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 ≤ 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 ≥ 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):

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

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):

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

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

where $\hat{y}$ is the model prediction, $\hat{\bar{y}}$ 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 ≤ 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 ≤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