A Survival Model for Individual Shortleaf Pine Trees in Even-Aged Natural Stands

Thomas B. Lynch, Michael M. Huebschmann, and Paul A. Murphy

ABSTRACT.-A model was developed that predicts the probability of survival for individual shortleaf pine (*Pinus echinata* Mill.) trees growing in even-aged natural stands. Data for model development were obtained from the first two measurements of permanently established plots located in naturally occurring shortleaf pine forests on the Ouachita and Ozark National Forests in western Arkansas and eastern Oklahoma. The logistic function was used to model survival probability. Parameters were estimated by using logistic regression in which the dependent variable was "1" for trees alive during both inventories and "0" for trees that died prior to the second inventory. Examination of several combinations of independent variables (representing tree size, relative position of the tree in the stand, stand density, and stand age) yielded the following model:

$$POS = (1 + \exp(-(b_0 + b_1/DR + b_2SBA + b_3DH)))^{-1}$$

where *POS* is annual probability of survival, *DR* is the ratio of quadratic mean diameter to individual tree dbh, *SBA* is stand basal area in square feet per acre, *DH* is average height of dominant and codominant trees in feet, b_0, b_1, b_2, b_3 are parameter estimates. A chi-square evaluation was performed to test model performance. This tree survival model is being used to estimate probability of individual tree survival in a distance-independent individual tree simulator for shortleaf pine.

A survival model is an integral part of the ShortLeaf Pine Stand Simulator (SLPSS) (Huebschmann *et al.* 1998), an individual tree model that has been developed for even-aged natural shortleaf pine forests. Other components of the model include an individual tree basal area growth model (Hitch 1994) and a compatible height prediction and projection system for shortleaf pine trees in even-aged natural stands (Lynch and Murphy 1995). Hamilton (1974) proposed that the following logistic equation be used to model probability of individual tree mortality:

$$P = \left[1 + \exp\left(-\left(\beta_0 + \sum_{i=1}^{m} \beta_i x_i\right)\right)\right]^{-1} + \varepsilon \qquad (1)$$

where P is probability of mortality, x_i are independent variables, β_0 and β_i are parameters, and ε is an error term with mean zero. Hamilton (1974) suggested that parameters in the model be estimated by logistic regression in which the dependent variable is assigned "0" for trees surviving a designated time interval and "1" for trees dying in the interval. Modelers who wish to predict probability of mortality designate mortality trees as "1" and surviving trees as "0," while modelers who wish to predict probability of survival designate surviving trees as "1" and mortality trees as "0." The logistic regression procedure would minimize the sum of squared errors between the dependent variable and model predictions. Predictions are constrained by the logistic model form to range between 1 and 0 so that they may be interpreted as mortality probabilities. Hamilton modified the "RISK" program developed by Walker and Duncan (1967) to accomplish parameter estimation. Hamilton and Edwards (1976) applied this methodology to mortality modeling for several tree species in northern Idaho. They used the SCREEN program (Hamilton and Wendt 1975) to identify potentially important independent variables. Among the variables they used to predict probability of mortality were dbh, height, age, crown class, and basal area per acre. Many other individual tree growth and yield systems use this or a very similar method to estimate probability of mortality or survival. Amateis *et al.* (1989) and Daniels and Burkhart (1975) fitted parameters to a survival model by assigning "1" as the dependent variable for surviving trees and "0" as the dependent variable for mortality trees, but they did not use the logistic model form as the prediction equation.

The authors are Associate Professor (telephone 405-744-5447, FAX 405-744-9693 email: tlynch@okway.okstate.edu) and Senior Research Specialist, Department of Forestry, OSU, Ag Hall Room 008C, Stillwater, OK 74078-60 13, USA; and Research Forester, Southern Research Station, University of Arkansas at Monticello, P.O. Box 35 16, UAM Monticello, AR 71656-35 16, USA.

Most growth and yield information on naturally occurring shortleaf pine stands has been based on stand-level tables or equations that do not include models for probability of survival of individual trees. USDA Misc. Publ. 50 (USDA Forest Service 1929) contains normal yield tables for fully stocked natural stands of shortleaf pine. Schumacher and Coile (1960) provide yield tables for shortleaf pine based on 74 well-stocked temporary plots in the North Carolina Piedmont, which are also based on a normal stocking concept. Brinkman (1967) developed shorteaf pine stand volume equations based on periodic remeasurements in 57 stands. Data obtained from the Southern Research Station Forest Inventory and Analysis unit are the basis of stand-level growth and yield models for natural shortleaf pine stands developed by Murphy and Beltz (198 1) and Murphy (1982), which can be applied to a variety of stand densities. Lynch *et al.* (1991) developed stand volume equations for shortleaf pine. Murphy and Farrar (1985) developed a stand-level growth and yield system for uneven-aged shortleaf pine forests. Growth and yield information for shorteaf pine stands can be obtained from the central states version of TWIGS (Miner *et al.* 1989), an individual tree simulator that uses equations developed with inventory data from Indiana, Illinois, and Missouri. Central states TWIGS does contain individual tree mortality models. An individual tree simulation system for forests in the south that provides information on shortleaf pine has been developed by Bolton and Meldahl (1990).

DATA

Prior to 1985, most sources of growth and yield information for naturally occurring shortleaf pine stands were based on fully stocked plots representing a limited range of densities (normal stocking) or on inventory data that primarily represented unmanaged stands. Therefore, in 1985, a cooperative study between the USDA Forest Service Southern Research Station at Monticello, Arkansas, and the Department of Forestry at Oklahoma State University (hereafter termed the OSU-Forest Service Co-op study) was initiated to establish growth and yield plots in even-aged natural shortleaf pine stands representing a range of ages, densities, and site qualities. These data were used to develop a shortleaf pine survival model. The study plan consisted of four levels of age (20, 40, 60, and 80 years), four levels of base age 50 site index (<56, 60, 70, and >75 feet), and four levels of basal area per acre (30, 60, 90, and 120 square feet per acre). The original plan specified establishment of three plots in each combination of age, site index, and density for a total of 192 plots. Only 19 1 of these were established. Some were lost due to a wind storm and failure to execute thinning treatments at some locations, so 183 plots remained at the second measurement. Plots were established during 1985 to 1988 and remeasured from 1990 to 1992, with intervals of 4 or 5 years between measurements.

A 33-foot circular buffer strip surrounds a 0.2-acre measurement plot at each plot location. The measurement plot and buffer strip were thinned from below to the same residual basal area per acre level and received the same chemical herbicide control treatment for competing vegetation. Within the 0.2-acre plot all shortleaf pine trees 1 inch in dbh or larger were numbered, marked at the dbh measurement point, and the dbh, crown class, distance to plot center, and azimuth from plot center were recorded. A representative sample of trees within each dbh class on the plot was selected for measurement of total height and height to live crown. Each dominant or codominant tree on every plot was cored with an increment borer to determine age. Site index curves developed by Graney and Burkhart (1973) were used to estimate the base age 50 year site index on each plot.

Thirty-four plots were available from a shortleaf pine thinning study established by Frank Freese during 1963-1964. Each plot consisted of a 0.2-acre measurement plot, and each was originally thinned to a basal area of 45, 65, 85, 105, or 125 square feet per acre. By combining plots from the OSU-Forest Service Co-op study with plots from the Freese study, a total of 2 17 plots were available for development of a mortality model for individual shortleaf pine trees. All plots in the OSU-Forest Service Co-op study were established on the Ouachita or Ozark National Forests. Although some of the Freese study plots were established on industrial forest ownerships, all remaining plots are located on the Ouachita National Forest. Plot locations range from the Boston Mountains near Russellville, Arkansas, in the Ozark National Forest to the West Gulf Coastal Plain in the Ouachita National Forest near Broken Bow, Oklahoma. Most plots are located in the Ouachita Mountain region of eastern Oklahoma and western Arkansas. Table 1 provides a summary of the plot-level data used for analysis. A summary of the data at the individual tree level is given in table 2.

MODEL DEVELOPMENT

Data from individual shortleaf pine trees were used to fit parameters to a logistic survival model having the general form of equation 1. To assess variables potentially useful for prediction of survival probabilities, the SCREEN program (Hamilton and Wendt 1975) was run on the shortleaf pine data set. Among variables considered were dbh, quadratic mean dbh (QMD), ratio of QMD to dbh, site index, crown ratio, average height of dominant and codominant trees, age, plot basal area per acre, and proportion of basal area in trees as large or larger than the tree for which survival is to be predicted.

Table 1.-*Summary statistics for 217 shortleafpine growth study plots*

Variable	Mean	Std. dev.	Minimum	Maximum
Initial basal area (ft^2/acre)	81.88	35.07	27.32	174.24
Final basal area (ft^2/acre)	93.04	37.91	14.87	180.00
Mid-period basal area (ft^2/acre)	87.46	36.24	22.53	177.12
Initial age (years)	55.2	18.7	18	93
Final age (years)	60.3	18.7	23	99
Site index, height at age 50 (ft)	62.2	10.7	38.9	87.1

Table 2.—*Mean, standard deviation, minimum and maximum for dbh, quadratic mean dbh (QMD), and ratio of QMD to dbh from 9,238 individual shortleafpine trees used to fit parameters to a logistic survival model*

Variable	Mean	Std. dev.	Minimum	Maximum
QMD (in.)	7.70	3.31	3.10	17.90
Dbh (in)	7.51	3.73	1.1	24.4
Ratio QMD to dbh	1.10	0.33	0.44	4.42

For fitting parameters to equation 1, the dependent variable for each individual tree was coded as "1" if the tree survived the measurement interval and "0" if the tree did not survive the measurement interval. It was desired to obtain a model that would predict survival on an annual basis for use in an individual tree growth model that uses annual steps. Since plot measurement intervals in the shortleaf pine data were either 4 or 5 years, the following transformation discussed by Hamilton and Edwards (1976) was used to fit parameters on an annual basis:

$$P_t = \left[+ \exp(-(\beta_0 + \sum_{i=1}^{m} \beta_i x_i)) \right]^{-t} + \varepsilon \qquad (2)$$

where t is the length of the measurement period in years, P_t is the probability of survival for time period t, and other symbols are as defined for equation 1. Nonlinear weighted regression was used to fit parameters. The weight used was 1/(*P*(I-P)), which is the inverse of the variance of a Bernoulli (1 or 0) random variable. Trials were run using both the RISK program (Hamilton 1974) and SAS PROC NLIN (SAS Institute 1988) with similar results. Various models with various combinations of independent variables were evaluated. The model resulting from this evaluation process is:

$$POS = (1 + \exp(-(b_0 + b_1/DR + b_2SBA + b_3DH)))^{-1} \qquad (3)$$

where *POS* is annual probability of survival, *DR* is the ratio of quadratic mean diameter to individual tree dbh, *SBA* is stand basal area in square feet per acre, *DH* is average height of dominant and codominant trees in feet, and b_0, b_1, b_2, b_3 are parameter estimates. Initially, the data were divided into two subsets, one for fitting parameters and a randomly selected independent data set used for model evaluation. After selection of model 3, the two subsets were combined and used to obtain final parameter estimates. These estimated parameter values, together with their standard errors, are given in table 3. Ratios of parameter estimates to standard errors show values of t=12.2 for b_1 (1/*DR*) and t=4.9 for b_2 (*SBA*), both of which are significant at the α=0.05 level. It might be expected that for a given QMD, trees having a smaller dbh resulting in a larger value of the ratio *DR* = QMD/dbh would have a smaller chance of survival, while trees in stands having a lower basal area per acre would have a greater chance of survival. The signs of the coefficients b_1 (1/*DR*) and b_2 (*SBA*) in the context of equation 3 are consistent with these expectations. The average height of dominants and codominants *DH* was calculated by evaluating the site index curves of Graney and Burkhart (1973) at plot age and site index. The coefficient associated with *DH*, b_3, has t=- 1.4, which would be significant

Table 3.-*Parameter estimates for a logistic model of survival probability for individual shortleaf pine trees*

Coefficient	Estimate	Standard error
b_0	2.912370652	0.44483029972
b_1	4.789284600	0.39415378476
b_2	-0.015129972	0.00310011084
b_3	-0.006680302	0.00465072624

at the α=0.16 level but is not significant at the α =0.05 level. Still,.it was decided that significance was high enough to be beneficial for predictions. *DH* is a function of site quality and age, so it reflects the influence of both of these variables.

MODEL EVALUATION

A chi-square test was proposed by Hamilton and Edwards (1976) for evaluation of mortality models based on logistic regression. Hamilton (1974) showed why the mean square error, often used to evaluate regression models, is not appropriate for evaluation of logistic regression with a Bernoulli (0 or 1) dependent variable. Neter *et al.* (1989) recommend the following chi-square statistic for evaluation of logistic regression models:

(4)

where c is the number of categories, O_{jk} is the prediction for "success" ("1") in category *j* when k=1, and for "failure" ("0") in category j when k=0. When applied to individual tree survival models, a natural categorization makes c the number of dbh classes, with O_{j1} the number of surviving trees in class j and O_{j0} the number of mortality trees in dbh class j. E_{j1} is obtained by summing the individual predicted probabilities of survival for trees in dbh class j, while E_{j0} is found by summing the individual predicted probabilities of mortality for trees in dbh class j.

Neter *et al.* (1989) recommend comparison of X^2 to a tabulated value of the chi-square distribution χ^2 with c − 2 degrees of freedom. Hosmer and Lemeshow (1980) showed through simulation studies that two degrees of freedom should be subtracted from the number of categories when evaluating the logistic regression model with chi-square computed on the data set used for parameter fitting (see also Hosmer and Lemeshow 1989).

Comparisons between observed and predicted number of surviving and mortality trees are given in table 4. Predicted survival is obtained by summation of equation 3 for all trees within the indicated dbh class. Probability for mortality can be obtained by subtracting the survival prediction from 1. These mortality predictions are used to obtain predicted mortality for each dbh class in table 4. Since the plot data are based on measurement intervals of 4 and 5 years, survival probabilities were obtained by raising equation 3 to the power of the number of years in the measurement period, as indicated in equation 2. Thus, the values in table 4 do not represent annual survival or mortality rates; rather, they are a mixture of 4- and 5-year rates. Chi-square values for survival and mortality corresponding to each dbh class are also given in table 4. The computed chi-square for model evaluation according to equation 4 can be obtained by summation of chi-square components corresponding to each dbh class in table 4:

$$X^2 = 2.0364 + 57.1077 = 59.1441$$

Table 4.-*Predicted and actual survival and mortality by dbh class with chi-square contribution, based on measurement intervals of 4 and 5 years*

Dbh class	No. trees	No. surv. trees	Predicted survivors	Chi-square	No.mort. trees	Predicted mortality	Chi-square
1	24	19	19.09	0.0004	5	4.91	0.0016
2	443	391	398.79	0.1520	52	44.21	1.3714
3	902	872	856.66	0.2746	30	45.34	5.1882
4	1,123	1,085	1,083.89	0.0011	38	39.11	0.0318
5	910	849	877.28	0.9117	61	32.72	24.4436
6	856	822	826.44	0.0239	34	29.56	0.6669
7	787	760	759.80	0.0001	27	27.20	0.0015
8	812	804	786.29	0.3989	8	25.71	12.1987
9	712	701	692.16	0.1129	11	19.84	3.9383
10	647	631	632.20	0.0023	16	14.80	0.0968
11	554	547	543.52	0.0222	7	10.48	1.1526
12	495	491	487.19	0.0298	4	7.81	1.8604
13	354	351	349.21	0.0092	3	4.79	0.6686
14	237	231	234.37	0.0485	6	2.63	4.3281
>14.5	382	374	378.30	0.0487	8	3.70	1.1592
Totals	9,238	8,928	8,925.18	2.0364	310	312.82	57.1077

This should be compared to a table value of the chi-square distribution χ^2 with 15 2 = 13 degrees of freedom. This table value for the a = 0.05 level of significance is 22.36. Since the computed value $X^2 = 59.1441 > 22.36 = \chi^2$, the hypothesis that the model fits the data should be rejected.

Inspection of table 4 shows that the model seems to tit survival rates well. Summation of individual chi-square components for survival over all dbh classes is 2.0364, which is quite low. Rejection of the model is due to the mortality results, especially in the 5-inch and 8-inch dbh classes. The chi-square component for the 5-inch class is 24.4436, while the chi-square component for the 8-inch class is 12.1987. Chi-square components for mortality in other dbh classes is quite low. Inspection of mortality in the 5-inch class indicates that the rate is much higher than in the adjoining classes. The reason for this is not known, but it would appear difficult to fit the mortality trend through the 5-inch class in a manner consistent with results from other classes.

In general, mortality rates are quite low in these data, since they come from thinned stands. Except for Freese thinning study data, most plots were thinned from below at the beginning of the measurement interval, so that most suppressed trees where removed. The Freese study plots also had a history of thinning, but the measurement interval used did not occur immediately after thinning treatments. Only 3 10 of the 9,238 trees present at the first measurement died, leaving 8,928 at the second measurement. An overall annual mortality rate of between 0.7 and 0.8 percent would be indicated by compound interest formulas, although an exact computation is complicated by the fact that measurement intervals were not the same for all plots.

SUMMARY AND CONCLUSIONS

A logistic model for prediction of survival probabilities for individual shortleaf pine trees growing in even-aged natural stands was developed on the basis of data from remeasured plots located in eastern Oklahoma and western Arkansas. Although a chi-square test failed to accept the model, this was primarily due to results in only 2 of 15 dbh classes examined. Two of the independent variables used in the model, ratio of QMD to dbh and basal area per acre, are highly significant on the basis of the ratio between parameter estimates and associated standard errors. In all but the 5- and 8-inch classes, predictions from the model correspond to observed data reasonably well. Trials with models having a number of other combinations of independent variables failed to produce better survival estimates. Since this model is a better alternative than a constant mortality rate for all trees, it was selected for use in the ShortLeaf Pine Stand Simulator (SLPSS) (Huebschmann *et* al. 1998), an individual tree simulator for naturally occurring shortleaf pine stands.

ACKNOWLEDGMENTS

The following people reviewed this manuscript: Lawrence R. Gering and Robert F. Wittwer, Department of Forestry, Oklahoma State University, Stillwater, OK, USA.

LITERATURE CITED

Amateis, R.L.; Burkhart, H.E.; Walsh, T.A. 1989. Diameter increment and survival equations for loblolly pine trees growing in thinned and unthinned plantations on cutover, site-prepared lands. Southern Journal of Applied Forestry. 13(4): 170- 174.

Bolton, R.K.; Meldahl, R.S. 1990. User's guide to a multipurpose forest projection system for southern forests. Bull. 604. Auburn University, Alabama: Alabama Agricultural Experiment Station. 15 p.

Brinkman, K.A. 1967. Stand volume equations for shortleaf pine in Missouri. Res. Note NC-24. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 4 p.

Daniels, R.F.; Burkhart, H.E. 1975. Simulation of individual tree growth and stand development in managed loblolly pine plantations, Publ. FWS-5-75. Virginia Polytechnical Institute and State University, Division of Forestry and Wildlife Research. 69 p.

Graney, D.L.; Burkhart, H.E. 1973. Polymorphic site index curves for shortleaf pine in the Ouachita mountains, Res. Pap. SO-85. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 12 p.

Hamilton, D.A. 1974. Event probabilities estimated by regression. Res. Pap. INT-152. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 18 p.

Hamilton, D.A.; Edwards, B.M. 1976. Modeling the probability of individual tree mortality. Res. Pap. INT- 185. Ogden, UT: U.S. Department of Agriculture. Forest Service, Intermountain Forest Range Experiment Station. 22 p.

Hamilton, D.A.; Wendt, D.L.R. 1975. SCREEN: a computer program to identify predictors of dichotomous dependent variables. Gen. Tech. Rep. INT-22. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 20 p.

Hitch, K.L. 1994. A distance-independent individual tree basal area growth model for natural, even-aged stands of shortleaf pine in eastern Oklahoma and western Arkansas. Stillwater, OK Oklahoma State University. 78 p. M.S. thesis.

Hosmer, D.W.; Lemeshow, S. 1980. A goodness-of-fit test for the multiple logistic regression model. Communications in Statistics. A10: 1043-1069.

Hosmer, D.W.; Lemeshow, S. 1989. Applied logistic regression. New York: John Wiley and Sons. 307 p.

Huebschmann, M.M.; Lynch, T.B.; Murphy, P.A. 1998. Shortleaf pine stand simulator, an even-aged natural shortleaf pine growth and yield model: user's manual. Res. Rep. P-967. Stillwater, OK: Oklahoma State University, Agricultural Experiment Station. 25 p.

Lynch, T.B.; Murphy, P.A. 1995. A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. Forest Science. 41(1): 194-209.

Lynch, T.B.; Murphy, P.A.; Lawson, E.R. 1991. Stand volume equations for managed natural even-aged shortleaf pine in eastern Oklahoma and western Arkansas. Res. Rep. P-92 1. Stillwater, OK: Oklahoma State University Agricultural Experiment Station. 12 p.

Miner, C.L.; Walters, N.R.; Belli, M.L. 1989. A technical guide to TWIGS for the North Central U.S. Gen. Tech. Rep. NC-125. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 105 p.

Murphy, P.A. 1982. Sawtimber growth and yield for natural even-aged stands of shortleaf pine in the West Gulf. Res. Pap. SO-181. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 13 p.

Murphy, P.A.; Beltz, R.C. 198 1. Growth and yield of shot-deaf pine in the West Gulf region. Res. Pap. SO-169. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 15 p.

Murphy, P.A.; Farrar, R.M., Jr. 1985. Growth and yield of uneven-aged shot-deaf pine in the Interior Highlands. Res. Pap. SO-218. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 11 p.

Neter, J.; Wasserman, W.; Kutner, M.H. 1989. Applied linear regression models, 2d ed. Homewood, IL: Irwin. 667 p.

SAS Institute Inc. 1988. SAS/STAT™ user's guide, release 6.03 Ed. Cary, NC: SAS Institute Inc. 1028 p.

Schumacher, F.X.; Coile, T.S. 1960. Growth and yield of natural stands of the southern pines. Durham, NC: T.S. Coile, Inc. 115 p.

USDA Forest Service. 1929 (rev. 1976). Volume, yield and stand tables for second-growth southern pines. Misc. Publ. 50. Washington, DC: U.S. Department of Agriculture. 202 p.

Walker, S.H.; Duncan, D.B. 1967. Estimation of the probability of an event as a function of several independent variables. Biometrika. 54: 167- 179.