Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr Type III distribution

JEFFREY H. GOVE[1]*, MARK J. DUCEY[2], WILLIAM B. LEAK[1] AND LIANJUN ZHANG[3]

[1]USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, USA
[2]Department of Natural Resources, 221 James Hall, University of New Hampshire, Durham, NH 03824, USA
[3]College of Environmental Sciences and Forestry, State University of New York, One Forestry Drive, Syracuse, NY 13210, USA
*Corresponding author. E- mail: jgove@fs.fed.us

Summary

Stand structures from a combined density manipulation and even- to uneven-aged conversion experiment on the Bartlett Experimental Forest (New Hampshire, USA) were examined 25 years after initial treatment for rotated sigmoidal diameter distributions. A comparison was made on these stands between two probability density functions for fitting these residual structures: the Burr Type III and a finite mixture of Weibulls. All stands exhibited some form of uneven-aged structure (either reverse J or rotated sigmoid) after 25 years. The Burr distribution fits the final distributions as well as the more complicated finite mixture model in all cases.

Introduction

As we learn more about the structure of uneven-aged forests, we find that often the diameter distribution structure encountered may not fit the widely accepted, quintessential reverse J-shaped model. An example of such a departure is the so-called rotated sigmoid form. Rotated sigmoid-shaped tree diameter distributions have a slight to pronounced plateau or even a mild hump in the mid-diameter range of the distribution, often associated with small sawtimber size classes. These structures, when plotted with the logarithm of number of trees *vs* diameter, show a more exaggerated plateau or bump in the mid-diameters. In contrast, reverse J-shaped distributions plot as a straight line on the log scale and have a smooth, constant negative exponential decrease in number of trees from small to large diameters. Such distributions are often characterized by de Liocourt's q quotient (de Liocourt, 1898).

A consequence of the departure from the theoretically reverse J condition as found in stands characterized by rotated sigmoid structures is that traditional diameter distribution models are generally not flexible enough to model the plateau or hump in the mid-diameter classes (e.g. Zhang *et al.*, 2001). Traditional models such as the negative exponential (Meyer, 1952; Leak, 1964) do well with reverse J-shaped stands following

© Institute of Chartered Foresters, 2008. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

Forestry, doi:10.1093/forestry/cpm025
Advance Access publication date 5 February 2008

a given q. Weibull distributions (Bailey and Dell, 1973) contain the negative exponential as a special case, but offer a wider range of reverse J and unimodal shapes. Both of these models are relatively easy to work with and have been used extensively in forestry in the past several decades. Other distribution models, such as Johnson's S_B, have been put forth (Hafley and Schreuder, 1977), but mostly as general models, not specifically aimed at uneven-aged stands, and their use is somewhat more complicated. Recently, Zhang *et al.* (2001) have shown that finite mixtures of Weibull distributions may be useful in modelling rotated sigmoid distributions. Finite mixture distributions can take on a wide variety of shapes, including bimodal structures (Liu *et al.*, 2001), but they can also be more demanding to work with, at least in the estimation phase.

While rotated sigmoid distributions seem to turn up with increasing frequency in uneven-aged management, the cause for their genesis still appears to be open to research. Apparently open also is the question of whether or not they are characteristic of an equilibrium, or steady state, condition. For example, Goff and West (1975) argued that the rotated sigmoid form is a characteristic equilibrium condition in old-growth stands. These authors supported their hypothesis by examining 49 small old-growth stands, which they found to be characteristic of one of several rotated sigmoid forms. This evidence also seems to be supported by various modelling exercises. For example, growth models for northern hardwoods that have been set in the context of a non-linear optimization have produced rotated sigmoid structures when they have not been constrained otherwise (e.g. Adams and Ek, 1974; Gove and Fairweather, 1992). Similarly, Hansen and Nyland (1987) used simulations to arrive at a diameter distribution of pure sugar maple (*Acer saccharum* Marsh.) that was of the rotated sigmoid form. Both the optimization and simulation methods resulted in target stands that were sustainable in these modelling studies.

Other authors report that stands that appear as rotated sigmoid have been caused by disturbance of one form or another. Lorimer and Frelich (1984), for example, contend that the existence of a plateau or hump in such stands is due to past disturbance and provide simulations that support their contention that the plateau will disappear in time. Leak (1996), who examined several managed and old-growth stands, also notes that the existence of a rotated sigmoid structure may be due to past disturbance; however, he does not conjecture as to the equilibrium status of such stands. In a 20-year simulation study with several combinations of growth and mortality schedules acting on different initial stand structures, Leak (2002) found that different combinations would lead to, or maintain, rotated sigmoidal structures over time, while others would maintain or create a balanced negative exponential distribution. These simulations were designed to address the origins and maintenance of rotated sigmoid stands over the period of a cutting cycle, rather than the long-run steady state.

In Europe, the wisdom of strong management adherence to reverse J conditions is also being questioned, while rotated sigmoid structures are evidently beginning to be considered (Kerr and O'Hara, 2000; Westphal *et al.*, 2006). An example of this is the plenter management system, which often prescribes rotated sigmoid structures (O'Hara and Gersonde, 2004). Plentering has been around for over a century, and is described as a fully irregular forest, characterized by stems of all sizes and age classes (Schütz, 1997a). Anthropogenic in origin, plenter stands begin with a reverse J structure as the initial aim, which is then adjusted based on a simple growth model biased towards large-tree retention, with the overall goal of sustainability and stand volume maximization (Schütz, 1997b). Schütz (1997a, b) contends that there must be continuous surveillance to establish and realize sustainability of the target structure. Evidently, however, once established, it is largely self-regulating; therefore, the rotated sigmoid structures defined by plentering appear to be disturbance oriented initially, with apparent longer-term steady-state properties (Schütz, 1997a).

Whatever the cause of rotated sigmoid stand structure – and there may be many – it is clear that these are important distributional forms in uneven-aged management and silviculture.

Equally important then, is the ability to model such structures by relatively straightforward methods. In this paper, we apply a particular form of the Burr distribution, known as the Type III Burr for use in fitting rotated sigmoid diameter distributions. It is compared against a more flexible, but more complicated, finite mixture model (FMM). Both models are applied to 25-year remeasurements from a single long- term density management case study located in the White Mountains of central New Hampshire, to help assess the usefulness of the Burr Type III distribution in modelling rotated sigmoid stand structures. We look at the untransformed rather than the log scale as this is what is used for management and growth simulation models. Although distributions that appear reverse J, but with steep initial slope in the smaller classes, and nearly level numbers in the larger classes also plot as rotated sigmoid on the log scale, we restrict attention here to the stand structures found in this study, all of which appear rotated sigmoid on the untransformed scale.

Study area and methods

The Bartlett density study

The Bartlett density study was established in 1964 in a second- growth northern hardwood stand on the Bartlett Experimental Forest in Bartlett, New Hampshire, USA. The stand originated in the late 1800s from clear-cutting. At study inception, it was a typical even-aged northern hardwood stand composed primarily of American beech (*Fagus grandifolia* Ehrh.), red maple (*Acer rubrum* L.) and paper birch (*Betula papyrifera* Marsh), with minor components of sugar maple (*A. saccharum* Marsh.), white ash (*Fraxinus americana* L.) and yellow birch (*Betula alleghaniensis* Britton); scattered stems and localized clusters of eastern white pine (*Pinus strobus* L.), red spruce (*Picea rubens* Sarg.) and eastern hemlock (*Tsuga canadensis* (L.) Carr.) also occur. The soils in the study area were predominantly well drained, loamy sands derived from granite; over time, beech with an admixture of hemlock has become the predominant climax species on these soils (Leak, 1982). The results from this study apply particularly to stand/site conditions similar to these. The original purpose for the study was twofold: (1) to determine treatment combinations that would efficiently convert even-aged northern hardwood stands to an uneven-aged structure and (2) to determine the level of stand density that produced the best residual growth.

Prior to treatment, 48 ⅓- acre (0.13 ha) plots were established, with 50 ft (~15 m) wide buffers on all sides. The plots and adjoining buffers were treated conformably. Twelve treatment levels were applied in a randomized design resulting in four replications per treatment. The treatment levels correspond to four residual basal area density levels of 40, 60, 80 and 100 $ft^2\,acre^{-1}$ (9.2, 13.8, 18.4, 23 $m^2\,ha^{-1}$, respectively) and three levels of residual stand structure at 30, 45 and 60 per cent sawtimber. The residual per cent sawtimber levels were in terms of basal area of trees larger than 10.5 in (26.7 cm) diameter at breast height (d.b.h.). Unfortunately, no pre-treatment inventory was undertaken on the study area, hence we have no information on the pre-treatment stand structure. However, post-treatment (1964) measurements and several subsequent remeasurements of all trees greater than 4.5 inches (11.4 cm) d.b.h. have been conducted. Initial post-treatment stand structures for the 12 treatments are shown in Figure 1. Note that some of the stands show a clear rotated sigmoidal structure for the merchantable stems, while others are unimodal, or even approaching bimodal. Note also that the maximum d.b.h. in the stand post-treatment was 19 inches (48 cm). The stand was treated a second time in 1990, but not consistently with the first treatment. Therefore, we concentrate on the stand structures after the first 25 years of growth (the 1989 remeasurement), just prior to retreatment, and attempt to address the success of the conversion strategy to uneven-aged structure.

The Burr distribution

A flexible family of probability distributions, which can be derived from a single differential equation, was developed by Burr (1942). Two members of the family, the Burr Types III and

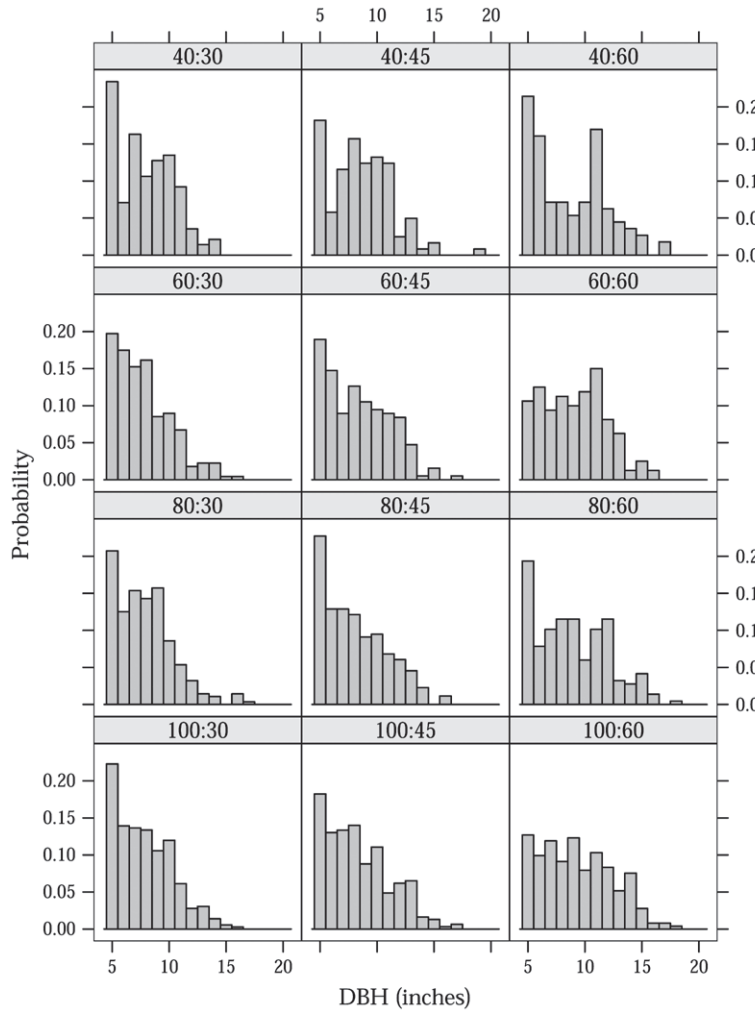

Figure 1. Residual 1964 post-treatment stand structures by treatment (basal area density:per cent sawtimber) on the Bartlett density study. (Note: 1 inch equals 2.54 cm.)

XII have been introduced to forestry by Lindsay *et al.* (1996). These distributions are important because they are inherently more flexible than the Weibull, which is often used in forestry applications. Both the Burr Types III and XII cover a much larger area of the skewness–kurtosis plane than the Weibull (Fry, 1993; Lindsay *et al.*, 1996; Rodriguez, 1977; Tadikamalla, 1980), with the Type III being the most flexible of the three. In this paper, we therefore restrict attention to the Type III Burr.

To derive several members of this family of distributions, Burr (1942) considered the simple differential equation on the cumulative distribution function (CDF) and a shaping function $g(x)$, with basic form

$$dF = F(1 - F)g(x)dx,$$

where $0 \leqslant F \leqslant 1$ is the CDF. When this expression is integrated, the solution becomes

$$F(x) = \left(1 + \mathrm{e}^{-\int g(x)\mathrm{d}x}\right)^{-1}.$$

Letting $g(x) = cx^{-1}$ and generalizing the inverse exponent above to $-k$ produces the CDF for the Type III form of the distribution presented by Burr (1942). Burr (1942) suggests that the CDF can be further modified by adding scale and location parameters for greater flexibility. Doing so yields the form of the Burr Type III CDF given by Lindsay *et al.* (1996) as

$$F(x;\boldsymbol{\theta}) = \left(1 + \left(\frac{b}{x-a}\right)^{c}\right)^{-k}, \quad x > 0,$$

with shape parameters $c, k > 0$, location parameter $a < x$ and scale parameter b such that $\boldsymbol{\theta} = (a, b, c, k)'$; the CDF is zero otherwise. The corresponding probability density function (PDF) is

$$f(x;\boldsymbol{\theta}) = \frac{kc}{b}\left(\frac{b}{x-a}\right)^{c+1}\left(1 + \left(\frac{b}{x-a}\right)^{c}\right)^{-(k+1)}.$$

This particular version of the Burr distribution is notable, not only in the general flexibility of form mentioned above, but specifically because it is able to fit certain rotated sigmoid distributional forms. Figure 2 presents a set of Type III PDFs with scale parameters $5 \leqslant c \leqslant 9$ and $0.09 \leqslant k \leqslant 0.13$, illustrating a number of possible rotated sigmoid and reverse J shapes over a scaled region of the random variable x. Given this range of k, distributions with $c < 5$ tend more to reverse J shaped, while those with $c > 9$ tend to be more unimodal.

Parameter estimation was accomplished via maximum likelihood (ML). The log likelihood for the Burr Type III is particularly simple and is given as

$$\begin{aligned}\mathcal{L}(\mathbf{x};\boldsymbol{\theta}) &= n(\log k + \log c - \log b)\\ &\quad + (c+1)\sum_{i=1}^{n}\log\left(\frac{b}{x_i - a}\right)\\ &\quad + (1-k)\sum_{i=1}^{n}\log\left(1 + \left(\frac{b}{x_i - a}\right)^{c}\right),\end{aligned}$$

where n is the sample size. This expression can be maximized directly for the parameter estimates $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c}, \hat{k})$. In addition, the location parameter, a, can be held fixed and the expression $\mathcal{L}(\mathbf{x};\boldsymbol{\theta})$ is then maximized for the remaining three parameters.

The Burr Type III distribution has been used in other fields and goes by various synonyms. For example, in economics it is known as the Dagum distribution (Dagum, 1977), and McDonald (1984) noted that it was related to the generalized beta of the second kind. Mielke (1973) and Mielke and Johnson (1974) have used this in meteorology and water resources applications and shown it to be a special case of their Kappa distribution (Mielke and Johnson, 1973; Tadikamalla, 1980).

Finite mixtures of Weibull distributions

In general, finite mixtures are composed of g individual PDFs, not necessarily all of the same form or family, which are individually 'mixed', or weighted by proportions $0 \leqslant \pi_i \leqslant 1$ such that $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_g)$ and $\sum_{i=1}^{g}\pi_i = 1$. Here, we use Weibull densities as the components of the mixture and take $g = 2$ as in Zhang *et al.* (2001). The three- parameter Weibull PDF is given as

$$f(x;\boldsymbol{\xi}) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \mathrm{e}^{-((x-a)/b)^{c}},$$

where $a < x$ is the location parameter, with scale and shape parameters $b > 0$ and $c > 0$, respectively, and $\boldsymbol{\xi} = (a, b, c)'$. Note that these parameters are distinct from those in the Burr distribution presented earlier and we assume in what follows that the context will allow differentiation between the different models. The corresponding CDF is simply

$$F(x;\boldsymbol{\xi}) = 1 - \mathrm{e}^{-((x-a)/b)^{c}}.$$

In general, the g- component mixture density is defined as

$$f(x;\boldsymbol{\Psi}) = \sum_{i=1}^{g}\pi_i f_i(x;\boldsymbol{\xi}_i),$$

where $f_i(x;\boldsymbol{\xi}_i)$ is the ith component density, $i = 1, \ldots, g$. Note that each component density is parameterized in terms of an m-dimensional vector of parameters $\boldsymbol{\xi}_i$ such that $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_g)$. Therefore, collecting all parameters,

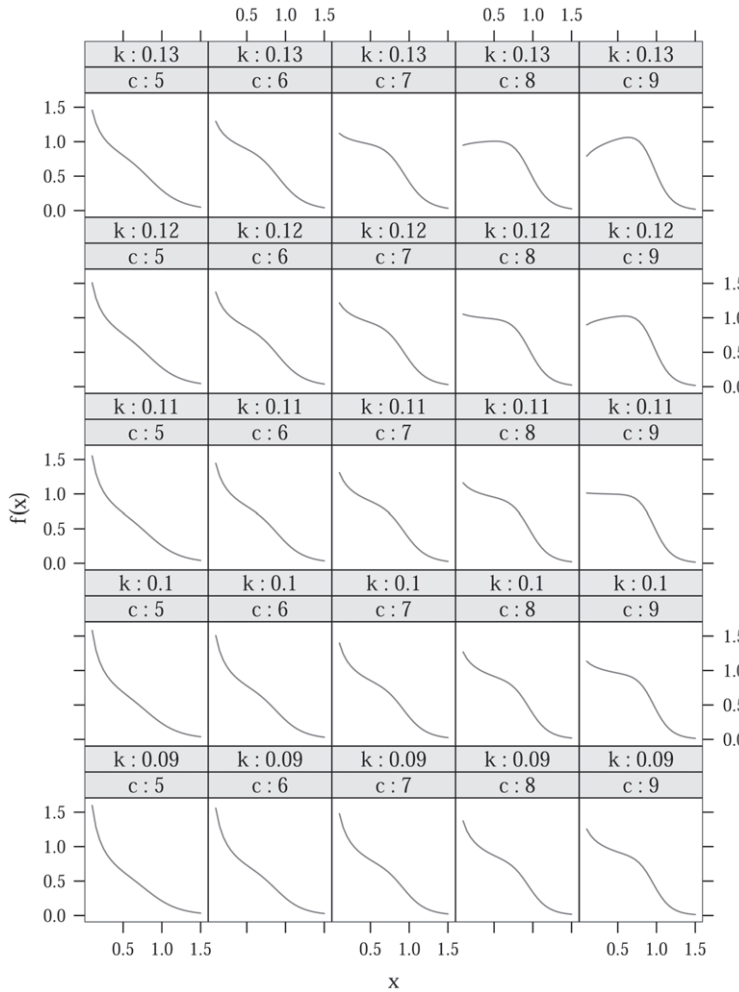

Figure 2. Examples of possible rotated sigmoid structures for the Burr Type III distribution with shape parameters $5 \leqslant c \leqslant 9$ and $0.09 \leqslant k \leqslant 0.13$, over the scaled random variable x on the abscissa.

$\boldsymbol{\Psi} = (\boldsymbol{\pi}, \boldsymbol{\Xi})$. The log likelihood follows naturally as (McLachlan and Peel, 2000, p. 47)

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\Psi}) = \sum_{j=1}^{n} \log f(x; \boldsymbol{\Psi}) = \sum_{j=1}^{n} \log \left\{ \sum_{i=1}^{g} \pi_i f_i(x; \boldsymbol{\xi}_i) \right\},$$

where random vector $\mathbf{x} = (x_1, \ldots, x_n)$ denotes a sample of diameters. Conceptually, parameter estimation for FMMs is straightforward, but tedious, as it is complicated by multimodality of the likelihood surface (McLachlan and Peel, 2000, p. 54). We maximize the log likelihood indirectly by driving the sums of squares of the gradient equations to zero using the following objective.

$$\min_{\{\boldsymbol{\pi}, \boldsymbol{\Xi}\}} \sum_{i=1}^{g-1} \left(\frac{\partial \mathcal{L}(\mathbf{x}, \boldsymbol{\Psi})}{\partial \pi_i} \right)^2 + \sum_{k=1}^{m} \sum_{i=1}^{g} \left(\frac{\partial \mathcal{L}(\mathbf{x}, \boldsymbol{\Psi})}{\partial \boldsymbol{\xi}_{i,k}} \right)^2, \tag{1}$$

where in the second term the derivatives with respect to $\boldsymbol{\xi}_{i,k}$ are the individual parameter derivatives with respect to the kth parameter, summed over all g mixture components. As with the Burr Type III, the location parameter, a, for each Weibull component density may be fixed rather than estimated, in which case, the respective gradient components do not contribute to the second term above. One constraint was applied to this minimization problem to ensure that the means for each component were ordered in ascending order; that is, $\bar{x}_1 \leqslant \bar{x}_2 \leqslant, \ldots, \leqslant \bar{x}_g$. This constraint is merely a convenience to assure that the component PDFs appear to increase in X and does not affect the resulting solutions.

For the two- component mixture used in this study, $\pi_2 = 1 - \pi_1$. Therefore, our mixture density is simply

$$f(x; \boldsymbol{\Psi}) = \pi_1 f_1(x; \boldsymbol{\xi}_1) + (1 - \pi_1) f_2(x; \boldsymbol{\xi}_2).$$

In general, estimating all g proportions makes for an overdetermined system; therefore, only π_1 is estimated in equation (1) with this two-component mixture. Hence, in the following we simply drop the subscript and refer to π as the known proportion or $\hat{\pi}$ as its maximum likelihood estimate (MLE). The full vector of estimated parameters under the two- component Weibull mixture is $\hat{\boldsymbol{\Psi}} = (\hat{\pi}, \hat{a}_1, \hat{b}_1, \hat{c}_1, \hat{a}_2, \hat{b}_2, \hat{c}_2)$. Therefore, if both location parameters are to be estimated, this FMM requires seven total parameters to be estimated whereas the Burr only requires four.

Results

Figure 3 presents the histograms and fitted Burr and FMM PDFs for the Bartlett density study in 1989; the histograms may be compared directly with Figure 1. The first point to notice is with regard to the stand structural evolution over the 25- year growth period from 1964 to 1989. As previously mentioned, the initial stand diameter distributions after treatment ranged from unimodal rotated sigmoid to bimodal, to a few that do not fall readily into any apparent smooth form. However, over the 25- year growth period, growth and mortality have smoothed the stand structures such that the terminal stand conditions fall into one of two reasonably well-defined structures; i.e. either rotated sigmoid or approximate reverse J shaped with the former predominating. In addition, the maximum tree d.b.h. has increased from 19 to 26.3 inches (66.8 cm) over the 25-year period.

The well- defined stand structures at the end of the growth period facilitated the estimation of the proposed PDF models in Figure 3. Perhaps, the most striking result of the model fits is that the Burr and FMM curves are so similar for every treatment level. Indeed, in a few of the treatments, such as the 100-ft^2 level, the curves appear to be almost coincident. For the other stands, the trend seems to be that the Burr decreases more rapidly in the smaller diameter classes and does not reach as well- defined a 'plateau' in the larger classes as the FMMs. The only exception to this is the 80:45 treatment, where the appearance of a small spike in the number of trees in the 9- inch class is enough to cause the FMM to actually develop a mode in this region.

In general, the higher basal area treatments tend to all have developed well-defined rotated sigmoidal diameter distributions 25 years after treatment (Figure 3). However, the low-density (40 ft^2) treatments have a somewhat different characteristic. Note that in the 40:30 treatment, the Burr tends to fit a slight 'hump' in the mid-diameter range, but in the FMM this is less pronounced. This slight difference between the two model fits may point to a stand in transition to a reverse J-shaped form. The other two residual sawtimber treatments in this low density set of treatments show a clearly defined transition to approximate reverse J in both of the model fits. Note that the stand is not in the quintessential reverse J condition since the FMM is still composed of two components: if the stand were exactly negative exponential, then the best fit would be a single Weibull density with unity shape parameter. Instead, both the Burr and the FMM PDFs tend to have a slightly smaller negative slope in the mid-diameter classes than a reverse J.

The best fits for both the Burr (Table 1) and FMMs (Table 2) were with fixed location

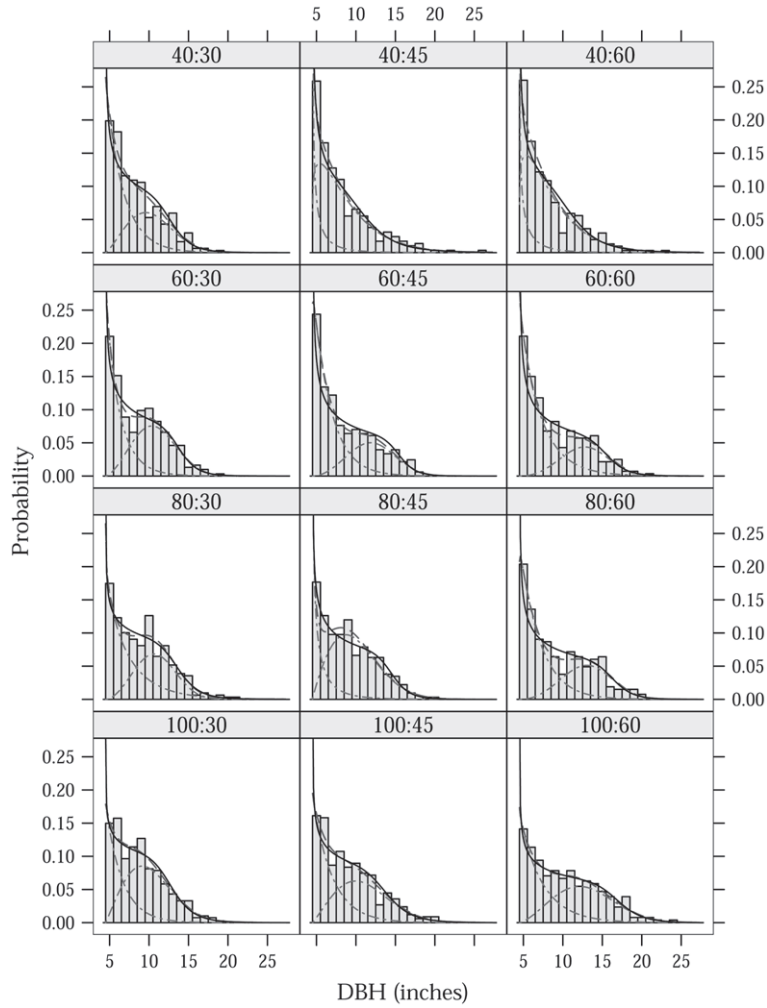

Figure 3. Stand structures with fitted Burr (solid) and FMM (dashed, component PDFs light dot-dashes) by treatment (basal area density:per cent sawtimber) after 25 years of growth on the Bartlett density study. (Note: 1 inch equals 2.54 cm.)

parameters. In the case of the Burr distribution, parameter estimates converged with the location parameter free in all cases, but the Akaike information criteria (AIC) (Burnham and Anderson, 1998) was minimized in the limit as this parameter is shifted towards the minimum diameter value. Thus, a value close to the minimum was fixed as shown in the table. The minimum diameter value itself cannot be used due to numerical considerations in the likelihood equations in both the Burr and FMMs. It is well-known that a simple three-parameter Weibull distribution, with location, shape and scale parameters, is more challenging to fit than the two-parameter version without the location parameter (Murthy *et al.*, 2004, p. 81). Therefore, it should not be surprising that, when desiring to estimate two free Weibull location parameters in a FMM context, the challenge becomes somewhat Herculean. Indeed, estimates for only a few of the stands converged after numerous attempts with differing starting values when estimating the location

Table 1: ML parameter estimates for the Burr Type III distribution by treatment in 1989 (location parameters fixed)

Treatment	a	$\hat{k}$	$\hat{c}$	$\hat{b}$
40:30	4.495	0.108	6.666	8.801
40:45	4.495	0.178	3.699	8.088
40:60	4.495	0.182	3.828	7.895
60:30	4.495	0.084	8.422	9.703
60:45	4.495	0.058	10.802	11.291
60:60	4.495	0.073	8.838	11.492
80:30	4.495	0.117	6.906	9.634
80:45	4.495	0.077	9.206	10.422
80:60	4.495	0.072	9.505	12.206
100:30	4.495	0.138	6.188	8.887
100:45	4.495	0.129	6.077	9.993
100:60	4.495	0.093	8.252	13.015

Table 2: ML parameter estimates for the FMM distribution by treatment in 1989 (location parameters fixed)

Treatment	$\hat{\pi}$	a_1	$\hat{c}_1$	$\hat{b}_1$	a_2	$\hat{c}_2$	$\hat{b}_2$
40:30	0.571	4.495	0.993	2.234	4.495	2.246	6.609
40:45	0.193	4.495	0.632	0.963	4.495	1.175	4.569
40:60	0.166	4.495	0.695	1.207	4.495	1.166	4.339
60:30	0.482	4.495	0.956	1.779	4.495	2.574	7.100
60:45	0.632	4.495	1.018	2.209	4.495	2.946	8.484
60:60	0.679	4.495	0.976	2.661	4.495	3.255	9.265
80:30	0.561	4.495	0.930	3.082	4.495	2.813	7.156
80:45	0.265	4.495	0.701	1.185	4.495	1.904	6.280
80:60	0.602	4.495	1.025	2.541	4.495	3.182	9.615
100:30	0.372	4.495	0.999	2.088	4.495	2.050	6.421
100:45	0.475	4.495	0.996	2.476	4.495	2.099	7.425
100:60	0.478	4.495	0.989	2.926	4.495	2.436	9.452

parameters as well. It should be noted that in those cases where the free location fits did converge, the corresponding fixed- location fits proved better, as in the case with the Burr. Thus, the location parameters for the FMMs were also fixed at the same value as the Burr (Table 2). Even with this constraint, the resulting FMM likelihoods can be multimodal, as will be discussed subsequently.

Many indexes of fit have been used for judging the adequateness of PDF models. For example, Zhang *et al.* (2001) used both root mean square error (RMSE) and likelihood ratio χ^2 statistics. Information–theoretic approaches, however, provide another, perhaps better, alternative as detailed in Burnham and Anderson (1998). These authors suggest several such criteria and provide a very cogent discussion of their strengths. In general, these criteria are applicable to both nested and non- nested models, which is often not the case with other statistics (Burnham and Anderson, 1998, p. 63). They also have an appealing interpretation of providing a method for selecting the best model from a pool of candidates based on the 'strength of evidence' from the data supporting a given model. We use AIC since the ratio of n/K is large for each treatment (Burnham and Anderson, 1998, p. 76); the model with the minimum AIC value is the best of the two candidate models. AIC takes a very simple form that is easy to calculate from the optimization results (Burnham and Anderson, 1998, p. 60):

Table 3: AIC and RMSE (in trees per acre) values for the Burr and FMM fits by treatment in 1989

	AIC			RMSE	
Treatment	FMM	Burr	Best fit	FMM	Burr
40:30	1399.3	1394.2	Burr	2.6	3.3
40:45	1349.5	1345.7	Burr	1.8	2.2
40:60	1410.4	1411.3	FMM	2.7	3.1
60:30	1439.6	1434.8	Burr	2.0	2.6
60:45	1571.8	1577.2	FMM	2.0	3.0
60:60	1384.4	1380.4	Burr	1.8	2.6
80:30	1525.2	1517.5	Burr	2.3	2.5
80:45	1540.9	1532.1	Burr	2.3	2.1
80:60	1352.2	1351.0	Burr	1.8	2.5
100:30	1905.8	1901.0	Burr	3.0	3.2
100:45	1678.9	1673.5	Burr	2.6	3.0
100:60	1382.2	1374.4	Burr	1.2	1.4

$$\mathrm{AIC} = \begin{cases} -2\mathcal{L}(\mathbf{x}; \hat{\boldsymbol{\theta}}) + 2K, & \text{Burr Type III,} \\ -2\mathcal{L}(\mathbf{x}, \hat{\boldsymbol{\Psi}}) + 2K, & \text{FMM,} \end{cases}$$

where K is the number of estimated model parameters in each case. Because we have fixed the location parameters, these do not count in the determination of AIC for either model. Therefore, $K = 3$ for the Burr Type III, while $K = 5$ for the Weibull FMM. Notice that the number of parameters acts as a type of penalty on the likelihood, so that models with fewer parameters benefit. Table 3 shows the AIC values for both models and for each of the treatments. This table also indicates the 'best' model based on choosing the model with lowest AIC value. It is interesting to note that the simpler Burr model tends to be preferred in all but two cases. However, the AIC values for the two models are quite close in many of the treatments – in fact all – suggesting that either model could be used. This conclusion is validated by the fitted curves in Figure 3, where, for most practical purposes, the results are quite similar.

It is also instructive to look at the so-called residuals from each of the model fits. These are defined as simply the predicted minus the actual number of trees for each model and treatment, by diameter class, and are shown in Figure 4. These plots essentially echo what can be gleaned from the differences between the histograms and the fitted curves in Figure 3. They do, however, help to highlight the fact that the major differences between the model predictions and the actual data are in the smaller diameter classes. They also show that, with few exceptions, there are only minor differences between the two models in all treatments, representing only a few trees per acre at worst in any given diameter class.

For completeness, we also present the RMSE statistic used by Zhang *et al.* (2001). It is simply the square root of the mean- squared residuals and so may be thought of as a summary index of the residual graphs – a standard deviation of sorts. The RMSE values are also presented in Table 3. In almost all cases, the RMSEs are slightly lower for the FMM. However, it should be noted that RMSE is in terms of trees per acre; therefore, the overall error is quite small. Indeed, if we average the RMSE values over all treatments, we find a difference of only 0.46 trees per acre (~1 tree per ha) between the two models. Again, these results support the conclusion that there is very little difference – none in practical terms – between the two models.

Discussion

Residual stand structure

The treatments applied in this study have resulted in two fairly distinct stand structures over

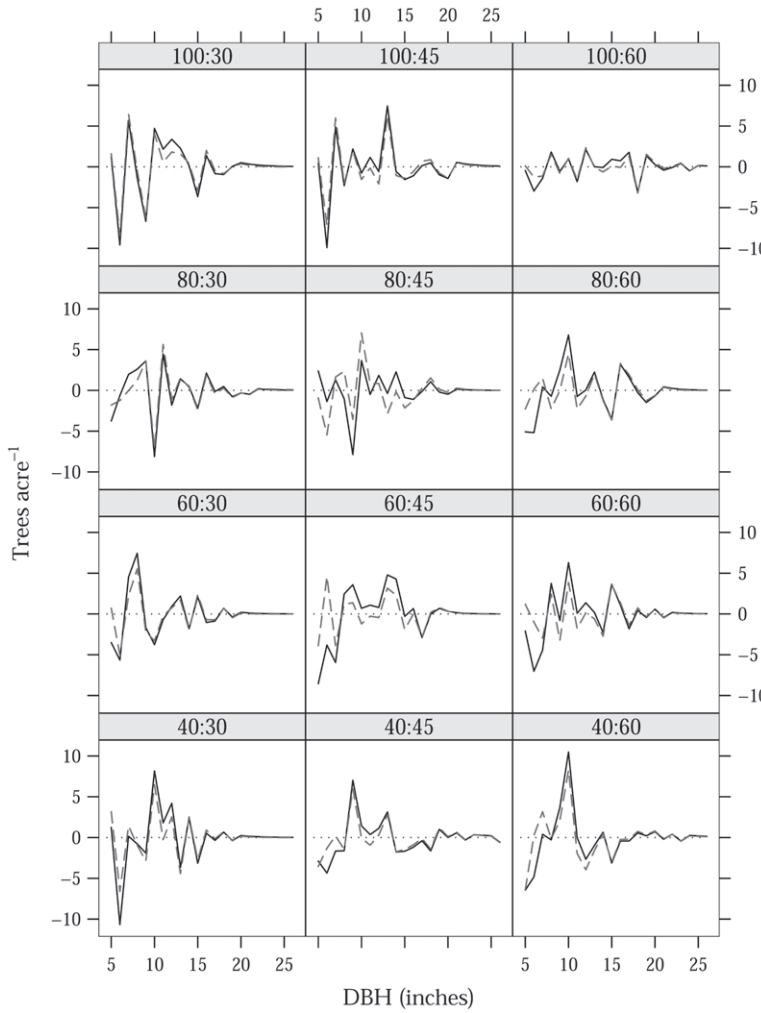

Figure 4. Residuals from fitted Burr (solid) and FMM (dashed) by treatment (basal area density:per cent sawtimber). (Note: 1 inch equals 2.54 cm.)

time: those evidently approaching reverse J and rotated-sigmoid. The quasi-reverse J-shaped structures are characteristic of the low-density treatments, although the 40:30 treatment with low residual sawtimber component still has retained some of its sigmoidal form. As previously mentioned, all of the initial stand structures were somewhat similar post-treatment in 1964 (Figure 1). Therefore, the question arises as to why the low-density treatments have tended to the reverse J structure? Unfortunately, while density may play a role, it is undoubtedly not the sole driver, but may contribute subtly to changes in species composition in certain diameter classes whose growth rates could push the stand in one structural direction or the other. Trends from the remeasurements over time do, however, show that the 40- and 60-ft^2 treatments have had a very large amount of ingrowth into the smallest diameter class, inflating the distribution in these lower diameter classes, while the higher density treatments show little change in these classes over time. Indeed, Leak and Gove (2008) have reported that the ingrowth in the 40- and 60-ft^2 classes are two to

three times that of the 80- and 100-ft^2 classes in terms of basal area. However, no measurements of sapling or seedling size classes were taken until after the second treatment in 1990, so we have no early quantitative information on the structure and dynamics of these smaller trees.

Another interesting question is whether the rotated sigmoid structure may only be transitional in northern hardwoods, and that given enough time, they will eventually move to a more reverse J form regardless of initial density. Leak (1996) looked at a number of uneven-aged compartment studies on the Bartlett Experimental Forest, plus one old-growth stand in New Hampshire's White Mountains known as the Bowl. The Bowl showed a stand distribution that could only be characterized as reverse J, while the other stands had either reverse J or rotated sigmoidal structures. Leak (1996) concluded, in accordance with Schmelz and Lindsey (1965), that the rotated sigmoid form was caused by disturbance and that something close to reverse J was the 'steady-state' condition for old-growth northern hardwoods. Unfortunately, we may not be able to answer this question with any degree of confidence based on remeasurements from the current study, since the retreatment in 1990 left only one replicate for long-term tracking of the original treatments. We can, however, address whether the initial treatments were successful in structural conversion from even- to uneven-aged stands. Whether transitional, or steady state, both the reverse J and rotated-sigmoid diameter distributions are now generally regarded as being indicative of the uneven- aged condition and all of our stands in the current study fit into one of these two structures. This, coupled with the fact that there are at least three extant age classes on the study (Leak and Solomon, 1975; Leak, 2004), allows us to conclude that at least one of the initial goals of the study was indeed successful after 25 years.

Stand structural modelling

The 25-year residual stand structures encountered in this study offer a good range of at least quasi steady-state structures encountered in uneven-aged stands. The simplicity and flexibility of the Burr Type III distribution commends it as a useful model for reverse J and rotated-sigmoidal stand structures, as well as many unimodal structures often found in even-aged stands. However, the Burr Type III does not possess the ability to contort, for example, to multiple modes like the FMMs do, making it less than suitable for some of the initial structures found, for example, in Figure 1. The ease with which the Burr Type III is able to be fit using simple optimization methods directly on the log likelihood is also a very important feature of this distribution.

FMMs allow added flexibility in general when compared with simple probability models like the Burr. For one, as we have seen here (e.g., the 80:45 treatment), and has been shown more lucidly in other studies (Liu *et al.*, 2001; Zhang *et al.*, 2001), the ability to model bimodality in the diameter distribution is often useful, especially in mixed-species stands. In addition, Zasada and Cieszewski (2005) indicate that the component distributions might be used to discriminate other biologically relevant information, such as tree height classes. This may indeed be possible, as these authors demonstrate, depending on the component PDFs used and the type of information desired. It should be noted, however, that Zasada and Cieszewski (2005) were unable to use the Weibull distribution as a component model for their unimodal stands because the estimation method used, which attempts to estimate the means and standard deviations of the component PDFs rather than estimate the parameters directly (Macdonald and Pitcher, 1979), can fail when there is strong overlap in the component densities.

It would be interesting to think that the weight parameter played an intuitive role in the final form of the mixture distribution. For example, one would think that π would weight more heavily towards the reverse J-shaped component in the 40:45 and 40:60 treatments since that seems to be the predominant structure, but exactly the opposite is the case: the J-shaped component density (reverse J PDFs are those whose shape parameter is in the range $0 < c \leqslant 1$ for Weibulls) receives little weight in these cases (Table 2). This may be in

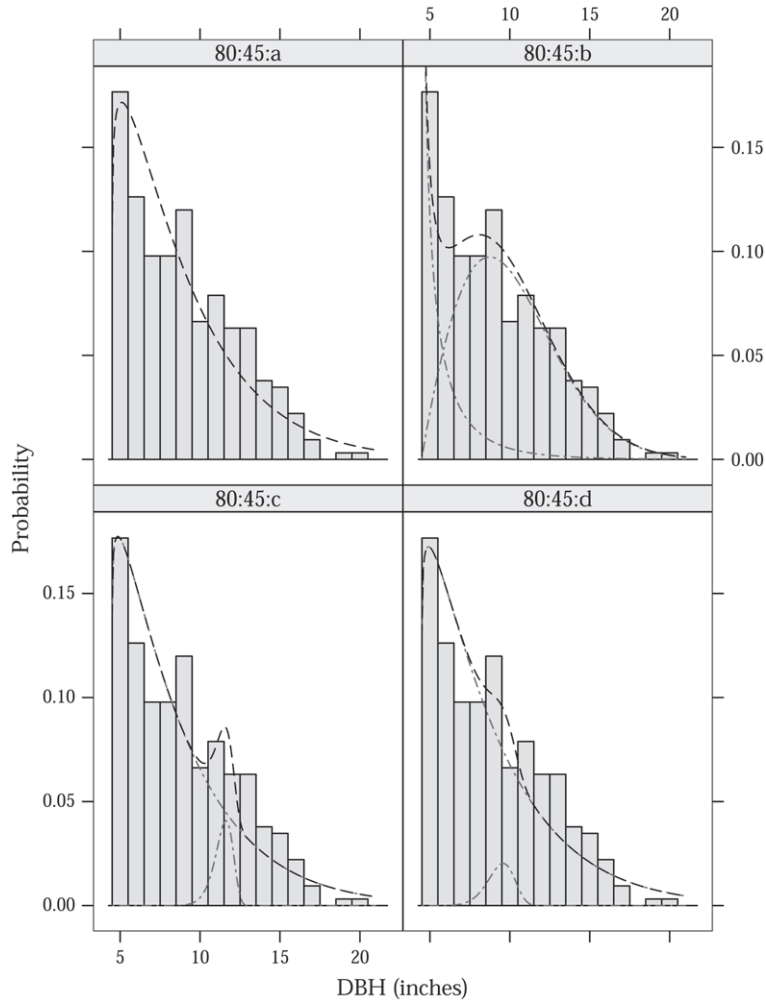

Figure 5. Four ML solutions for the FMM (heavy dashes, component densities light dot-dash) in the 80:45 treatment illustrating the multimodality of the likelihood with AIC values (a) 1594.8, (b) 1540.9, (c) 1587.4 and (d) 1593.4. (Note: 1 inch equals 2.54 cm.)

part due to the multimodality of the likelihood – perhaps we did not find the overall optimum – but is more likely due to the fact that the model does not necessarily translate as well to biological interpretation as one would like to hope. In other words, it is fitting the data as best it can, with no constraints or preconceptions with respect to foreknowledge of inherent reverse J conditions.

The added flexibility of the FMM approach does also bring with it some associated concerns to be aware of that may make estimation more difficult than with simple non-mixture PDF models such as the Burr. First, as we have already mentioned, with distribution models like the Weibull, where there is an inherent difficulty fitting one of the parameters in the single-PDF setting, these results will be magnified in the mixture setting. With the three- parameter Weibull model, the concern is with the location parameters of the mixture. The inability of a time-tested non-linear solver, such

Table 4: Correlations between parameter estimates over all treatments

	FMM					Burr		
	$\hat{\pi}$	$\hat{c}_1$	$\hat{b}_1$	$\hat{c}_2$	$\hat{b}_2$	$\hat{k}$	$\hat{c}$	$\hat{b}$
$\hat{\pi}$	1.0	0.86	0.82	0.94	0.84	−0.75	0.69	0.62
$\hat{c}_1$		1.0	0.82	0.79	0.78	−0.59	0.56	0.55
$\hat{b}_1$			1.0	0.77	0.79	−0.47	0.42	0.62
$\hat{c}_2$				1.0	0.89	−0.84	0.81	0.73
$\hat{b}_2$					1.0	−0.81	0.79	0.93
$\hat{k}$						1.0	−0.98	0.79
$\hat{c}$							1.0	0.80
$\hat{b}$								1.0

as GRG2 (Lasdon *et al.*, 1978) to converge in all but a few treatments when the location parameters were free, speaks to the difficulty of estimation in such situations. In those treatments where it did converge, only one of the components was selected ($\pi \equiv 0$ or 1). As mentioned earlier, Zasada and Cieszewski (2005) had similar problems fitting component Weibulls with the Macdonald and Pitcher (1979) approach.

A second potential concern with the mixture approach, which has also been mentioned previously, is the evident multimodality of the log likelihood surface. Based on the experience of fitting the Weibull mixtures to the Bartlett density study data, it is important to try multiple starting values when endeavouring to find the optimum, as the likelihood surface appears to be multimodal, at least in some cases. In fact, it might not be a bad idea to embed the optimization within another directed search algorithm to try to find the global maximum; however, in such a scenario, it should be kept in mind that the embedded mixture optimization may not always converge from every starting point, and non-convergent results must be rejected. Finally, it is often the case that a likelihood solution mode is coincident with the boundary case of $\pi = 0$ or 1 and such solutions are easily distinguished. This is illustrated in Figure 5a, along with three other convergent solutions. These four solutions illustrate that there are indeed local optima in the likelihood surface for these data. Figure 5b shows the final solution with minimum AIC as reported in Table 2. Note that it is relatively easy to choose the best fitting FMM by eye based on these results. However, consider the case where the ML algorithm was run in a batch mode for fitting FMMs to many stands, say in a large modelling effort. This example illustrates the need for testing AIC values at multiple starting points in such a setting. If the likelihood surface were not suitably explored, one could end up with one of the less desirable fits shown in this figure. Again, to our knowledge, the Burr Type III likelihood does not suffer from such multimodality concerns.

In all FMMs, the first component was some form of reverse J (or very close, e.g. the 60:45 and 80:60 treatments have $\hat{c}_1 > 1$, but only slightly) and the second possessed a clear mode as indicated by the respective shape parameters $\hat{c}_1$ and $\hat{c}_2$ (Table 2). These components are clearly visible in the component densities shown in Figure 3. It is interesting to note that correlation among the parameters of both distributions is high not only within the given distribution but also between the FMMs and Burrs in some cases (Table 4). For example, all three of the Burr parameters are fairly strongly correlated with both the scale and shape parameters of the second Weibull component ($\hat{b}_2$ and $\hat{c}_2$). These two Weibull parameters are also correlated to a lesser extent with the corresponding parameters for the first component in the mixture. In addition, the second component's shape parameter, $\hat{c}_2$, is very highly correlated with the FMM weight, $\hat{\pi}$. It is difficult to draw any conclusions with regard to causality from these correlations. However, it might be inferred that the reason for the high

correlations between the Burr and the second FMM component parameters goes back to Figure 2, where we note that the Burr is rotated sigmoidal in this range of parameters (compare with Table 1) and that the second Weibull component defines the associated plateau in the fitted FMM. Because the plateau in the rotated sigmoid fits is a strong definer of the overall distribution shape, it is also plausible that this might explain the correlation between the estimated FMM weight and $\hat{c}_2$. Regardless of the possible explanations, the relatively high correlations may be useful in subsequent modelling contexts where parameters are predicted from a combination of both stand variables and associated parameter estimates from previously fitted stands.

Conclusions

The diameter distributions for the northern hardwood stands in this study were equally well fit by both the Burr Type III and a finite mixture of Weibull distributions. We have shown that while the FMM distribution is a more flexible distribution for modelling overall, it is significantly more difficult to find correct parameter estimates for it than for the Burr. Therefore, while either method could be used for fitting reverse J shaped to rotated sigmoidal stand structures similar to those found in this study, the Burr Type III distribution can be recommended in such stands based on parsimony and ease of fit for the modeller. If, however, a pronounced hump or other mode occurs in the mid range of the diameter distribution, finite mixtures will probably be a better choice.

Acknowledgements

The authors would like to thank Professor Kevin O'Hara for his helpful review of the manuscript. We would also like to thank Dr Dale S. Solomon (deceased) for his foresight and untiring work on establishing and running this study for almost 40 years.

Conflict of Interest Statement

None declared.

References

Adams, D.M. and Ek, A.R. 1974 Optimizing the management of uneven-aged forest stands. *Can. J. For. Res.* **4**, 274–286.

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

Burnham, K.P. and Anderson, D.R. 1998 *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer.

Burr, I.W. 1942 Cumulative frequency functions. *Ann. Math. Stat.* **13**, 215–232.

Dagum, C. 1977 A new model of personal income distribution: specifications and estimation. *Econ. Aplique* **30**, 413–436.

de Liocourt, F. 1898 *De l'amenagement des sapinières*. Technical Report, Bulletin trimestriel, Société forestière de Franche-Comtè et Belfort. http://www.snr.missouri.edu/forestry/larsen.html.

Fry, T.R.L. 1993 Univariate and multivariate Burr distributions: a survey. *Pak. J. Stat.* **9**, 1–24.

Goff, F.G. and West, D. 1975 Canopy-understory interaction effects on forest population structure. *For. Sci.* **21**, 98–108.

Gove, J.H. and Fairweather, S.E. 1992 Optimizing the management of uneven-aged stands: a stochastic approach. *For. Sci.* **38**, 623–640.

Hafley, W.L. and Schreuder, H.T. 1977 Statistical distributions for fitting diameter and height data in even-aged stands. *Can. J. For. Res.* **7**, 481–487.

Hansen, G.D. and Nyland, R.D. 1987 Effects of diameter distribution on the growth of simulated uneven-aged sugar maple stands. *Can. J. For. Res.* **17**, 1–8.

Kerr, G. and O'Hara, K.L. 2000 Uneven-aged silviculture: common myths explored. *Q. J. For.* **94**, 145–150.

Lasdon, L.S., Warren, A.D., Jain, A. and Ratner, M. 1978 Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Trans. Math. Softw.* **4**, 34–50.

Leak, W.B. 1964 The J-shaped probability distribution. *For. Sci.* **11**, 405–409.

Leak, W.B. 1982 Habitat mapping and interpretation in New England. *Research Paper* NE-496. US Forest Service, Northeastern Forest Experiment Station. 28 p.

Leak, W.B. 1996 Long-term structural change in uneven-aged northern hardwoods. *For. Sci.* **42**, 160–165.

Leak, W.B. 2002 Origin of sigmoid diameter distributions. *Research Paper* NE-718. USDA Forest Service, Northeastern Research Station, Newtown Square, PA.

Leak, W.B. 2004 Diameter growth in even- and uneven-aged northern hardwoods in New Hampshire under partial cutting. *North. J. Appl. For.* **21**, 160–163.

Leak, W.B. and Gove, J.H. 2008 Growth of northern hardwoods in New England – a 25-year update. *North. J. Appl. For.* (in press).

Leak, W.B. and Solomon, D.S. 1975 Influence of residual stand density on regeneration of northern hardwoods. *Research Paper* NE- 310. USDA Forest Service, Northeastern Research Station, Upper Darby, PA.

Lindsay, S.R., Wood, G.R. and Woollons, R.C. 1996 Modelling the diameter distribution of forest stands using the Burr distribution. *J. Appl. Stat.* **23**, 609–619.

Liu, C., Zhang, L., Davis, C.J., Solomon, D.S. and Gove, J.H. 2001 A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. *For. Sci.* **48**, 653–661.

Lorimer, C.G. and Frelich, L.E. 1984 A simulation of equilibrium diameter distributions of sugar maple (*Acer saccharum*). *Bull. Torrey Bot. Club* **111**, 193–199.

Macdonald, P.D.M. and Pitcher, T.J. 1979 Age-groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. *J. Fish. Res. Board Can.* **36**, 987–1001.

McDonald, J.B. 1984 Some generalized functions for the size distribution of income. *Econometrica* **52**, 647–664.

McLachlan, G. and Peel, D. 2000 *Finite Mixture Models*. John Wiley and Sons.

Meyer, H.A. 1952 Structure, growth, and drain in balanced uneven-aged forests. *J. For.* **50**, 85–92.

Mielke, P.W., Jr. 1973 Another family of distributions for describing and analyzing precipitation data. *J. Appl. Meteorol.* **12**, 275–280.

Mielke, P.W., Jr. and Johnson, E.S. 1973 Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests. *Mont. Weather Rev.* **101**, 701–707.

Mielke, P.W., Jr. and Johnson, E.S. 1974 Some generalized beta distributions of the second kind having desirable application in hydrology and meteorology. *Water Resour. Res.* **10**, 223–226.

Murthy, D.N.P., Xie, M. and Jiang, R. 2004 *Weibull Models*. John Wiley and Sons.

O'Hara, K.T. and Gersonde, R.F. 2004 Stocking control concepts in uneven-aged silviculture. *Forestry* **77**, 131–143.

Rodriguez, R.N. 1977 A guide to the Burr type XII distributions. *Biometrika* **64**, 129–134.

Schmelz, D.V. and Lindsey, A.A. 1965 Size-class structure of old-growth forests in Indiana. *For. Sci.* **11**, 258–264.

Schütz, J.- P. 1997a The Swiss experience: more than one hundred years of experience with a single-tree selection management system in mountainous mixed forests of spruce, fir and beech from an empirically developed utilization in small-scale private forest to an elaborate and original concept of silviculture. In *Proceedings of the IUFRO Interdisciplinary Uneven-Aged management Symposium*. W.H. Emmingham (ed.). Oregon State University, pp. 21–34.

Schütz, J.- P. 1997b Conditions of equilibrium in fully irregular, uneven-aged forests: the state-of-the-art in European plenter forests. In *Proceedings of the IUFRO Interdisciplinary Uneven-Aged Management Symposium*. W.H. Emmingham (ed.). Oregon State University, pp. 455–467.

Tadikamalla, P.R. 1980 A look at the Burr and related distributions. *Int. Stat. Rev.* **48**, 337–344.

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

Zasada, M. and Cieszewski, C.J. 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. Manage.* **204**, 145–158.

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