Forest Sci., Vol. 26, No. 2, 1980, pp. 291–300

Predicting Mortality with a Weibull Distribution

GREG L. SOMERS
RICHARD G. ODERWALD
WILLIAM R. HARMS
O. GORDON LANGDON

ABSTRACT. The Weibull distribution is presented as a flexible model of survival in even-aged stands. The data necessary for both the complete and censored Weibull distribution and their applications are discussed. A case study is presented using the Weibull distribution to model survival in even-aged stands of loblolly pine from ages 3 to 14. Within the limitations of the data an adequate model was obtained. FOREST SCI. 26:291–300.

ADDITIONAL KEY WORDS. Loblolly pine, mortality prediction.

RELIABLE GROWTH INFORMATION has long been recognized as an essential part of forest management. Growth in a forest stand can be divided into many interrelated components, the most important being survivor growth, ingrowth, and mortality. Short-range prediction of growth may contain a negligible component of ingrowth and mortality, but these components may constitute an important part of growth in any long-range prediction.

A common approach to survival prediction has been to model the percent survival or rate of survival on the basis of the initial number of trees rather than to model the actual number of trees directly. Linear functions have been used by Staebler (1953) to predict percent survival from age, site, and mean diameter and by Lee (1971) to predict mortality rate from stand age or mean diameter. Restrictions often have to be imposed upon the range of the dependent variables to ensure values of percent survival between 0 and 100 (or between 0 and 1). These restrictions may or may not be acceptable depending upon the scope of the project. Instead of predicting percent survival directly Lenhart (1972) predicted the probit of survival, which can take on any value, as a linear function of initial number of trees and age. The probit as used by Lenhart (1972) is the standard normal deviate plus 5 and can be transformed to percent survival by a table of the standard normal deviate.

Many nonlinear functions are implicitly defined on the range of 0 to 1 and can, therefore, be used without any restrictions on the range of the dependent variables. The logistic function used by Hamilton (1974) to predict Lee's (1971) mor-

---

Authors are, respectively, Ph.D. Graduate Assistant and Assistant Professor at Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, and Plant Physiologist and Principal Silviculturist for the U.S.D.A. Forest Service, Southeastern Forest Experiment Station, Forestry Sciences Laboratory, Charleston, SC 29407. This study was partly funded by a U.S. Department of Agriculture program entitled "The Expanded Southern Pine Beetle Research and Applications Program" grant number CSRS 704-15-1. The findings, opinions, and recommendations expressed herein are those of the authors and not necessarily those of the U.S. Department of Agriculture. Manuscript received 26 June 1979.

tality rates as a quadratic function of mean diameter is an example of a nonlinear function defined on the range from 0 to 1. The negative exponential function has been used by Moser (1972) in uneven-aged stands and Hett (1971) who compared the negative exponential and the power function in survival of sugar maple seedlings. Again both the negative exponential and power function are restricted between 0 and 1.

A large subclass of nonlinear functions restricted between 0 and 1 are cumulative distribution functions. The negative exponential used by both Moser (1972) and Hett (1971) belongs to this class as it is the cumulative distribution function of the exponential distribution. The exponential distribution assumes a constant mortality rate over time, an assumption that both Hett (1971) and Moser (1972) concluded may not always be true.

The Weibull distribution provides a logical extension of the exponential distribution having the exponential distribution as a special case. Pinder and others (1978) gave the instantaneous mortality rate as a function of time and the parameters of the Weibull distribution. It can be shown that for the shape parameter equal to 1 the mortality rate is constant, greater than 1 the mortality rate is monotonically decreasing, and for values less than 1 it is monotonically increasing. The Weibull distribution has been used extensively in life testing research in engineering (e.g., Procassini and Romano 1961) and medical research (e.g., Peto and Lee 1973). It has also been used to study survival in animal populations (Pinder and others 1978) and to describe survival in fertilized stands of loblolly pine (Glover and Hool 1979).

The Weibull distribution is a three-parameter function with location parameter "a", a scale parameter "b", shape parameter "c", and probability density function (Johnson and Kotz 1970)

$$f(x) = cb^{-1}[(x - a)/b]^{c-1}\exp[-((x - a)/b)^c]. \tag{1}$$

This function can be used to model the proportion of mortality for each age x. The cumulative proportion of mortality up to age x is represented by the cumulative density function (Johnson and Kotz 1970)

$$F(x) = 1 - \exp[-((x - a)/b)^c]. \tag{2}$$

The proportion of survival at each age x is then 1 minus the cumulative density or

$$G(x) = \exp[-((x - a)/b)^c]. \tag{3}$$

The location parameter "a" indicates the beginning point of the distribution. It can either be estimated from data or assumed to be a constant. For forest tree mortality, "a" can be taken as 0 if mortality is assumed to begin at age 0, the moment of stand inception, or at some age greater than 0 when the stand is considered established and it can be assumed that no new seedlings are germinating to replace those that die. The parameters "b" and "c" can be estimated by maximum likelihood techniques. With "a" equal to 0 these estimators are (Johnson and Kotz 1970)

$$\hat{b} = \left[\left(\sum_{i=1}^{n} f_i\right)^{-1}\left(\sum_{i=1}^{n} f_i x_i^{\hat{c}}\right)\right]^{1/\hat{c}} \tag{4}$$

$$\hat{c} = \left[\left(\sum_{i=1}^{n} f_i x_i^{\hat{c}} \log x_i\right)\left(\sum_{i=1}^{n} f_i x_i^{\hat{c}}\right)^{-1} - \left(\sum_{i=1}^{n} f_i\right)^{-1}\left(\sum_{i=1}^{n} f_i \log x_i\right)\right]^{-1} \tag{5}$$

where n = number of age classes,
x = age, and
f = number of deaths recorded for each age.

One possible use of the Weibull distribution is to allocate a known number of deaths within a given time period. Unless this time period is from "a" to infinity, a truncated distribution should be used, as this would force all the deaths to occur within the specified time period. A much more interesting use of the Weibull distribution would be the prediction of mortality given only the initial number of trees. Unless the age of death is known for all the trees in the data set, a censored Weibull distribution should be used to estimate the parameters. Censored distributions are used widely in life testing studies when the experiment is terminated at either a specified time limit or after a certain number of failures are recorded. The maximum likelihood estimators with "a" equal to 0 are the same as the complete Weibull with the following changes (Johnson and Kotz 1970)

$\sum_{i=1}^{n} f_i x_i^{\hat{c}}$ from equations (4) and (5) becomes $\sum_{i=1}^{n} f_i x_i^{\hat{c}} + \left(N_a - \sum_{i=1}^{n} f_i\right) x_t^{\hat{c}}$ and

$\sum_{i=1}^{n} f_i x_i^{\hat{c}} \log x_i$ from equation (5) becomes $\sum_{i=1}^{n} f_i x_i^{\hat{c}} \log x_i + \left(N_a - \sum_{i=1}^{n} f_i\right) x_t^{\hat{c}} \log x_t$

where N_a = initial number of trees at age a, x_t = terminal age, and the other variables as defined above. If complete mortality data are available then $\sum_{i=1}^{n} f_i$ equals N_a. These parameter estimates can then be used to predict survival as

$$N = G(x)N_a \tag{6}$$

with the variables as previously defined.

A Case Study

This study was performed to demonstrate the use of the censored Weibull distribution in young even-aged natural stands of loblolly pine. Hett (1971) modeled survival in sugar maple seedlings from age 1 to 3 and Smalley and Bailey (1974) modeled survival in loblolly plantations from age 10 to 40, predicting survival from the known initial density. Little quantitative work has been done to model mortality for the ages below 10 years. This study was done to partially fill this gap.

Data for this study were from a previously reported study by Harms and Langdon (1976). Briefly, the study consisted of twenty 0.04-ha (0.1-ac) plots on the Santee Experimental Forest in the Lower Coastal Plain of South Carolina near Charleston, all with site index 32 m (105 ft) at 50 years. The 20 plots were thinned at age 3 to five densities: 2.5, 5, 10, 20, and 40 thousand trees per ha (1, 2, 4, 8, and 16 thousand trees per ac), with four plots at each density level. Henceforth, all plots will be identified by indicating the number of trees per ha to which the plots were initially thinned. Since the data only extend to age 14, a censored Weibull was used to model the survival. The location parameter "a" was taken to be either 0 (referred to as base age 0) or 3 (base age 3). The parameters were estimated from the maximum likelihood estimators given earlier. Parameter estimates for the density totals for both base age 0 and base age 3 are given in Table 1.

To predict survival given only the initial number of trees, linear regression models were developed to relate the values of "b" and "c" to the initial densities. Graphs of "c" for base age 0 are shown in Figure 1. The graph shows the estimates of "c" for each plot and the plots pooled for each density. Notice for density 2.5 only two plots are shown because the other two plots did not have

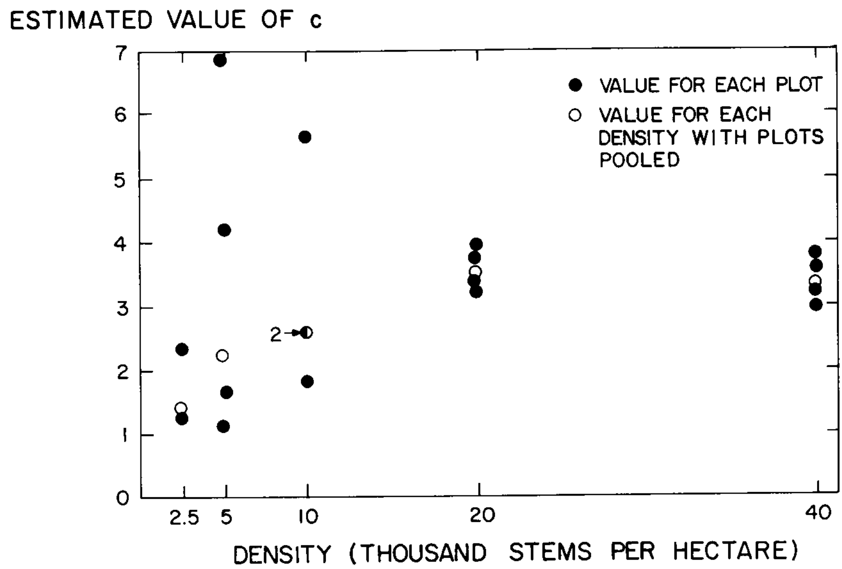

FIGURE 1. Scatter diagram of estimated value of "c" by density for the censored Weibull with "a" equal to 0.

enough deaths to allow estimation. The pooled estimate does include all four plots. For each of the lower densities, the values of "c" show a wide range of variability. This is to be expected because there are relatively few deaths in these densities and the predicted survival is very insensitive to the predicted value of "c". The two higher densities show a much smaller range in the value of "c" indicating a greater sensitivity. The two higher densities also gave approximately

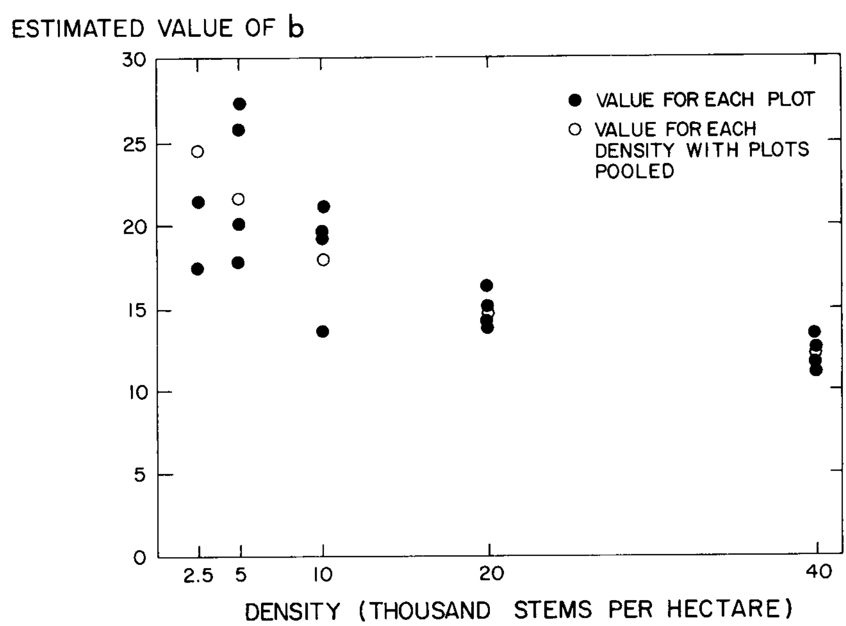

FIGURE 2. Scatter diagram of estimated value of "b" by the density for the censored Weibull with "c" = 2.9561 and "a" equal to 0.

TABLE 1. Parameter estimates for the censored Weibull distribution with all plots in each density pooled.

Density (thousand stems per ha)	$a = 0$		$a = 3$	
	c	b	c	b
2.5	1.3744	47.0331	0.5746	253.5511
5	2.2440	24.9532	0.9902	51.1814
10	2.5628	18.6168	1.1903	25.5648
20	3.4394	14.4249	1.6493	14.4730
40	3.2025	12.1510	1.6503	10.2999

	$a = 0$, $c = 2.9561$	$a = 3$, $c = 1.4663$
	b	b
2.5	24.3352	42.3746
5	21.6018	32.9551
10	17.8130	22.3467
20	14.6810	14.8836
40	12.1581	10.3295

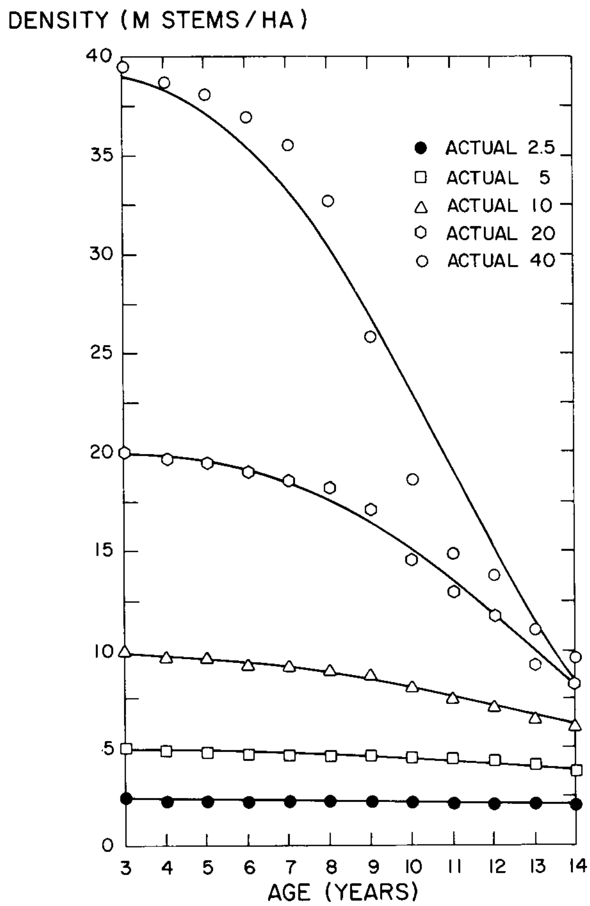

FIGURE 3. Comparison of actual and predicted survival for the censored Weibull with variable "c" and "a" equal to 0.

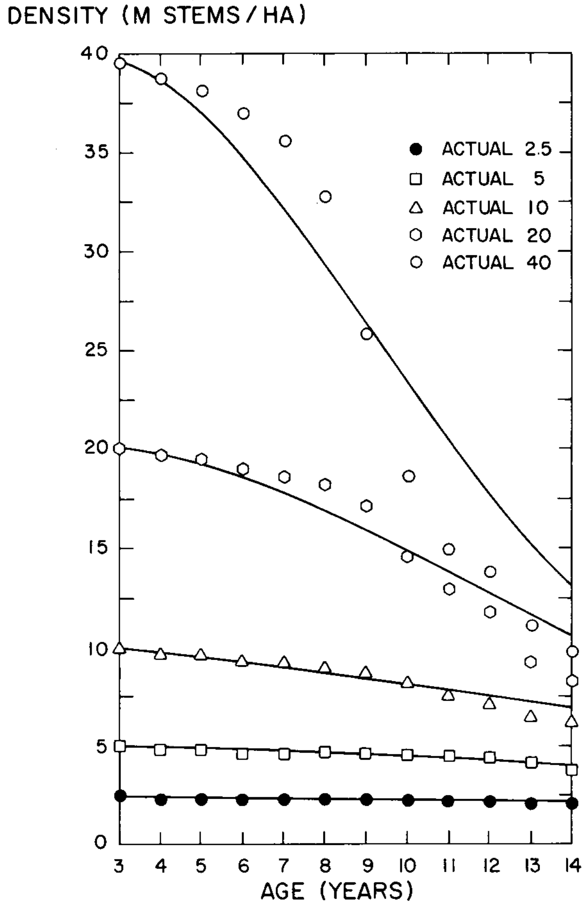

FIGURE 4. Comparison of actual and predicted survival for the censored Weibull with variable "c" and "a" equal to 3.

the same range in values for "c" and since the predicted outcome in the lower densities is insensitive to the value of "c", a common value was assumed for all densities. The common value of "c" was obtained by maximum likelihood estimation for the combined data of all 20 plots. The common values for "c" were 2.9561 for base age 0 and 1.4663 for base age 3. New values for "b" were then estimated using these common values for "c" (Table 1).

Graphs of the values of "b" in relation to density with a common "c" showed good correlation with the initial density for base age 0 (Fig. 2). Several possible equations relating the parameter "b" to initial density were examined. A log-log equation of the form

$$\ln(b) = \beta_0 + \beta_1 \ln N_a \tag{7}$$

where β_0, β_1 = constants to be estimated by regression, and N_a as previously defined, gave the best fit for both base age 0 and 3. The estimates of the parameters for these equations appear in Table 2.

Figures 3–6 are a graphical comparison between the actual and predicted survival for each of the five density totals. Figures 3 and 4 are the censored Weibull with a variable "c" for base age 0 and 3, respectively, and Figures 5 and 6 are with a common "c". The graphs show close fits for each alternative or combi-

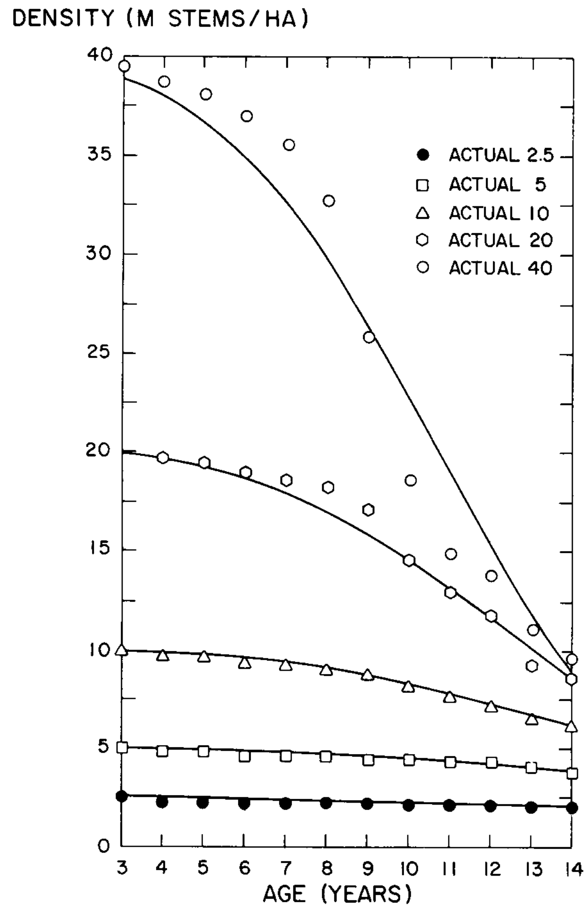

FIGURE 5. Comparison of actual and predicted survival for the censored Weibull with "c" = 1.4663 and "a" equal to 0.

nation of alternatives over all densities, although a close examination shows a slightly better fit of base age 0 than base age 3 for both a common and variable value of "c".

Since the data set was too small to set aside an independent subset to test the predictive ability of the model, two alternative methods of validation not requiring an independent subset of data were used: the Kolmogorov goodness-of-fit test and the PRESS procedure. The Kolmogorov goodness-of-fit test compares the observed distribution with a hypothesized distribution with the test statistic being the maximum absolute difference between the observed and hypothesized cumulative distribution functions (Conover 1971). This test was performed for each

TABLE 2. Least squares estimates of coefficients for linear model used to predict censored Weibull parameter "b" [$\ln(b) = \beta_0 + \beta_1 \ln(\text{density})$].

Base age	β_0	β_1	R^2
0	4.9023	−0.2230	0.6329
3	7.2075	−0.4525	0.6348

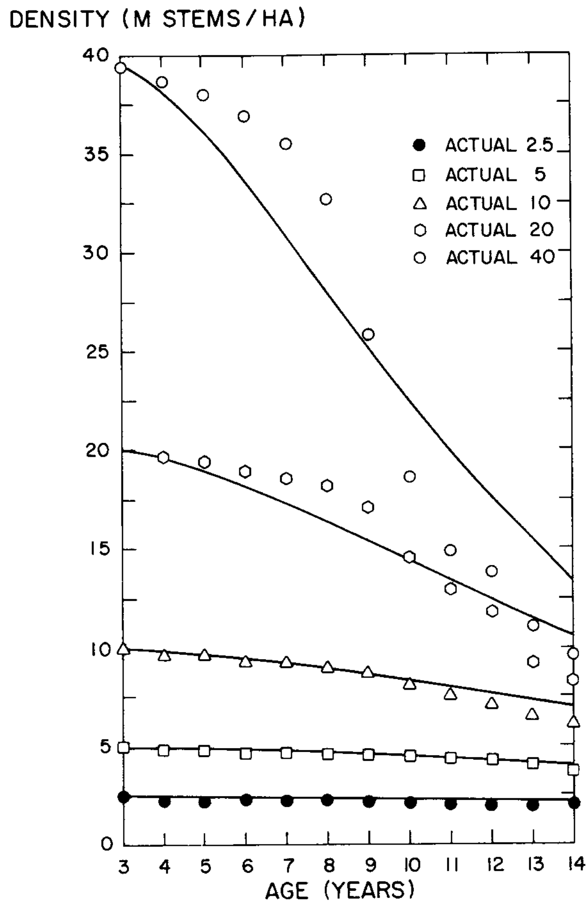

FIGURE 6. Comparison of actual and predicted survival for the censored Weibull with "c" = 2.9561 and "a" equal to 3.

density level separately for each of the alternatives. All the tests were nonsignificant at $\alpha = 0.05$ indicating that all four sets of alternatives gave valid representations of the observed data.

To compare the different alternatives for the Weibull distributions, a technique known as the PRESS procedure was used. PRESS is an acronym for PREdicted Sum of Squares and is described by Allen and others (1973) as a technique to screen variables in regression analysis. PRESS is defined as

$$\text{PRESS} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_{ij})^2 \tag{8}$$

where Y_{ij} = ij^{th} observation on the dependent variable,
$\hat{Y}_{ij}$ = estimator of $E(Y_{ij})$ excluding the i^{th} observation,
n_i = number of age groups, and
m = number of plots.

For this procedure a model is fitted with only $m - 1$ observations, and the observation omitted is then "predicted." This procedure is repeated for each observation. PRESS is then defined as above. Since PRESS simulates prediction,

TABLE 3. Predicted sum of squares (PRESS) for the censored Weibull distribution with variable and common "c" *and* "a" *equal to 0 and 3.*

Density (thousand stems per ha)	Base age 0		Base age 3	
	Variable c	Common c	Variable c	Common c
2.5	207.906	205.150	206.495	198.391
5	445.516	351.783	493.290	364.733
10	3,338.075	3,166.514	3,393.092	3,289.019
20	3,889.308	4,114.914	6,266.481	6,529.347
40	36,101.837	35,040.528	46,995.604	48,369.992

the alternative with the lowest PRESS value can be viewed as having better predictive ability. In this study an observation consisted of the total plot ranging over the entire span of years. A value of PRESS was obtained for each alternative and these values were compared. The values of PRESS for each alternative are shown in Table 3. The best alternative differed for each density and since the best alternative is needed over all the densities, the criterion for choice was

$$W = \sum_{i=1}^{m} \frac{\text{Alt } 1_i - \text{Alt } 2_i}{\text{Alt } 1_i + \text{Alt } 2_i} \tag{9}$$

where $m = 20$, number of plots,
Alt 1_i = value of PRESS for alternative 1, plot i, and
Alt 2_i = value of PRESS for alternative 2, plot i.

If W was positive, then alternative 2 must be smaller on the average than alternative 1 and therefore constitutes the better alternative, and vice versa if W was negative. The results (Table 4) show base age 0 was better than base age 3 for both a common and variable "c" and that a common "c" was better than a variable "c" for both base age 0 and 3. The better set of alternatives was therefore a common value for "c" with "a" equal to zero.

In conclusion, the prediction of survival in young natural stands of loblolly pine was done in two stages. First, the values of "c" and "a" were held constant at 2.9561 and 0, respectively. The value of "b" was then predicted using the regression equation,

$$\ln(b) = 4.9023 - 0.2230 \ln N_a$$

TABLE 4. Calculated values of W *for the censored Weibull distribution to compare alternatives.*

Alternative 1	Alternative 2	Category[1]	W[2]
Base age 0	Base age 3	Variable c	−1.0380
Base age 0	Base age 3	Common c	−0.3150
Variable c	Common c	Base age 0	0.8198
Variable c	Common c	Base age 3	1.5549

[1] Subclass of models under which the alternatives are compared.

[2] $W = \sum_{i=1}^{m} \frac{\text{Alternative } 1_i - \text{Alternative } 2_i}{\text{Alternative } 1_i + \text{Alternative } 2_i}$

where m is the number of plots. Positive values indicate alternative 2 is the better alternative.

where N_a = initial density in trees per ha at age 3. The number of survivors at any age within the range of the data was then predicted from

$$N_i = N_a \exp[-(x_i/b)^c]$$

where N_i = number of trees surviving per ha at age i,
N_a = initial number of trees per ha at age 3,
x_i = age,
"b" = predicted value from regression equation, and
"c" = estimated common value = 2.9561.

Discussion and Conclusions

For a stand level model the Weibull distribution provides a logical extension to the exponential distribution used by Moser (1972) and Hett (1971), and one in which the mortality rate as derived by Pinder and others (1978) is monotonically decreasing for values of "c" greater than 1. This study investigated the use of the Weibull distribution as a model of survival in young stands of natural loblolly pine, an age range of stand development largely unquantified, and found the Weibull gave an adequate fit to the data. The value of "c" was held at a common value for all densities in this data but this was not felt to be generally the case. The value of "b" was related to initial number of trees alone as this was the only real difference between plots. The values of these parameters may be related, it was felt, to other initial stand characteristics such as site and initial spacing but due to the limitations in the data, this could not be tested.

Literature Cited

Allen, D. M., J. O. Street, and A. Mazzoleni. 1973. PRESS used to study relationsips between component cardiac weights and electrocardiographic patterns. *In* Proceedings of Univ of Kentucky Conf on regression with a large number of predictor variables (W. O. Thompson and F. B. Cady, eds), p 124–135.

Conover, W. J. 1971. Practical nonparameteric statistics. John Wiley and Sons, New York. 462 p.

Glover, G. R., and J. N. Hool. 1979. A basal area ratio predictor of loblolly pine plantation mortality. Forest Sci 25:275–282.

Hamilton, D. A. 1974. Event probabilities estimated by regression. USDA Forest Serv Res Pap INT-152, 18 p.

Harms, W. R., and O. G. Langdon. 1976. Development of loblolly pine in dense stands. Forest Sci 22:331–337.

Hett, J. M. 1971. A dynamic analysis of age in sugar maple seedlings. Ecology 52:1071–1074.

Johnson, M. L., and S. Kotz. 1970. Continuous univariate distributions 1. Houghton Mifflin, Boston. 399 p.

Lee, Y. J. 1971. Predicting mortality for even-aged stands of loblolly pine. For Chron 47:29–32.

Lenhart, J. D. 1972. Predicting survival of unthinned, old-field loblolly pine plantations. J For 70:754–755.

Moser, J. W. 1972. Dynamics of an uneven-aged forest stand. Forest Sci 18:184–191.

Peto, R., and P. Lee. 1973. Weibull distributions for continuous-carcinogenesis experiments. Biometrics 29:247–270.

Pinder, J. E, J. F. Wiener, and M. H. Smith. 1978. The Weibull distribution: a new method of summarizing survivorship data. Ecology 59:175–179.

Procassini, A. A., and A. Romano. 1961. Transistor reliability estimates improved with Weibull distribution function. Motorola Engineering Bull 9, No 2, p 16–18.

Smalley, G. W., and R. L. Bailey. 1974. Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama, and Georgia Highlands. USDA Forest Serv Res Pap SO-96, 81 p.

Staebler, G. R. 1953. Mortality estimation in fully-stocked stands of young Douglas-fir. USDA Forest Serv, Pac Northwest Forest and Range Exp Stn Res Pap 4, 8 p.