Maximum Likelihood Estimation for Predicting the Probability of Obtaining Variable Shortleaf Pine Regeneration Densities

Thomas B. Lynch, Jean Nkouka, Michael M. Huebschmann, and James M. Guldin

ABSTRACT. A logistic equation is the basis for a model that predicts the probability of obtaining regeneration at specified densities. The density of regeneration (trees/ha) for which an estimate of probability is desired can be specified by means of independent variables in the model. When estimating parameters, the dependent variable is set to 1 if the regeneration density (trees/ha) on the plot is greater than the specified density (trees/ha), otherwise 0. Since it is desired to estimate parameters for a range of probability densities, traditional estimation techniques for logistic models cannot be used. Multiple regeneration densities require a multinomial distribution, for which maximum likelihood estimates are obtained. Counts of shortleaf pine regeneration taken 9-10 years after thinning on 182 plots established in naturally occurring shortleaf pine forests are used to estimate parameters. FOR. SCI. 49(4):577–584.

Key Words: Natural regeneration, *Pinus echinata*, logistic regresson.

NATURAL REPRODUCTION OF SHORTLEAF PINE (*Pinus echinata* Mill.) is essential for continued management of naturally occurring shortleaf pine forests. Such forests are common in eastern Oklahoma and western Arkansas on the USDA Forest Service Ouachita and Ozark National Forests as well as on land in nonindustrial private ownership and some industrial holdings. Restoration of the shortleaf pine-bluestem grass forest type, thought to be the presettlement forest type in parts of the Ouachita mountain region, has contributed to sustained interest in management of naturally occurring shortleaf pine on the Ouachita National Forest (Huebschmann 2000). Lawson (1986) stated that about one-half to one-third of primarily shortleaf or mixed loblolly (*Pinus taeda* L.)-shortleaf forests were expected to be naturally regenerated.

Shelton (1997) studied understory development in pine and pine-hardwood shelterwood forests in the Ouachita Mountains during a 3 yr period after a shelterwood cut. Sound pine seed production monitored with seed traps ranged from marginal to good during the 3 yr period after cutting (96,371/ha in 1990, 69,190/ha in 1991, and 383,013/ha in 1992). Pine seedlings 3 yr after the shelterwood cut averaged 3,830 stems/ha, with no significant differences between overstory or hardwood control treatments (Shelton 1997). The minimum density limit for natural regeneration of even-aged shortleaf pine is usually considered to be 1,730 stems/ha

Thomas B. Lynch is Professor, Department of Forestry, Room 008C Ag Hall, Oklahoma State University, Stillwater, OK 74078—Phone: 405-744-5447; Fax: 405-744-3530; E-mail: tlynch@okstate.edu. Jean Nkouka is (former) OSU Graduate Research Assistant—Phone: 617-405-3625; E-mail: nkouka@hotmail.com. Michael M. Huebschmann is Natural Resource Economist, P.O. Box 50, Boise Corp., Boise, ID 3728-0001— Phone: 208-384-7212; E-mail: MikeHuebschmann@bc.com. James M. Guldin is Project Leader/Research Forest Ecologist, USDA Forest Service, Southern Research Station, P.O. Box 1270, Hot Springs, AR 71902—Phone: 501-623-1174; E-mail: jguldin@fs.fed.us.

Acknowledgments: The authors are grateful for the long-term support of the USDA Forest Service Southern Research Station, the Ouachita National Forest and the Ozark National Forest. The long-term shortleaf pine growth study that provided data used here would not exist without the many contributions of the late. Paul A. Murphy, Southern Research Station, which are gratefully acknowledged. Approved for publication by the Director, Oklahoma Agricultural Experiment Station. This research was supported under project MS-1887.

Manuscript received May 8, 2002, accepted June 13, 2001. Copyright © 2003 by the Society of American Foresters

(Shelton 1997, Campbell and Mann 1973, Grano 1967).

A study of shortleaf pine regeneration 3 yr after implementation of uneven-aged silviculture in a shortleaf pine-oak forest also illustrates the influence of factors such as variation in seed crop, summer droughts, and hardwood competition that affect shortleaf pine regeneration in the Ozark and Ouachita Mountain region (Shelton and Murphy 1997). Harvesting occurred on the plots from December 1988 to early March 1989. Shelton and Murphy (1997) stated "More and taller seedlings were observed in the 0-hardwood treatment, and values progressively declined as additional hardwoods were retained." Shortleaf seedling mortality for the period 1990–1991 ranged from 50% with no hardwoods to 86% where 6.9 m^2/ha of hardwoods were retained. Shelton and Cain (2000) give a comprehensive discussion of regeneration for uneven-aged stands of loblolly and shortleaf pine.

Shelton and Wittwer (1996) reported the results of a study of shortleaf pine regeneration in the Ouachita and Ozark Mountains from southern Missouri to southeastern Oklahoma during the 9 yr period from 1965 to 1974. During this period, they reported one bumper crop and two good crops. Their analysis indicated that seed production tended to decrease with increasing overstory pine and hardwood basal area, while tending to increase with stand age. Shelton and Wittwer (1996) concluded that in most of their study area, seed production should be sufficient for natural regeneration. Wittwer and Shelton (1992) summarized studies of shortleaf pine seed yields from Virginia to east Texas. They concluded that good seed crops could be expected to occur from 2 to 4 of 10 yr but occur less frequently near the limits of the natural range for shortleaf pine.

Analytical tools for predicting desired regeneration densities should be useful to foresters managing natural shortleaf pine stands. Shelton and Murphy (1994) present equations that predict loblolly pine seedling and sapling density (trees/ha) 5 yr after implementing uneven-aged silviculture. The regression techniques used to estimate parameters in their equations are designed to predict mean values of the dependent variable (regeneration stems per hectare). However, it would be equally useful to know the probability of obtaining a minimum density (trees/ha) sufficient to regenerate a pine forest. Larsen et al. (1997, 1999) used logistic models of the following form to predict the probability of occurrence for oak regeneration in the Missouri Ozarks:

$$P_s = \frac{\exp(b_0 + b_1 BA)}{1 + \exp(b_0 + b_1 BA)} \qquad (1)$$

where P_s is the estimated probability of attaining the specified number of reproduction density stems within a defined size class, BA is basal area per unit area for the overstory, and b_0 and b_1 are estimated coefficients. Estimates of coefficients in Equation (1) were obtained by using standard logistic regression techniques. The dependent variable was set to 1 for data plots that contained at least the specified number of reproduction stems in the specified size class and set to 0 for plots that did not satisfy these criteria. The resulting estimates can be used in Equation (1) to predict probabilities, which are restricted by the mathematical form of Equation (1) to be between 0 and 1.

Larsen et al. (1997) fitted separate sets of coefficients in Equation (1) for different specified minimum reproduction densities ranging from 125 stems/ha to 1,000 stems/ha. However, it may be desirable to fit parameters to a single equation in which the minimum density can be varied. The objectives of this article are to develop a model form that can be used to predict probability of regeneration density at specified, variable densities and to apply this model to naturally occurring shortleaf pine stands.

Data

Plots were permanently established in naturally occurring shortleaf pine forests on the Ouachita and Ozark National Forests during the period from 1985 to 1987. These plots are part of a cooperative study by the USDA Forest Service Southern Research Station, the Ouachita National Forest, the Ozark National Forest, and the Department of Forestry at Oklahoma State University. Plot locations range from areas on the Ozark National Forest near Russellville, Arkansas (latitude 35.3°N, longitude 93.1°W) to areas on the Ouachita National Forest near Broken Bow (latitude 34.0°N, longitude 94.7°W) in southeastern Oklahoma. Most of the plots are in the Ozark and Ouachita mountain highlands, while some of the southerly plots are in the West Gulf coastal plain.

The original study design specified establishment of three plots in each combination of four age classes (midpoints 20, 40, 60, and 80 yr), four site index classes (midpoints 15.2, 18.3, 21.3, and 24.4 m at 50 yr), and four density (m^2/ha) classes (midpoints 6.9, 13.8, 20.7, and 27.6 m^2/ha). Plots were remeasured during the periods from 1990 to 1992 and again from 1995 to 1997. Of the 192 plots originally specified by the design, 182 remaining for the third measurement were suitable for this study. In actual plot establishment, there were also deviations from planned age, site index, and density (m^2/ha) classifications. Each plot location consists of a 0.08 ha circular measurement plot surrounded by a 10 m buffer strip. This 10 m buffer strip received the same treatment as the measurement plot. At plot establishment, each plot and surrounding buffer strip was thinned from below to a specified residual shortleaf pine density (m^2/ha). Plots which were already below specified densities were not thinned—these occurrences were mostly at the higher densities (27.6 m^2/ha). Hardwoods in the plot and surrounding buffer strip were controlled by herbicide using injection or felling with stump spray. All shortleaf pine remaining in the 0.08 ha plot after thinning and greater than 2.5 cm dbh were mapped, classified by crown position, and had their dbh's measured. Total heights and heights to crown base were measured on a subsample of shortleaf on the 0.08 ha plot. Dominant and codominant shortleaf pine measured for total height were aged from increment cores. This enabled estimation of plot site indices using the curves developed by Graney and Burkhart (1973).

At the second remeasurement (1995 to 1997), two 0.002 ha (5 milacre) subplots were established within each

0.08 ha overstory plot for measurement of understory conditions. One 0.002 ha plot was established due north and a second due south of the center of the 0.08 ha plot. Each 0.002 ha subplot was circular, with a 2.53 m radius, and was located midway between the center and the boundary of the 0.08 ha plot. All shortleaf pine regeneration and hardwood understory stems 1.37 m tall or taller within each 0.002 ha plot radius were tallied by species and 2.5 cm dbh class. Counts of shortleaf pine regeneration for the two 0.002 ha subplots were averaged on each 0.08 ha overstory plot and expanded to a per hectare basis for model development. The interval between plot establishment and understory measurement was 9–10 yr. Means, ranges, and standard deviations for the data used to fit the model of probability for regeneration at specified densities are given in Table 1.

Probability of Regeneration with Varying Minimum Trees per Hectare

An expression for the probability of obtaining regeneration at varying minimum densities can be obtained by modifying the logistic function. The model should be such that the probability of desired regeneration density increases monotonically as the minimum regeneration density decreases. The probability of obtaining at least zero regeneration stems should be one while the limiting probability of attaining desired density should approach zero as the minimum regeneration density goes to infinity. The following model form satisfies these conditions:

$$P_{\mathbf{b}}(T_j \geq t_k) = \left[1 + t_k^{b_{02}+\sum_i b_{i2}X_{ij}} \exp(b_{01} + b_{11}X_{1j} + \cdots + b_{n1}X_{nj})\right]^{-1} \quad (2)$$

rearranging terms:

$$P_{\mathbf{b}}(T_j \geq t_k) = \left[1 + \exp(b_{01} + (b_{02} + \sum_i b_{i2}X_{ij})\ln(t_k) + b_{11}X_{1j} + \cdots + b_{n1}X_{nj})\right]^{-1} \quad (3)$$

where

$$b_{02} + \sum b_{i2}X_{ij} > 0$$

$P_{\mathbf{b}}(T_j \geq t_k)$ is the estimated probability that the regeneration trees/ha are greater than or equal to t_k for plot j,

T_j is number of shortleaf regeneration stems per hectare for plot j,

X_{ij} is the value for the independent variable i on plot j,

t_k is regeneration threshold k,and

$\mathbf{b} = (b_{01}, b_{11}, \cdots b_{n1}, b_{02}, b_{12}, \cdots b_{n2})$ is a vector of coefficients estimated from the data.

Equation (2) is constrained to predict a probability of 1 for obtaining at least $t_k = 0$ regeneration stems. As t_k approaches infinity, the probability of obtaining t_k or more regeneration stems goes to zero in Equation (2). Equation (3) is mathematically equivalent to Equation (2); However, this form is more similar to that used in commonly available logistic regression programs. Lowell and Mitchell (1987) used a predefined percent dbh growth as an independent variable in a logistic regression to estimate the probability of exceeding or equaling a specified dbh in 5 yr. Since several levels of growth percent were considered, their approach resembles Equation (3). Johnson (1984) used logistic regression to estimate the probability that northern red oak (*Quercus rubra* L.) regeneration exceeds a specified height. This also seems to be similar to Equation (3) since the specified height level was used as an independent variable, and multiple height levels were used. These authors apparently stacked or combined replicates of the dataset representing variable success levels. This approach can provide workable estimates but ignores correlation among replicates. The observation associated with a tree evaluated at the first success level is correlated with the observation representing the same tree evaluated at the second success level.

The multinomial distribution can be used to develop a correct likelihood function for estimating parameters with Equation (3). Using Equation (3) with three minimum densities, $t_1 = 740$, $t_2 = 1{,}235$ and $t_3 = 1{,}730$, it is evident that the regeneration density on each plot must fall into one of four categories so that one of the following Bernoulli variables will be one while the others will be zero:

Table 1. Summary of data from 182 plots used to estimate parameters in a model of probability of attaining desired regeneration density (trees/ha) for naturally occurring shortleaf pine.

Variable	Mean	SD	Min	Max
Basal area (m^2/ha)	17.19	7.49	6.27	29.61
Age (yr)	54.96	20.14	18	93
Site index (m at 50 yr)	18.75	3.44	11.86	26.55
Regen. trees/ha	859.44	2293.28	0	13343.69
Y_1 = 1 if regen. < 740/ha, else 0	0.791	0.408	0	1
Y_2 = 1 if 740/ha ≤ regen.< 1235/ha, else 0	0.0275	0.164	0	1
Y_3 = 1 if 1,235/ha ≤ regen.< 1,730/ha, else 0	0.0495	0.217	0	1
Y_4 =1 if regen ≥ 1,730/ha, else 0	0.132	0.339	0	1

Y_{1j} = 1 if less than 740/ha shortleaf regeneration on plot j, otherwise 0,

Y_{2j} = 1 if less than 1,235/ha but at least 740/ha shortleaf regeneration on plot j, otherwise 0,

Y_{3j} = 1 if less than 1,730/ha but at least 1,235/ha shortleaf regeneration on plot j, otherwise 0, and

Y_{4j} = 1 if at least 1,730/ha shortleaf regeneration on plot j, otherwise 0.

The multinomial distribution associated with these random variables leads to the following likelihood function:

$$L(\mathbf{b}) = \prod_{j=1}^{n} \left[P_{\mathbf{b}}(T_j \geq t_3)\right]^{Y_{4j}} \left[P_{\mathbf{b}}(t_3 > T_j \geq t_2)\right]^{Y_{3j}} \left[P_{\mathbf{b}}(t_2 > T_j \geq t_1)\right]^{Y_{2j}} \left[P_{\mathbf{b}}(t_1 > T_j)\right]^{Y_{1j}} \quad (4)$$

Taking natural logarithms in Equation (4) yields:

$$\log(L(\mathbf{b})) = \sum_{j=1}^{n} \{Y_{4j} \log(P_{\mathbf{b}}(T_j \geq t_3) + Y_{3j} \log(P_{\mathbf{b}}(t_3 > T_j \geq t_2) + Y_{2j} \log(P_{\mathbf{b}}(t_2 > T_j \geq t_1) + Y_{1j} \log(P_{\mathbf{b}}(t_1 > T_j)\} \quad (5)$$

where t_1 = 740/ha, t_2 = 1,235/ha and t_3 = 1,730/ha shortleaf regeneration stems. For parameter estimation, Equation (3) is substituted for

$$P_{\mathbf{b}}(T_j \geq t_k)$$

and

$$P_{\mathbf{b}}(t_{k+1} > T_j \geq t_k) = P_{\mathbf{b}}(T_j \geq t_k) - P_{\mathbf{b}}(T_j \geq t_{k+1}).$$

Equation (4) is based on the likelihood function for multinomial choice models equation 18.3.1 given by Judge et al. (1985, p. 768). The use of the multinomial distribution to obtain a likelihood function is similar to the development of the multinomial logit model (Judge et al. 1985, p. 768–772). However, the formulation for probabilities used in Equations (4) and (5) is different from the multinomial logit model in several ways, one being that the same function with common parameters is utilized in all categories. The method discussed above differs from logistic regression techniques used for ordinal data in that the variable t (regeneration level) in the final model can be varied continuously. Maximum likelihood estimates of parameters were obtained by maximizing Equation (5) with the LOGDEN function in SHAZAM econometrics software (White 1993, p. 255–266).

Results

Nkouka (1999) found that variables likely to be correlated with shortleaf pine regeneration density include plot age, site index, and plot basal area per hectare. Shelton and Wittwer (1996) have documented the positive correlation of stand age and the negative correlation of basal area with shortleaf seed production. Lawson (1986) noted the difficulty for growth and survival that shortleaf regeneration experiences on better sites due to rapid hardwood regrowth. Thus, plot basal area per hectare, site index, and age were selected as potential independent variables.

Regression analysis was used to investigate prediction models for prediction of the number of shortleaf pine regeneration stems per hectare. Parameters in the following model were estimated by using nonlinear least squares:

$$\hat{T}_j = \exp(\underset{(1.0441)}{8.4964} - \underset{(0.0274)}{0.1356} X_{1j} + \underset{(0.00520)}{0.0175} X_{2j} - \underset{(0.0399)}{0.1411} X_{3j} + \underset{(0.8238)}{2.1453} X_{4j}) \quad (6)$$

where

$\hat{T}_j$ is the predicted number of shortleaf pine regeneration stems per hectare, plot j

X_{1j} is plot j basal area per hectare after thinning (m^2/ha),

X_{2j} is plot j age (years),

X_{3j} is plot j site index (average total height of dominants and codominants in meters at 50 years),

X_{4j} is 1 if plot j was thinned and treated in 1985 or 1986, otherwise 0.

Standard errors are given in parentheses beneath parameter estimates. The standard error for the residuals (root mean square error) from this model was 1,878. The model explains 34% of the total variation (as measured by the corrected sum of squares) in the number of shortleaf pine regeneration stems per acre ("fit index" 0.34). The standard error for this model is rather high. As a result, the model explained a relatively low proportion of the total variability in shortleaf pine regeneration stems. Equation (6) estimates the mean or expected number of shortleaf pine regeneration stems and does not directly estimate the probability of attaining desired levels of regeneration. Therefore Equation (3) was the basis of a model for predicting the probability of attaining desired shortleaf pine regeneration densities.

Preliminary analyses of the data confirmed that variables important for predicting probability of attaining desired regeneration density included plot age, site index, and plot basal area per hectare (Nkouka 1999). These analyses included use of PROC LOGISTIC (SAS Institute Inc. 1997) on dependent variables associated with $T_j \geq 741$, $T_j \geq 1{,}235$ and $T_j \geq 1{,}730$ separately as well as with a "stacked" dataset including all three levels. STEPWISE and BACKWARD options in PROC LO-

Table 2. Goodness of fit for predicting probability of shortleaf regeneration stems/ha on randomly selected data independent of parameters used for prediction.

Basal area (m^2/ha), Site index (m)	No. of plots	740/ha or more		1,235/ha or more		1,730/ha or more	
		Observed	Expected	Observed	Expected	Observed	Expected
Below or equal 16, below or equal 17	8	5	3.9	5	3.4	5	3.1
Below or equal 16, above 17	7	3	2.5	2	1.9	1	1.6
Above 16, below or equal 17	20	4	4.6	2	3.8	1	3.3
Above 16, above 17	26	2	4.3	2	3.1	2	2.4
Total*	61	14	15.3	11	12.2	9	10.4

NOTE: For 740/ha: Chi-square = 2.29, *P*-value = 0.68; for 1,235/ha: Chi-square = 2.78, *P*-value = 0.59; for 1,730/ha: Chi-square = 4.23, *P*-value = 0.38. Chi-square values include expected and observed failure results.

* Deviation of totals and column sums due to rounding.

GISTIC were used as aids in variable selection. Basal area per hectare was consistently the most highly significant variable. The significance of the relationship between basal area per hectare and shortleaf pine regeneration has long been noted in the literature, as indicated above. This is due to the fact that shortleaf pine is a shade intolerant species, so that high levels of basal area per hectare tend to be correlated with lower densities of shortleaf pine regeneration. Nkouka (1999) indicated that regeneration densities on plots established in 1985–1986 differed substantially from those established in 1987. This could be a result of the annual shortleaf pine seed crop variability in the Ozark-Ouachita Mountain region (Shelton and Wittwer 1996). Even during years when adequate seed crops occur, the hot dry summers common to the region may impede seedling establishment. Further analysis of a model for probability of attaining regeneration density based on Equation (3) indicated a significant relationship between site index and the minimum regeneration density. Since some plots were measured on a 9 yr interval while others were measured on a 10 yr interval, preliminary tests were conducted using PROC LOGISTIC to evaluate the significance measurement interval on the model. Tests using PROC LOGISTIC STEPWISE and BACKWARD with a significance level of $\alpha = 0.05$ and the dependent variables discussed above consistently resulted in a model that did not include measurement interval. These considerations led to the following model for estimated probability of attaining desired regeneration for shortleaf pine, 9–10 yr after thinning and hardwood control:

$$P_b(T_j \geq t_k) = \frac{1}{1+\exp(b_0 + b_1 X_{1j} + b_2 X_{2j} + b_3 X_{3j} + b_4 X_{3j}^2 \ln(t_k) + b_5 X_{4j})} \quad (7)$$

where t_k is minimum threshold regeneration density (trees/ha), and b_0, b_1, b_2, b_3, b_4, b_5 are coefficients estimated using the data.

In order to obtain an independent dataset to test coefficient estimates obtained by using the log-likelihood Equation (5), one-third of the data were randomly removed as an independent dataset for testing, while the remaining two-thirds of the data were a "fitting" dataset used to obtain coefficient estimates. This procedure provided test data independent of coefficient estimates.

The independent dataset was used to evaluate model performance in Table 2 by comparing actual shortleaf regeneration stems 9–10 yr after thinning to the expected number predicted by Equation (7) in four overstory shortleaf basal area per hectare—site index classes. These four classes are: (1) basal area equal to or below 16m^2/ha and site index less than or equal to 17 m; (2) basal area equal to or below 16m^2/ha and site index greater than 17 m; (3) basal area above 16m^2/ha and site index less than or equal to 17 m; (4) basal area above 16m^2/ha and site index greater than 17 m. Basal area specifications for categories are based on overstory densities after thinning at the time of plot establishment. The expected number of plots exceeding the specified number t_k of shortleaf regeneration stems per hectare in each category $E(T_C)$ is:

$$E(T_C) = \sum_{i=1}^{n_c} P(T_i > t_k)$$

where Equation (7) is used to predict values of $P(T_i > t_k)$ and n_c is the number of plots in the basal area – site index class of interest. Inspection of Table 2 shows reasonable correspondence between predicted and actual numbers of plots achieving specified regeneration thresholds, 740/ha, 1235/ha, and 1730/ha. A chi-square statistic was calculated for each regeneration density (trees/ha) in Table 2. Neter et al. (1989, p. 613–615) and Hosmer and Lemeshow (1980, 1989 p. 140–145) discuss evaluation of logistic models by using a chi-square test. For each regeneration density in Table 2, the number of site index – basal area per

Table 3. Agreement data to estimate Cohen's measure of agreement *K* (standard error), $\hat{K}$ = 0.38 (0.16) for predicting that randomly selected plots independent of parameter estimates will have 740 or more shortleaf pine regeneration stems per hectare, with *P*-value = 0.02 for H_0: *K* = 0.

	Actual 740 or more	Actual less than 740	Total
Predicted 740 or more	5	2	7
Predicted less than 740	9	45	54
Total	14	47	61

Table 4. Parameter estimates of a model that predicts probability of attaining specified densities of naturally occurring shortleaf pine regeneration, based on the combined dataset consisting of 182 plots.

Variable	Coefficient	Estimate	SD	*P*-value
Intercept	b_0	3.6412	1.7685	0.041
X_1, basal area (m^2/ha)	b_1	0.13238	0.030571	0.000025
X_2, age (yr)	b_2	−0.01741	0.010104	0.087
X_3, site index (m at 50 yr)	b_3	−0.43772	0.15666	0.0058
$(X_3)^2\ln(t_k)$, (site index)^2 × ln(regeneration density*)	b_4	0.0023892	0.0006014	0.00010
X_4,thin year	b_5	−1.2300	0.45333	0.0073

* Trees/ha specified.

hectare categories c is 4. Therefore, the *P*-values given at the bottom of Table 2 are based on 4 degrees of freedom. These *P*-values indicate the hypothesis that the model fits the data cannot be rejected at the $\alpha = 0.05$ significance level for any of the three minimum regeneration densities. Thus, these tests fail to reject Equation (7) as a model for probability of achieving a specified regeneration density.

The independent dataset was also used to evaluate the ability of the model to predict which individual plots in this dataset would contain 740, 1,235, or 1,730 or more shortleaf pine regeneration stems per hectare. Plots were designated likely to attain a specified number of shortleaf regeneration stems per hectare when Equation (7) predicted a probability for attaining that density of one-half or greater. Table 3 summarizes the agreement between predicted and actual numbers of plots in the independent dataset attaining 740 or more shortleaf pine regeneration stems per hectare. Based on these data, Cohen's coefficient of agreement *K* (Cohen 1960, Bishop et al. 1975, p. 395–397) was estimated to be (standard errors in parentheses) 0.38 (0.16). Since the estimated coefficient of agreement was more than twice its standard error, the coefficient of agreement was significantly different from zero at the $\alpha = 0.05$ level of significance (also see *P*-value given in Table 3). Similar tests indicated that Cohen's *K* was not significant for the probability of attaining 1,235 or 1,730 trees per hectare. However, Cohen's *K* coefficient should be interpreted in light of the nature of predicting the probability of events that have a less than 50% chance of happening. Suppose a collection of plots having a given set of independent variables is predicted to have a 30% chance of attaining 1,730 shortleaf regeneration stems per hectare or more. None of these individual plots would be predicted to attain 1,730 regeneration stems per hectare because none of them had a predicted probability of greater than 50%, yet we would expect that 30% of these plots would have 1,730 regeneration stems per hectare or more.

For comparison purposes, parameters for Equation (6) were reestimated using the same split fitting data discussed above [the parameter estimates shown in Equation (6) above are based on the complete dataset]. Predictions from the resulting equation were used to predict which plots in the independent data had 740 trees per hectare or more, and Cohen's *K* was calculated for this classification. The resulting $K = 0.40$ was statistically significant and similar in magnitude to that for Equation (7) discussed above. However, Equation (7) classified 80% of the plots in the independent data correctly as compared to 73% for Equation (6). Equations (6) and (7) were designed for different primary objectives. Equation (6) is not designed to directly estimate the probability of attaining specified levels of regeneration stems per hectare.

The "fitting" and "independent" datasets were combined to obtain estimated values of Equation (7) coefficients given in Table 4. The signs of the coefficients for age and thinning year are negative, indicating that the probability of attaining regeneration density will increase as the values of these variables increase. The positive sign associated with the basal area per hectare coefficient in Table 4 indicates a decreasing probability of attaining regeneration density as basal area per hectare increases. The sign of the coefficient associated with the interaction between regeneration density t_k and the square of site index is positive, indicating decreasing probability of attaining regeneration density as the threshold for desired density is increased. This is a logical property of the model.

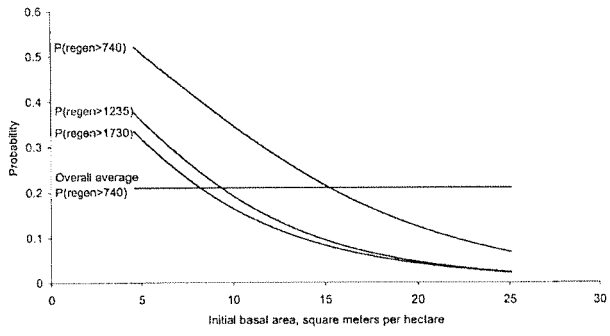

Figure 1. Probability of obtaining shortleaf regeneration greater than three specified densities 9–10 yr after thinning at age 50 yr on site index 21 m (base age 50 yr) stands, good conditions.

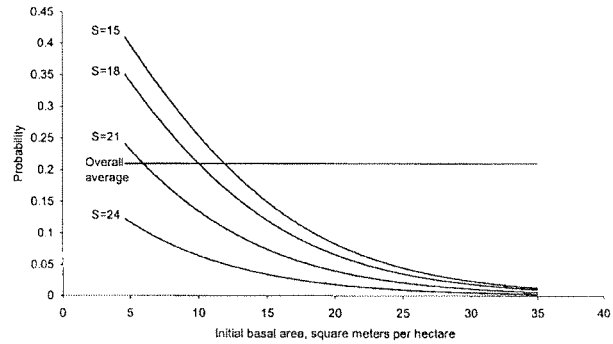

Figure 2. Probability of obtaining 740 or more shortleaf regeneration stems per hectare 9–10 yr after thinning at age 50 by site index (S) class, poor conditions.

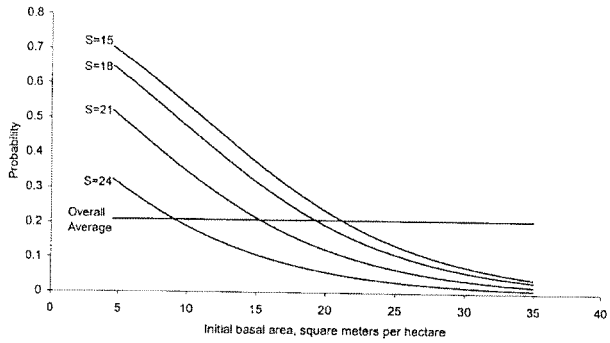

Figure 3. Probability of obtaining 740 or more shortleaf regeneration stems per hectare 9–10 yr after thinning at age 50 by site index (S) class, good conditions.

Site index is associated with two coefficients, b_3 and b_4, one positive and one negative. Since site index is squared in the term associated with b_4, trials show that the net effect of increasing site index in the model is to decrease the probability of attaining regeneration density for ranges of site index normally observed in shortleaf pine natural stands. These model properties are consistent with the relationships between shortleaf seed production, age, and basal area observed by Shelton and Wittwer (1996) and with Lawson's (1986) observations on site quality, hardwood competition, and survival of shortleaf regeneration.

Properties of Equation (7) are illustrated in Figures 1–3. For all figures and all threshold minimum densities, the probability of attaining that density (trees/ha) 9–10 yr after thinning declines sharply with increasing overstory shortleaf basal area per hectare. Figures 2 and 3 indicate that the probability of attaining at least 740 shortleaf regeneration stems per hectare 9–10 yr after thinning is greater on poorer sites. This corresponds with Lawson's (1986) observations regarding the effect of hardwood regrowth on shortleaf seedling survival on better sites. Shelton and Cain (2000) state that for uneven-aged stands, it is more difficult to obtain loblolly or shortleaf regeneration for good sites due to increased competition from nonpine species. Consequently, they discuss use of selective herbicides. Comparison of Figures 2 and 3 illustrate the possible differences represented by annual variation in good and poor conditions for seed production and seedling establishment. Figure 2 represents poor conditions ($X_{4j} = 0$), while Figure 3 shows good conditions ($X_{4j} = 1$). Comparison of these figures indicates that the differences in probability of attaining regeneration density (trees/ha) can be substantial. This may be due to the wide variation in annual seed crops that is typical in the Ozark-Ouachita Mountain region (Shelton and Wittwer 1996) or to differences in conditions for seedling establishment and survival.

Discussion and Conclusions

A model has been developed that can be used to predict the probability of attaining specified regeneration densities 9–10 yr after thinning in naturally occurring shortleaf pine forests. Parameters were estimated using specified regeneration densities between 740 and 1,730 regeneration stems per hectare. The model form allows for continuous variation of the regeneration density level desired. Variation between 740 and 1,730 stems per hectare would probably be justified for the model parameters presented here. Variables that significantly affected the probability of regeneration density included overstory basal area per hectare, site index, overstory age, and an indicator variable used to express the difference between good and poor years for seedling establishment. These factors are consistent with the results of previous research concerning variations in shortleaf pine seed production and seedling establishment.

Although the parameters here were estimated using data from naturally occurring shortleaf pine forests, the basic model form should be applicable to estimation of probability of attaining regeneration density for other tree species. Applying the model to other species would probably require the use of somewhat different sets of independent variables. However, the basic model structure and the method of parameter estimation presented here should be broadly applicable.

Literature Cited

BISHOP, Y.M.M., S.E. FIENBERG, AND P.W. HOLLAND. 1975. Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, Mass. 557 p.

CAMPBELL, T.E., AND W.F. MANN, JR. 1973. Regenerating loblolly pine by direct seeding, natural seeding, and planting. USDA For. Serv. Res. Pap. SO-84. 9 p.

COHEN, J. 1960. A coefficient of agreement for nominal scales. Educ. Psych. Meas. 20:37–46.

GRANEY, D.L., AND H.E. BURKHART. 1973. Polymorphic site index curves for shortleaf pine in the Ouachita mountains. USDA For. Serv. Res. Pap. SO-85. 12 p.

GRANO, C.X. 1967. Growing loblolly and shortleaf pine in the Midsouth. Farmer's Bull. 2102. USDA, Washington, DC. 27 p.

HOSMER, D.W., JR. AND S. LEMESHOW. 1989. Applied logistic regression. Wiley, New York. 307 p.

HOSMER, D.W., JR. AND S. LEMESHOW. 1980. A goodness of fit test for the multiple logistic regression model. Commun. Stat. A10:1043–1069.

HUEBSCHMANN, M.M. 2000. Restoring the shortleaf pine-bluestem grass ecosystem on the Ouachita National Forest: An economic evaluation. Ph.D. thesis, Oklahoma State University, Stillwater, OK. 141 p.

JOHNSON, P.S. 1984. Responses of planted northern red oak to three overstory treatments. Can. J. For. Res. 14:536–542.

JUDGE, G.G., W.E. GRIFFITHS, R.C. HILL, H. LÜTKEPOHL, AND T-C. LEE. 1985. The theory and practice of econometrics. Ed. 2. Wiley, New York. 1019 p.

LARSEN, D.R., E.F. LOEWENSTEIN, AND P.S. JOHNSON. 1999. Sustaining recruitment of oak reproduction in uneven-aged stands in the Ozark highlands. USDA For. Serv. Gen. Tech. Rep. NC-203. 11 p.

LARSEN, D.R., M.A. METZGER, AND P.S. JOHNSON. 1997. Oak regeneration and overstory density in the Missouri Ozarks. Can. J. For. Res. 27:869–875.

LAWSON, E.R. 1986. Natural regeneration of shortleaf pine. P. 53–63 *in* Proc. of a Symp. on the shortleaf pine ecosystem, P.A. Murphy (ed.). Arkansas Coop. Ext. Serv., Monticello, AR. 272 p.

LOWELL, K.E., AND R.J. MITCHELL. 1987. Stand growth projection: Simultaneous estimation of growth and mortality using a single probabilistic function. Can. J. For. Res. 17:1466–1470.

NETER, J., W. WASSERMAN, AND M.H. KUTNER. 1989. Applied linear regression models. Ed. 2. Irwin, Homewood, IL. 667 p.

NKOUKA, J. 1999. Evaluation of the effect of thinning levels, site index, and age on shortleaf pine (*Pinus echinata* Mill.) regeneration and hardwood understory. M.Sc. thesis, Oklahoma State University, Stillwater, OK.76 p.

SAS Institute Inc. 1997. SAS/STAT® software: Changes and enhancements through release 6.12. SAS Institute Inc., Cary, NC. 1167 p.

SHELTON, M.G. 1997. Development of understory vegetation in pine and pine-hardwood shelterwood stands in the Ouachita Mountains—the first 3 years. USDA For. Serv. Res. Pap. SRS-8. 18 p.

SHELTON, M.G., AND M.D. CAIN. 2000. Regenerating uneven-aged stands of loblolly and shortleaf pines: the current state of knowledge. For. Ecol. Manage. 129:177–193.

SHELTON, M.G.. AND P.A. MURPHY. 1994. Loblolly pine regeneration and competing vegetation 5 years after implementing uneven-aged silviculture. Can. J. For. Res. 24:2448–2458.

SHELTON, M.G., AND P.A. MURPHY. 1997. Understory vegetation 3 years after implementing uneven-aged silviculture in a shortleaf pine-oak stand. USDA For. Serv. Res. Pap. SO-296. 13 p.

SHELTON, M.G., AND R.F. WITTWER. 1996. Shortleaf pine seed production in natural stands in the Ouachita and Ozark Mountains. South. J. Appl. For. 20(2):74–80.

WHITE, K.J. 1993. SHAZAM user's reference manual version 7.0—second printing. McGraw-Hill, New York. 482 p.

WITTWER, R.F., AND M.G. SHELTON. 1992. Seed production in natural shortleaf pine stands. P. 113–123 *in* Proc. of the Shortleaf pine regeneration workshop. USDA For. Serv. Gen. Tech. Rep. SO-90. 236 p.