A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands

Lianjun Zhang, Jeffrey H. Gove, Chuangmin Liu, and William B. Leak

Abstract: The rotated-sigmoid form is a characteristic of old-growth, uneven-aged forest stands caused by past disturbances such as cutting, fire, disease, and insect attacks. The diameter frequency distribution of the rotated-sigmoid form is bimodal with the second rounded peak in the midsized classes, rather than a smooth, steeply descending, monotonic curve. In this study a finite mixture of two Weibull distributions is used to describe the diameter distributions of the rotated-sigmoid, uneven-aged forest stands. Four example stands are selected to demonstrate model fitting and comparison. Compared with a single Weibull or negative exponential function, the finite finite mixture model is the only one that fits the diameter distributions well and produces root mean square error at least four times smaller than the other two. The results show that the finite mixture distribution is a better alternative method for modeling the diameter distribution of the rotated-sigmoid, uneven-aged forest stands.

Résumé : La forme sigmoïde inversée est caractéristique des peuplements inéquiennes de vieille forêt. Elle est causée par les perturbations passées telles la coupe, le feu, les maladies et les attaques d'insectes. La distribution de fréquence diamétrale de forme sigmoïde inversée est bimodale avec le deuxième mode situé dans les classes intermédiaires de diamètre, plutôt qu'une courbe lisse, monotone et avec une forte pente négative. Dans l'étude, un mélange fini de deux distributions de Weibull est utilisé pour décrire la distribution diamétrale des peuplements inéquiennes de forme sigmoïde inversée. Quatre peuplements sont sélectionnés à titre d'exemple pour démontrer l'ajustement et la comparaison du modèle. Par rapport à la fonction simple de Weibull ou à la fonction exponentielle négative, la fonction de mélange fini est la seule qui s'ajuste bien à la distribution diamétrale et qui engendre une erreur standard au moins quatre fois plus petite que celles des deux autres fonctions. Les résultats montrent que la fonction de mélange fini constitue un meilleur choix pour modéliser la distribution diamétrale des peuplements inéquiennes de forme sigmoïde inversée.

[Traduit par la Rédaction]

Introduction

The diameter-class distribution (as given by the plot of tree frequency and diameter) is one of the four interrelated components, i.e., species composition, quality, volume, and diameter distribution, of uneven-aged forest stands (Leak 1964). Meyer (1950) defined a balanced, uneven-aged forest as one where an essentially constant yield can be removed periodically while maintaining the structure and volume of the forest. Leak (1996) pointed out that the balance concept means a diameter distribution and density that will be maintained over time in an unmanaged stand through mortality or a distribution and density that can be maintained through cutting such that the stand structure can be reconstructed again and again with essentially constant yields from each cut. Much effort has been invested in field and simulation research studies on the structural dynamics of uneven-aged forest stands and the notion of a sustainable equilibrium state (e.g., Leak 1964; Adams and Ek 1974; Adams 1976; Alexander and Edminster 1977; Hann and Bare 1979; Lorimer and Frelich 1984; Leak and Gottsacker 1985; Hansen and Nyland 1987; Gove and Fairweather 1992; Baker et al. 1996).

Although the reverse J-shaped form has been traditionally considered an essential feature of balanced, uneven-aged diameter distributions (Meyer 1950; Leak 1965), deviations from this descending, monotonic curve have also been recognized and studied. For example, Goff and West (1975) noted that in old-growth stands with moderate or severe past disturbances, the vigorous and mature trees just entering the upper canopy have relatively higher growth rate and relatively lower mortality rate for a period of years. Subsequently, these trees slow down in growth as they approach maximum age, and become more vulnerable to disease, wind, and other causes of death, increasing the mortality rate in these diameter classes. The resulting diameter frequency distributions are characterized by a broad, rounded peak in the midsized classes, rather than a smooth, steeply descending, monotonic distribution as in the traditional definition of reverse J-shaped structures. Goff and West (1975) suggested the name of "rotated-sigmoid" for this type of uneven-aged diameter distribution, because the log density versus diameter curve can have a steep drop in the smaller classes, fol-

Received January 2, 2001. Accepted May 3, 2001. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on September 8, 2001.

L. Zhang[1] and C. Liu. Faculty of Forestry, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, U.S.A.
J.H. Gove and W.B. Leak. U.S. Department of Agriculture Forest Service, Northeastern Research Station, Durham, NH 03824, U.S.A.

[1]Corresponding author (e-mail: lizhang@syr.edu).

 DOI: 10.1139/cjfr-31-9-1654

lowed by a nearly horizontal trend in the middle portion of the curve, and a sharp decline in the large size classes. They stated that the rotated-sigmoid form is biologically more reasonable as the characteristic equilibrium population structure in smaller or structurally uniform old-growth stands. Lorimer and Frelich (1984) simulated the diameter distributions of old-growth, uneven-aged stands of sugar maple (*Acer saccharum* Marsh.) in upper Michigan and indicated that, while these stands deviate markedly from traditional reversed J-shape and resemble rotated-sigmoid form, this may not be the long-run equilibrium structure for the stand. Schmelz and Lindsey (1965) showed that the rotated-sigmoid form in midwestern old-growth hardwoods was due to early disturbances. Recently Leak (1996) studied the long-term structural change in uneven-aged northern hardwoods. He concluded that the rotated-sigmoid characteristics of these stands 35 years after cutting treatments are caused by disturbances, possibly accompanied by an increase in tolerant softwoods.

The oldest mathematical model used for balanced, uneven-aged diameter distribution is the negative exponential (de Liocourt 1898; Meyer and Stevenson 1943). An important characteristic of this distribution is the constant reduction rate in number of trees from one diameter class to the next with increasing tree size. On a semilogarithmic plot of number of trees versus diameter classes, the negative exponential distribution yields a straight line of negative slope. Hett and Loucks (1976) argued that the constant mortality rate was not realistic and proposed a negative power function as an alternative model, which implies a continuously decreasing rate of attrition as tree size increases. Bailey and Dell (1973) used the Weibull function to fit uneven-aged shortleaf pine (*Pinus echinata* Mill.) – loblolly pine (*Pinus taeda* L.) – mixed hardwood stands, both before and after management, on the Crossett Experimental Forest in Arkansas (data from Davis 1966, p. 216). However, a single Weibull function did not reflect the observed rotated-sigmoid projection (Goff and West 1975). In their paper, Bailey and Dell (1973) also suggested a mixture of Weibull distributions should be considered when a stand has bimodal diameter distribution whose elements were not classified during data collection. Goff and West (1975) concurred that the rotated-sigmoid form of the diameter distribution could be fitted with a mixture of Weibull functions.

Recently, Liu et al. (2001) introduced the use of finite mixture distributions to model the bimodal diameter distribution arising from certain mixed-species forest stands in the Northeast. The Weibull function was assumed as the component probability density function (pdf) in the finite mixture model. They found that the finite mixture model was flexible enough to fit irregular, multimodal, or highly skewed diameter distributions. To date, however, no work that we are aware of has been published based on the suggestions of Bailey and Dell (1973) and Goff and West (1975), i.e., modeling the rotated-sigmoid diameter distribution form of uneven-aged stands using finite mixture models. Therefore, the purpose of this study is to demonstrate the usefulness of the finite mixture of two Weibull distributions using four stands as examples. The model fitting is evaluated and compared with fitting (*i*) a single Weibull pdf and (*ii*) a single negative exponential pdf, to the example stands.

First, however, it will be beneficial to briefly review the concept of the finite mixture distribution; more detailed information can be found in Liu et al. (2001). Assume a finite mixture distribution consists of k individual pdf components; then the distribution of the ith individual component is described by a specific pdf, $f_i(x)$; thus, the overall pdf, $f(x)$, for the mixture distribution can be expressed as

$$f(x) = \sum_{i=1}^{k} \rho_i f_i(x) = \rho_1 f_1(x) + \rho_2 f_2(x) \ldots + \rho_k f_k(x)$$

where ρ_i is the relative abundance of the ith component as a proportion of the total population and must satisfy the constraints $0 \le \rho_i \le 1$ and $\Sigma_{i=1}^{k} \rho_i = 1$. Exposition is restricted here to the simplest case, where $f_1(x), f_2(x), \ldots, f_k(x)$ have a common pdf with different means and, possibly, different variances.

In this study we assume that the component pdf in the finite mixture distribution of a random variable X (i.e., tree diameters) is a three-parameter Weibull function ($X \sim$ Weibull (α, β, γ)) given by

$$[1] \qquad f(x,\theta) = \left(\frac{\gamma}{\beta}\right)\left(\frac{x-\alpha}{\beta}\right)^{\gamma-1} e^{-[(x-\alpha)/\beta]^{\gamma}}$$

$$\infty > x \ge \alpha, \quad \beta > 0, \quad \gamma > 0$$

where $\theta = (\alpha, \beta, \gamma)'$ and α, β, and γ are the location, scale, and shape parameters, respectively. The associated cumulative distribution function (cdf) for the three-parameter Weibull is

$$F(x,\theta) = 1 - e^{-[(x-\alpha)/\beta]^{\gamma}}$$

In a special case, the Weibull becomes the negative exponential when the shape parameter $\gamma = 1$, such that

$$[2] \qquad f(x,\theta) = \left(\frac{1}{\beta}\right) e^{-(x-\alpha)/\beta}$$

Since we only consider a finite mixture distribution with two components following the Weibull distribution in this study, the pdf of the mixture distribution is

$$[3] \qquad f(x, \psi) = \rho f_1(x, \theta_1) + (1 - \rho) f_2(x, \theta_2)$$

where $\psi = (\rho, \theta_1, \theta_2)$ with $\theta_i = (\alpha_i, \beta_i, \gamma_i)'$, and $i = 1$ and 2, and $0 \le \rho \le 1$. Similarly, the corresponding cdf of the mixture distribution is

$$F(x, \psi) = \rho F_1(x, \theta_1) + (1 - \rho) F_2(x, \theta_2)$$

Therefore, this particular mixture distribution is characterized by seven parameters, a location, shape, and scale parameter for each of the two components (i.e., α_1, β_1, γ_1, α_2, β_2, and γ_2) and a parameter (i.e., ρ) characterizing the optimal mixture.

Data and methods

Two published data sets of uneven-aged stands were found in Davis (1966, p. 216) and used by Bailey and Dell (1973). The data were from a shortleaf pine – loblolly pine – hardwood stand on the Crossett Experimental Forest in Arkansas. The two data sets were the same stand before (stand 1) and 10 years after (stand 2) management and evidently followed the rotated-sigmoid form of uneven-aged stands. Because the numbers of trees by diameter class were so low in the actual data, each number was multiplied by 10 before estimating the model parameters as in Bailey and Dell (1973). The third data set was from a loblolly–shortleaf pine stand (stand 3) published in Murphy and Farrar (1981). The data for stand 4 was recently collected from compartment No. 42 on the Bartlett Experimental Forest, New Hampshire. Stand 4 is a managed uneven-aged northern hardwood stand and part of its past history is described by Filip (1978).

The empirical diameter distributions for the four example stands were fitted using the three models described above: (*i*) a single Weibull distribution (eq. 1); (*ii*) a single negative exponential distribution (eq. 2); and (*iii*) a finite mixture of two Weibull distributions (eq. 3). The lower bound of the smallest observed diameter classes (i.e., 3.5 in. for stands 1, 2, and 3 and 5.0 in. for stand 4; 1 in. = 2.54 cm) was used as the estimate for the location parameter α. In this study, the program MIX (Macdonald and Pitcher 1979; Macdonald 1987; Haughton 1997) was used to estimate the parameters for the three models. The key features of this commercial software include fitting the grouped data of mixture distribution by maximum likelihood (ML), for distributions of up to 15 components and 80 class intervals, with component distributions of normal, lognormal, gamma, exponential, or Weibull. The MIX is an interactive, menu-driven program and easy to use. The input is in the form of histogram frequencies with a maximum of 80 bins, and the user needs to provide initial values for the parameters (mixing proportions and the means and standard deviations of the component distributions). The output includes parameter estimates and their standard errors, good-of-fit test, and easy-to-print high-resolution graphics.

The criteria used for model comparison were the root mean square error (RMSE) and χ^2 goodness-of-fit test. Let the model residual (R_j) be defined as the difference between observed and predicted number of trees for each diameter-class in a plot; then:

$$R_j = N_j - \hat{N}_j$$

where N_j and $\hat{N}_j$ are observed and predicted number of trees, respectively, in the jth diameter class. Positive residuals represent underprediction by the model, and negative residuals represent overprediction by the model. Then, the RMSE was computed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{m}(N_j - \hat{N}_j)^2}{m}}$$

where m is the number of diameter classes. The likelihood-ratio χ^2 test was chosen for testing "goodness of fit" (Macdonald and Pitcher 1979) such that

$$\chi^2 = -2\sum_{j=1}^{m} N_j \times \log\left(\frac{\hat{N}_j}{N_j}\right)$$

where χ^2 has $(m - q - 1)$ degrees of freedom and q is the number of estimated parameters.

Results and discussion

Table 1 presents the estimated parameters for each of the three models and four example stands. Note that the estimates of the β parameter for the negative exponential model were close to those obtained by Bailey and Dell (1973) for stands 1 and 2 (given $\alpha = 3.5$ and $\gamma = 1$). The predicted frequencies by diameter classes were obtained from each model for each stand. Then, the predictions from each model were compared with the observed frequencies. The RMSE, χ^2, and p value for the χ^2 test were computed for each model and each stand (Table 2).

It appears that the finite mixture model was the only one to fit all four example stands well based on the p values of the χ^2 tests (Table 2). Neither the Weibull nor exponential models fit the four diameter distributions (the p values of the χ^2 tests were all less than 0.05). The RMSE of the finite mixture model was at least four times smaller than those of the Weibull and exponential models for the stand 1 and was at least six times smaller than those of the Weibull and exponential models for the stand 2. Similarly, the finite mixture model produced the RMSE at least five times smaller than those of the Weibull and exponential models for the stand 4. In an extreme case (stand 3), the RMSE of the finite mixture model was 31 times smaller than that of the Weibull model and 202 times smaller than that of the negative exponential model (Table 2). The observed frequency distribution (histograms) and the three prediction curves for each stand are given in Fig. 1.

The residuals computed across diameter classes for the three models are shown for each stand in Fig. 2. It is evident that the Weibull and negative exponential models overpredict the frequencies for small-sized trees (e.g., 15–25 cm diameter classes) as well as large-sized trees (e.g., >50 cm diameter classes) and underpredict the frequencies for midsized trees (e.g., 30–45 cm diameter classes). It is clear that both Weibull and negative exponential models may adequately fit the smooth, steeply descending, monotonic reverse J-shaped distribution typical of the quintessential balanced uneven-aged diameter distribution of de Liocourt (1898). However, the inherent flexibility of the finite mixture model is a better choice for fitting the rotated-sigmoid diameter distributions that have been shown to be widespread in many managed and unmanaged stands.

Rotated-sigmoid stand structures have been previously presented most often in the form of the semilogarithmic plot (e.g., Goff and West 1975; Leak 1996). Thus, for completeness, and to allow comparison of stand structures used in this paper with those of previous authors, Fig. 3 shows the observed diameter distributions and the predictions from the three models for each of the four stands as a semilog plot (logarithm of trees per hectare against diameter classes). Notice the observed distributions of the four stands show the plateau (nearly level or slightly negative slope) in the middle of the diameter at breast height classes (Leak 1996). Predictions from the finite mixture model closely followed the observed rotated-sigmoid curves, while both Weibull and negative exponential models produced straight lines (Fig. 3) as discussed in the literature.

Conclusion

For several decades now, silviculturalists have recognized that strict adherence to one given q distribution over multiple cutting cycles is unnecessary and can even be detrimental. Indeed, there is evidently a range of sustainable distributions available for a given stand that obviates the need for strict

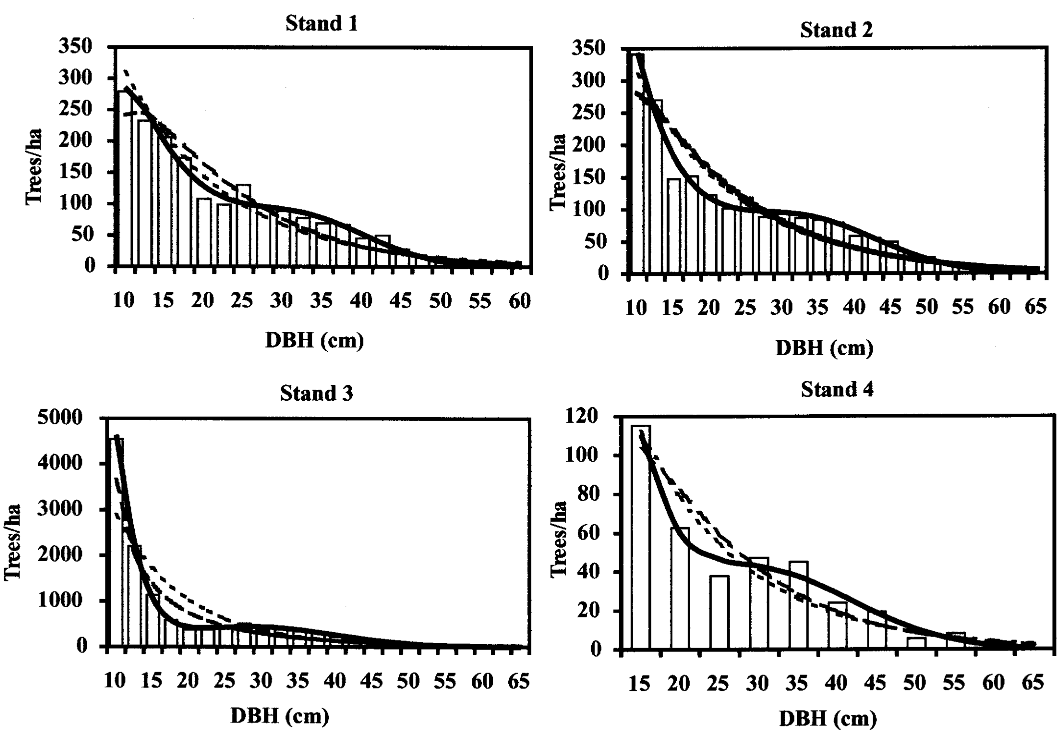
Table 1. Parameter estimates of the three models for the four example stands.

	Weibull			Exponential		Finite mixture						
Example stand	α	β	γ	α	β	ρ	α_1	β_1	γ_1	α_2	β_2	γ_2
Stand 1	8.89	13.94	1.16	8.89	13.26	0.66	8.89	8.08	1.15	8.89	26.67	3.16
Stand 2	8.89	14.88	1.06	8.89	14.58	0.58	8.89	6.93	1.01	8.89	28.42	2.80
Stand 3	8.89	8.19	0.74	8.89	9.80	0.65	8.89	3.12	1.09	8.89	25.17	2.46
Stand 4	12.7	14.40	1.10	12.7	13.87	0.48	12.7	5.18	0.99	12.7	24.49	2.30

Table 2. The root mean square error (RMSE) and χ^2 test of the three models for the four example stands.

	Weibull			Exponential			Finite mixture		
Example stand	RMSE	χ^2	*p*	RMSE	χ^2	*p*	RMSE	χ^2	*p*
Stand 1	90.72	66.80	<0.0001	82.73	85.60	<0.0001	19.54	9.50	0.80
Stand 2	147.95	95.27	<0.0001	118.78	98.46	<0.0001	18.61	8.71	0.92
Stand 3	5224.92	655.9	<0.0001	33 988.5	1209.6	<0.0001	168.08	21.51	0.12
Stand 4	24.43	15.43	0.0308	19.22	16.84	0.0318	3.55	3.73	0.44

Fig. 1. Model comparison for the four example stands. The histogram represents the observed diameter distribution, with the finite mixture model (solid line), a single Weibull function (long dashes), and negative exponential distribution (short dashes).



marking guides. This range of distributions encompasses not only a range of familiar *q* structures but also includes departures from *q* in the form of rotated-sigmoid structures. For example, Adams and Ek (1974) presented a range of stand structures that encompassed both reverse-J and rotated-sigmoid forms. The differences were based on the desired objective to be maximized. Similarly, Martin (1982) found a range of structures conforming to *q* for his guides. In both cases, the diameter distributions derived were not only optimal but also sustainable.

In natural stands and in stands that have been silviculturally manipulated by management activities, the rotated-sigmoid forms of the stand diameter frequency distribution may occur more frequently than previously thought. Such stands do not lend themselves to the traditional, simple reverse J-shaped diameter distribution model that has been used as the primary paradigm for over 100 years. More flexibility is required to fit the rotated-sigmoid stand condition and the finite mixture distribution is a promising alternative method for modeling the diameter distribution of such uneven-aged stands. Advantages of finite mixture distributions include not only this inherent flexibility over unimodal pdfs, but their use also avoids the necessity of classify the two or more components of a rotated-sigmoid or multimodal distribution a priori during data collection. However, the number of component distributions must be decided upon

Fig. 2. Residuals produced by the three models across diameter classes for the four example stands with the finite mixture model (solid line), a single Weibull function (long dashes), and negative exponential distribution (short dashes).

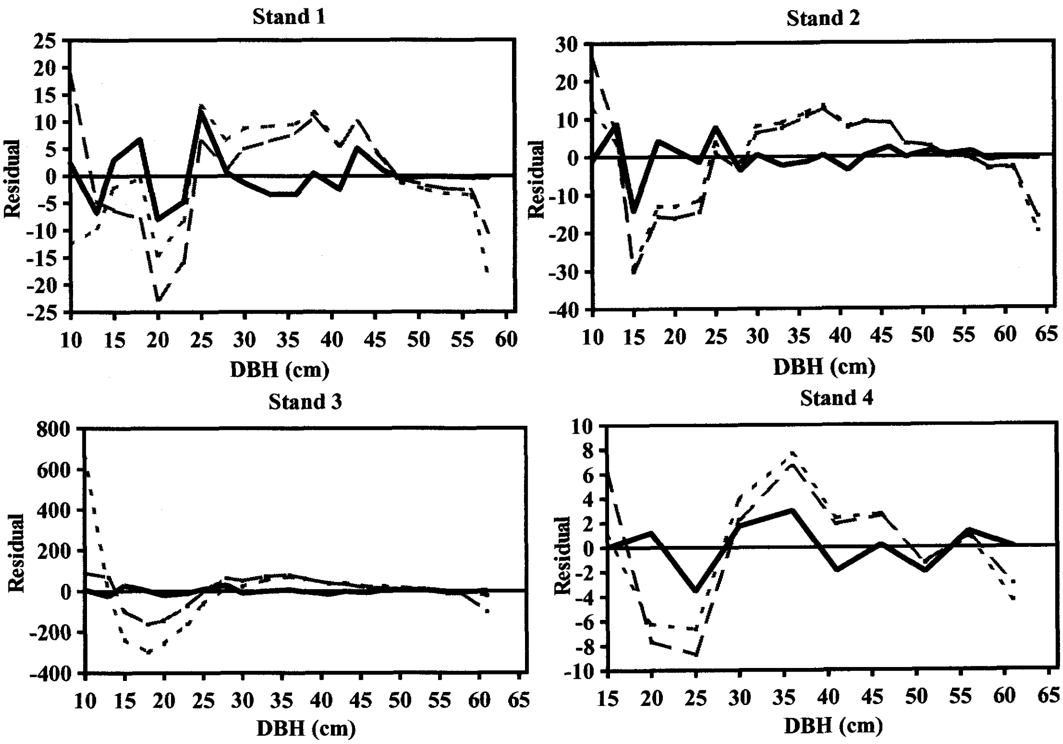
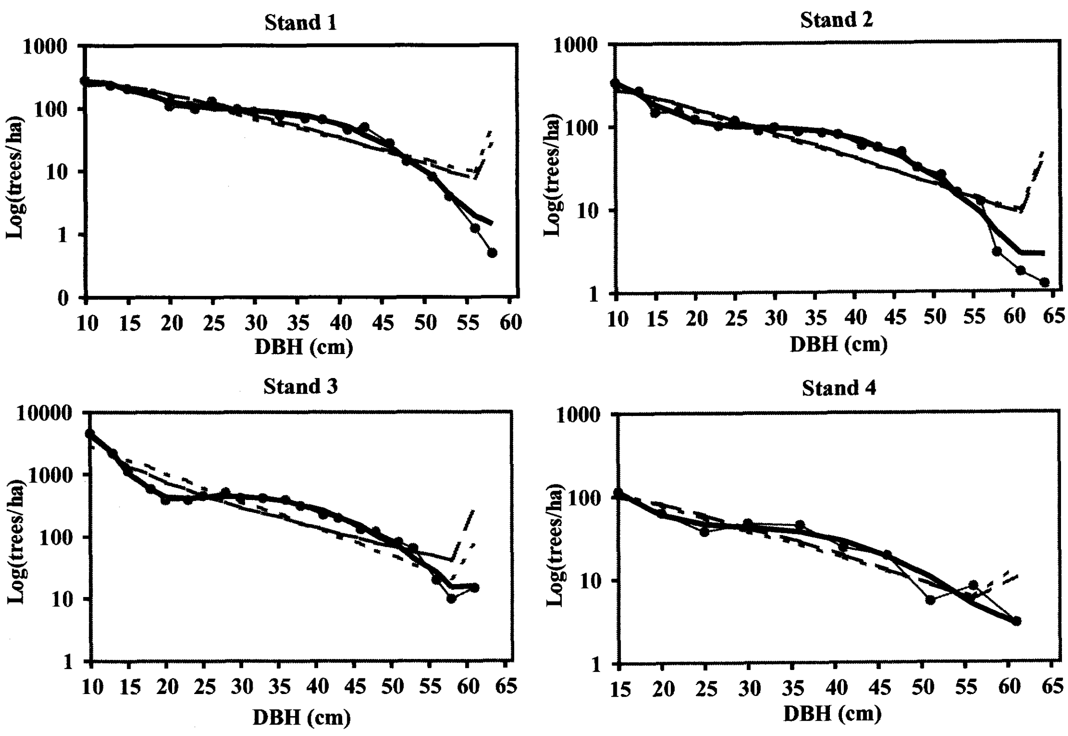


Fig. 3. Plot of logarithm of trees per hectare against diameter classes for the four example stands with the observed diameter distribution (solid line with points), the finite mixture model (thick line), a single Weibull function (long dashes), and negative exponential distribution (short dashes).



when specifying the underlying model to be fitted by ML, but this can be done in an exploratory way by fitting several finite mixtures and settling on the one with the best inherent fit statistics. Commercially available software such as the program used here makes this a relatively simple task for the modeler of stand structure.

Acknowledgments

The authors thank the Associate Editor, Dr. Mark J. Ducey, University of New Hampshire, and an anonymous reviewer for their constructive comments and suggestions.

References

Adams, D.M. 1976. A note on the interdependence of stand structure and best stocking in a selection forest. For. Sci. **22**: 180–184.

Adams, D.M., and Ek, A.R. 1974. Optimizing the management of uneven-aged forest stands. Can. J. For. Res. **4**: 274–287.

Alexander, R.R., and Edminster, C.B. 1977. Regulation and control of cut under uneven-aged management. USDA For. Serv. Res. Pap. RM-182.

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

Baker, J.B., Cain, M.D., Guldin, J.M., Murphy, P.A., and Shelton, M.G. 1996. Uneven-aged silviculture for loblolly and shortleaf pine forest cover types. USDA For. Serv. Gen. Tech. Rep. SO-118.

Davis, K.P. 1966. Forest management: regulation and valuation. 2nd ed. McGraw-Hill, New York.

de Liocourt, F. 1898. De l'amenagement des sapinieres, Soc. For. Franche-Comte Belfort Bull. **6**: 1169–1184.

Filip, S.M. 1978. Impact of beech bark disease on uneven-age management of a northern hardwood forest. USDA For. Serv. Gen. Tech. Rep. NE-45.

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 forest stands: a stochastic approach. For. Sci. **38**: 623–640.

Hann, D.W., and Bare, B.B. 1979. Uneven-aged forest management: state of the art (or science?). USDA For. Serv. Gen. Tech. Rep. INT-50.

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.

Haughton, D. 1997. Packages for estimating finite mixtures: a review. Am. Stat. **51**: 194–205.

Hett, J.M., and Loucks, O.L. 1976. Age structure models of balsam fir and eastern hemlock. J. Ecol. **64**: 1029–1044.

Leak, W.B. 1964. An expression of diameter distribution for unbalanced, uneven-aged stands and forests. For. Sci. **10**: 39–50.

Leak, W.B. 1965. The j-shaped probability distribution. For. Sci. **11**: 405–409.

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

Leak, W.B., and Gottsacker, J.H. 1985. New approaches to uneven-aged management in New England. North. J. Appl. For. **2**: 28–31.

Liu, C., Zhang, L., Davis, C.J., Solomon, D.S., and Gove, J.H. 2001. A finite mixture model for characterizing the diameter distribution of mixed-species forest stands. For. Sci. In press.

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

Macdonald, P.D.M. 1987. Analysis of length–frequency distributions. *In* Age and growth of fish. *Edited by* R.C. Summerfelt and G.E. Hall. Iowa State University Press, Ames. pp. 371–384.

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.

Martin, G.L. 1982. Investment-efficient stocking guides for all-aged northern hardwood forests. University of Wisconsin, Madison, Wis. Agric. Life Sci. Res. Rep. R3129.

Meyer, H.A. 1950. Structure, growth, and drain in balanced, uneven-aged forests. J. For. **52**: 85–92.

Meyer, H.A., and Stevenson, D.D. 1943. The structure and growth of virgin beech–birch–maple–hemlock forests in northern Pennsylvania. J. Agric. Res. **67**: 465–484.

Murphy, P.A., and Farrar, R.M. 1981. A test of the exponential distribution for stand structure definition in uneven-aged loblolly–shortleaf pine stands. USDA For. Serv. Res. Pap. SO-164.

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