.3	2517.6	1796.0	307.8	153.5	68.9	13946.2
Std error	4859.7	667.9	582.1	126.4	69.7	72.6	5033.2
Std error (revised)	1846.7	253.8	221.2	48.0	26.5	27.6	1912.6

ther growth to age 120 (S = 6000, B = 50) is indicative of a return to canopy-closure and the ensuing effects of decreased light availability in the understory (i.e., fewer, smaller seedlings).

Though some of the significant predictors in model (4) contributed little to improving model fit, the results may provide some insight into other possibly relevant factors that may affect seedling height distributions. Physiographic class, terrain position, slope, and forest type classification appeared to play relatively minor roles in determining seedling height distributions in this study. Alternatively, Vickers et al. (2011) used moisture availability classes to develop models for hardwood regeneration in the central Appalachian region of the U.S. It is noteworthy that physiographic class and forest type were included as indicator variables representing certain classifications. Other variables and classifications, i.e., a continuous moisture index (Iverson et al., 1997) to describe physiographic conditions, may provide more useful information and have a greater impact on the model fit. Some variables may also prove more useful for other locations as well, e.g., slope may become increasingly important in mountainous areas. Evaluation of these hypothesized outcomes are topics for further exploration.

Prediction of seedling height distributions for selected species subdomains was accomplished by refitting the model to reduced datasets based on browse palatability rankings. This would be the preferred method when possible. Otherwise, predictions for browse classifications would rely on applying the model fitted to all species and then reducing the data to only the species of interest. The degree to which this can be accurately accomplished depends on how closely the distribution for the selected species follows the overall distribution. It would be expected that more accurate results would be obtained as the number of species in the subdomain increases, e.g., the low-moderate browse group. Modeling distributions of individual species is the next logical step as the distribution for any given species may differ considerably from that of all seedlings due to factors such as available moisture, shade tolerance, and herbivory stress. Parameterization of species-level models should become more promising as more data are collected and sample sizes increase.

An obvious application of the results would be to use these models to estimate seedling height distributions for the other NRS-FIA plots with no P2+ measurements (88 percent of all plots). Theoretically, this would increase the sample size by 7× for estimates requiring seedling heights, e.g., number of seedlings by height class. Statistical formulae suggest a reduction of 62% in the standard error of the estimates would be achieved (Table 5). To be objective, the uncertainty due to the model should be included in the total error for the estimates. Generally this source of error has been shown to be small in comparison to sampling error (McRoberts and Westfall, 2014).

Another valuable use would be for growth and yield models where seedling information is included in stand projections or as part of regeneration establishment modules requiring predictions for initial conditions, e.g., the Northeast variant of the Forest Vegetation Simulator (FVS) (Dixon and Keyser, 2008). Currently, FVS options for regeneration establishment are planting, sprouting, and custom input. Modeled seedling height distributions could be appended to the entire NRS-FIA P2 dataset. This has been a limitation in applying the Northern variant for mixed deciduous forests of the study region.

5. Conclusion

Seedling populations can be highly variable due to the wide range of factors that affect survival and growth. The results of this study quantified the impacts of several stand attributes on expected seedling height distributions. Specifically, relationships between certain stand/site variables and distribution parameters provide insight that should help foresters better manage stands to achieve post-harvest regeneration goals. The results can be considered broadly applicable across the study area, even though different associations between predictor variables and distribution parameters may be encountered in other environments. Also, more refined relationships may be developed for the study area as these data continue to be collected on an annual basis by the NRS-FIA program. Empirical explanation of the seedling component remains an important area for further research with applications such as silvicultural prescriptions and inputs into growth and yield models for projections of future stand conditions. To this end, the methods present a modeling framework to potentially employ across a range of different applications (e.g., individual species or spatial extents).

As with all models, values used for the predictor variables should fall within a reasonable range of possible stand conditions, e.g., a stand of age 1 year would not have a basal area of 25 m^2/ha. This should generally not be a problem with observed data, with the exception of recent disturbance events that could substantially alter stand characteristics. In such cases, the number of seedlings expected as a response needs to be inferred or model usage should be delayed until the response can be observed. There are no specific guidelines in this regard, thus it is incumbent upon the user to be cautious in such situations.

Acknowledgements

The authors are grateful to anonymous reviewers for insightful suggestions that notably improved the draft manuscript.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Albrecht, M.A., McCarthy, B.C., 2006. Effects of prescribed fire and thinning on tree recruitment patterns in central hardwood forests. For. Ecol. Manage. 226, 88–103.

Bazzaz, F.A., Coleman, J.S., Morse, S.R., 1990. Growth responses of seven major co-occurring tree species of the northeastern United States to elevated CO_2. Can. J. For. Res. 20, 1479–1484.

Bechtold, W.A., Scott, C.T., 2005. The forest inventory and analysis plot design. In: Bechtold, W.A., Patterson, P.L. (Eds.), The Enhanced Forest Inventory and Analysis Program-National Sampling Design and Estimation Procedures. USDA For. Serv. Gen. Tech. Rep. SRS-80, pp. 27–42.

Benner, J.M., 2007. Browsing and Regeneration Monitoring Report for Pennsylvania's State Forests. Pennsylvania DCNR, Bureau of Forestry. 21 p.

Brose, P.H., 2008. Root development of acorn-origin oak seedlings in shelterwood stands on the Appalachian Plateau of northern Pennsylvania: 4-year results. For. Ecol. Manage. 255, 3374–3381.

Brose, P.H., Gottschalk, K.W., Horsley, S.B., Knopp, P.D., Kochenderfer, J.N., McGuinness, B.J., Miller, G.W., Ristau, T.E., Stoleson, S.H., Stout, S.L., 2008. Prescribing Regeneration Treatments for Mixed-Oak Forests in the Mid-Atlantic Region. USDA For. Serv. Gen. Tech. Rep. NRS-33. 100p.

Coates, K.D., 2002. Tree recruitment in gaps of various size, clearcuts and undisturbed mixed forest of interior British Columbia, Canada. For. Ecol. Manage. 155, 387–398.

Conover, W.J., 1999. Practical Nonparametric Statistics. Wiley, New York.

Decocq, G., Valentin, B., Toussaint, B., Hendoux, F., Saguez, R., Bardat, J., 2004. Soil seed bank composition and diversity in a managed temperate deciduous forest. Biodivers. Conserv. 13, 2485–2509.

Del Tredici, P., 2001. Sprouting in temperate trees: a morphological and ecological review. Botan. Rev. 67, 121–140.

Dixon, G.E., Keyser, C.E., (Comps.), 2008 (revised 2016). Northeast (NE) Variant Overview – Forest Vegetation Simulator. Internal Rep. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Management Service Center. 54p.

Horsley, S.B., Stout, S.L., deCalesta, D.S., 2003. White-tailed deer impact on the vegetation dynamics of a northern hardwood forest. Ecol. Appl. 13, 98–118.

Horsley, S.B., Marquis, D.A., 1983. Interference by weeds and deer with Allegheny hardwood reproduction. Can. J. For. Res. 13, 61–69.

Iverson, L.R., Dale, M.E., Scott, C.T., Prasad, A., 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (U.S.A.). Landsc. Ecol. 12, 331–348.

Jobidon, R., Roy, V., Cyr, G., 2003. Net effect of competing vegetation on selected environmental conditions and performance of four spruce seedling stock sizes after eight years in Québec (Canada). Ann. For. Sci. 60, 691–699.

Lieffers, V.J., Messier, C., Stadt, K.J., Gendron, F., Comeau, P.G., 1999. Predicting and managing light in the understory of boreal forests. Can. J. For. Res. 29, 796–811.

Lochhead, K.D., Comeau, P.G., 2012. Relationships between forest structure, understorey light and regeneration in complex Douglas-fir dominated stands in south-eastern British Columbia. For. Ecol. Manage. 284, 12–22.

Loftis, D.L., 1990. Predicting post-harvest performance of advance red oak reproduction in the Southern Appalachians. For. Sci. 36, 908–916.

Marquis, D.A., Ernst, R.L., Stout, S.L., 1992. Prescribing Silvicultural Treatments In Hardwood Stands of the Alleghenies. (Revised). USDA For. Serv. Gen. Tech. Rep. NE-96. 101p.

McRoberts, R.E., Westfall, J.A., 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. 60, 34–42.

McWilliams, W.H., Stout, S.L., Bowersox, T.W., McCormick, L.H., 1995. Adequacy of advance tree-seedling regeneration in Pennsylvania's forests. North. J. Appl. For. 12, 187–191.

McWilliams, W.H., Westfall, J.A., Brose, P.H., Dey, D.C., Hatfield, M., Johnson, K., Laustsen, K.M., Lehman, S.L., Morin, R.S., Nelson, M.D., Ristau, T.E., Royo, A.A., Stout, S.L., Willard, T., Woodall, C.W., 2015. A Regeneration indicator for Forest Inventory and Analysis: History, Sampling, Estimation, Analytics, and Potential Use in the Midwest and Northeast United States. USDA For. Serv. Gen. Tech. Rep. NRS-148. 74p.

Millington, J.D., Walters, M.B., Matonis, M.S., Laurent, E.J., Hall, K.R., Liu, J., 2011. Combined long-term effects of variable tree regeneration and timber management on forest songbirds and timber production. For. Ecol. Manage. 262, 718–729.

Petrides, G.A., 1941. Observations on the relative importance of winter deer browse species in central New York. J. Wildl. Manage. 5, 416–422.

Puhlick, J.J., Moore, M.M., Weiskittel, A.R., 2013. Factors influencing height-age relationships and recruitment of ponderosa pine regeneration in northern Arizona. West. J. Appl. For. 28, 91–96.

Rochefort, R.M., Little, R.L., Woodward, A., Peterson, D.L., 1994. Changes in sub-alpine tree distribution in western North America: a review of climatic and other causal factors. The Holocene 4, 89–100.

Stancioiu, P.T., O'Hara, K.L., 2006. Regeneration growth in different light environments of mixed species, multi-aged, mountainous forests of Romania. Eur. J. For. Res. 125, 151–162.

Standovár, T., Kenderes, K., 2003. A review on natural stand dynamics in beechwoods of East Central Europe. Appl. Ecol. Environ. Res. 1, 19–46.

Swanson, M.E., Franklin, J.F., Beschta, R.L., Crisafulli, C.M., DellaSala, D.A., Hutto, R.L., Lindenmayer, D.B., Swanson, F.J., 2011. The forgotten stage of forest succession: early-successional ecosystems on forest sites. Front. Ecol. Environ. 9, 117–125.

Teimouri, M., Gupta, A.K., 2013. On the three-parameter Weibull distribution shape parameter estimation. J. Data Sci. 11, 403–414.

U.S. Forest Service, 2013. Forest Inventory and Analysis National Core Field Guide, Volume I: field data collection procedures for phase 2 plots. Version 6.0.2, Northern Research Station Edition. <http://www.nrs.fs.fed.us/fia/data-collection/> (accessed 10 Feb., 2017).

Vickers, L.A., Fox, T.R., Loftis, D.L., Boucugnani, D.A., 2011. Predicting forest regeneration in the Central Appalachians using the REGEN expert system. J. Sustain. For. 30, 790–822.

Vickers, L.A., Larsen, D.R., Dey, D.C., Knapp, B.O., Kabrick, J.M., 2017. The impact of overstory density on reproduction establishment in the Missouri Ozarks: models for simulating regeneration stochastically. For. Sci. 63, 71–86.

Vonesh, E.F., Chinchilli, V.M., Pu, K., 1996. Goodness-of-fit in generalized nonlinear mixed-effects models. Biometrics 52, 572–587.

Woudenberg, S.W., Conkling, B.L., O'Connell, B.M., LaPointe, E.B., Turner, J.A., Waddell, K.L., 2010. The Forest Inventory and Analysis Database: Database Description and Users Manual Version 4.0 for Phase 2. U.S. For. Serv. Gen. Tech. Rep. RMRS-GTR-245. 339p.

York, R.A., Battles, J.J., Heald, R.C., 2003. Edge effects in mixed conifer group selection openings: tree height response to resource gradients. For. Ecol. Manage. 179, 107–121.

Zaczek, J.J., 2002. Composition, diversity, and height of tree regeneration, 3 years after soil scarification in a mixed-oak shelterwood. For. Ecol. Manage. 163, 205–215.