ESTIMATING THE PROBABILITY OF SURVIVAL OF INDIVIDUAL SHORTLEAF PINE (*Pinus echinata* Mill.) TREES

By

SUDIP SHRESTHA

Bachelor Science in Forestry

Tribhuvan University

Pokhara, Kaski, Nepal

2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2010

ESTIMATING THE PROBABILITY OF SURVIVAL

OF INDIVIDUAL SHORTLEAF PINE

(*Pinus echinata* Mill.) TREES

Thesis Approved:

Thomas B. Lynch

Thesis Adviser

Carla Goad

Difei Zhang

Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my major adviser Dr. Thomas B. Lynch for his time, encouragement and advice throughout the study period. I always found him whenever I needed statistical and computational support. It would be impossible to bring this research work at this point without his guidance. Dr. Lynch, I am very thankful to you.

I also would like to thank my respected committee members Dr. Carla Goad for statistical support and suggestions and Dr. Difei Zhang for valuable comments and input on each section of my research work. It is a pleasure to thank Doug Stevenson and Jomals Mathews John who provided worthwhile thoughts for the research work and supported in data editing.

This research is also supported by project OKLO-1897 of the Oklahoma Agricultural Experiment Station. I sincerely appreciate the members of the project for their cooperation in data collection and financial support.

Continuous support and suggestion received from Nabin Gyawali and Gary Hatfield are highly appreciated. They have been a great source of support in the course of statistical learning. I would like to remember the Department of Natural Resource Ecology and

Management (NREM) staff at Oklahoma State University (OSU) for their support and cooperation. They have been very amicable and helpful.

Thanks are also due to all my Nepalese friends for their encouragement and providing homely environment during my stay in Stillwater. I am so indebted to my nearest and dearest Binita K.C. for her never-ending support and encouragement in every step. She has always been a person of great importance for me.

Lastly, I owe deepest gratitude to my parents and all who supported me in any respect during completion of this research work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1. Shortleaf Pine

Shortleaf pine (*Pinus echinata* Mill.), a medium sized tree species, is also known as Southern Arkansas soft pine or yellow pine depending upon locale. The shortleaf pine is native to south-central and southeastern states in the U.S. It is the state tree of Arkansas (ODAFS, 2000). The bark of the shortleaf pine tree is brown to dark brown colored. Needles occur in clusters of 2 or 3 and the size of needles is about 3 inches to 8 inches long. The height of the tree is about 50-100 ft. with short, spreading branches forming a pyramidal shaped crown that opens with age. It is considered as one of the four major commercially valuable conifers in the southeastern United States. Shortleaf pine has second largest standing volume among the four southern pines. The tree is important for timber, lumber production, millwork and many other structural materials. Even the taproots are useful in making pulpwood. Shortleaf pine is commonly grown on non-industrial private ownership in Oklahoma and Arkansas. Shortleaf pine forests are also managed by some industrial private owners in the region such as Deltic Farm and Timber and Plum Creek. Shortleaf pine is a major component of the shortleaf pine-bluestem grass ecosystem which was the pre-settlement forest type in much of western Arkansas and southeastern Oklahoma. The shortleaf pine-bluestem grass ecosystem is now being restored on a portion of Ouachita National Forest.

Shortleaf pine can be found in a wide range of soil and site conditions (Lawson, 1990). Compared with other southern pines, shortleaf pine has the largest range of occurrence covering more than 440,000 square miles (Willett, 1986). According to Lawson (1990) shortleaf pine is distributed in 22 states ranging from southeastern New York and New Jersey west to Pennsylvania, southern Ohio, Kentucky, southwestern Illinois, and southern Missouri; south to eastern Oklahoma and eastern Texas; and east to northern Florida and northeast through the Atlantic Coast States to Delaware (Figure 1). Shortleaf pine grows best on deep, well-drained soils having fine sandy loam or silty loam textures.

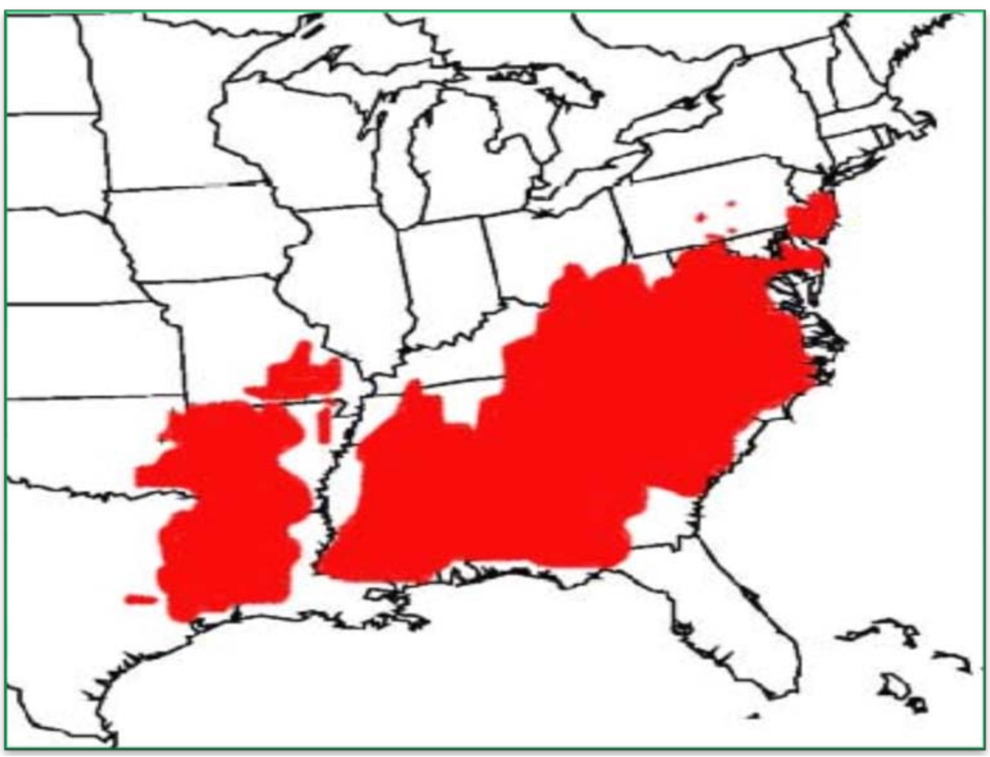

Figure 1. Shortleaf pine distribution in the U.S.

Source: http://www.nearctica.com/trees/conifer/pinus/Pechin.htm

In 1986, Smith reported that the largest shortleaf pine forest in the world is in the Ouachita Mountains of Arkansas and Oklahoma. Wide distribution and great commercial value of this species call for accurate estimation of growth and yield, which in turn calls for accurate prediction of future survival of shortleaf pine.

1.2. Purpose of the Study

The purpose of this study is to investigate the important independent variables and to develop an annual survival equation for individual shortleaf pine trees. Iteratively reweighted nonlinear regression was used to get the best estimates of the parameters. A nonlinear mixed model was also applied to investigate the model performance.

CHAPTER II

MANUSCRIPT I

ESTIMATING THE PROBABILITY OF SURVIVAL OF INDIVIDUAL SHORTLEAF PINE (*Pinus echinata* Mill.*)* TREES

Abstract

An individual tree survival model was developed for shortleaf pine (*Pinus echinata* Mill.*)* trees. Prediction of the probability of survival of an individual tree is essential when considering growth and yield of a stand. Data for this study were from more than 200 permanently established plots on even-aged natural shortleaf pine stands that were located in the Ozark and Ouachita National Forests. Plots were established during the period of 1985-1987. Plots have been remeasured every 4, 5 or 6 years, and individual tree survival or mortality was recorded at each measurement. These plots were selected to represent a range of ages, densities and site qualities. Logistic regression was used to find the best sets of significant predictor variables in which the response variable was a binary variable '1' for the survival tree and '0' for the mortality tree. Significant variables found in predicting the survival were mid-period basal area per acre (Mid-BA), inverse of ratio of quadratic mean diameter to diameter at breast height (DBH) (DRINV), their interaction and square of DBH. Parameters of the logistic equation were estimated using iteratively re-weighted nonlinear regression. A nonlinear mixed-effects approach was also applied to investigate the plot level effect on the model. Model performance was evaluated using chi-square goodness-of-fit test, and it was found that the model worked better while estimating the parameters using iteratively reweighted non linear regression than with the nonlinear mixed model. This individual tree survival model can be used to predict the annual survival rate of individual trees of even-aged shortleaf pine forests located in Ozark and Ouachita National Forests and in the surrounding regions.

2.1. Introduction

2.1.1. Individual Tree Survival/Mortality Model

Growth and yield are important factors in forest management. Accurate and reliable growth and yield information supports better forest management. An individual-tree modeling approach is one of the better methods available for predicting growth and yield as it provides essential information about particular tree species; tree size, tree quality and tree present status. An equation for estimating the probability of individual tree survival is important for management of shortleaf pine forests on a sustainable basis. Individual tree survival models simulate survival and growth of individual trees in a forest stand. They are important in predicting the growth and yield of trees and forests and in determining the development pattern of stand.

A survival model is a major component of the Shortleaf Pine Stand Simulator (SLPSS) (Huebschmann, 1998) which has been developed for even-aged natural shortleaf pine forests. SLPSS includes a prediction equation for probability of tree survival which is based on repeatedly measured plots permanently located in the Ozark and Ouachita National forests which have diverse ages, site qualities and densities. Other important components of SLPSS 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).

Annual survival equations predict the probability that a tree survives the following year. It is difficult to fit the annual survival equation to the data that were measured repeatedly at intervals longer than one year (Cao, 2000). However it is essential to build a good forest growth model from periodically measured data to obtain a good estimate of survival/mortality probability. Individual–tree survival/mortality models have been developed for various species in different forest types and in various site conditions. The most commonly used regression model to estimate the survival/mortality of individual tree is the logistic model (Cao, 2000). Use of the logistic function is often found to be a good approach compared to other regression models for the survival/mortality of an individual tree as well as for stand level survival.

2.1.2. Logistic Model

Regression methods have long been an important tool for explaining the relationship between a response variable and one or more explanatory variables. Sometimes, the dependent variable can take only one of two discrete values in which case the logistic regression model provides one of the better options for modeling the relationship between the binary (0 or 1) dependent variable and the independent variables (Hosmer et al., 2000; Neter and Maynes, 1970). In this study the dependent variable is binary i.e. if the subject tree survives to the end of the measurement period then the dependent variable is indicated by ‘1’ and if the tree does not survive it is indicated by ‘0’.

Kutner et al. (2004) listed some potential problems that would be encountered by applying ordinary regression methods to a binary dependent variable. Some of them are discussed below:

1. Non- normality of error terms: Unlike the situation for the error term in a general linear regression model, the assumption of a normal error term is not appropriate when the dependent variable is binary.
2. Non-constancy of error variance: The assumption of a constant error variance is not satisfied when the dependent variable is ‘0’ or ‘1’.
3. Limitation of response function: The mean response is forced to be between ‘0’ and ‘1’, that is, $0 \leq E[Y] \leq 1$, where E[Y] is the mean response function.

Hamilton (1986) specified the advantages of the logistic function as:

1. The logistic function provides a good description of the probability of survival or mortality. A variable transformation may be required to get the best outcome.
2. The logistic function can be used with iteratively reweighted nonlinear regression or equivalently maximum likelihood estimation can provide optimal estimates of regression parameters.

When using logistic regression the sum of squared errors (SSE) between the dependent variable and model prediction is minimized (Lynch et al., 1998).

The logistic model is formulated as:

$$APOS = [1 + \exp\{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_n)\}]^{-1} + \varepsilon \quad (1)$$

where;

APOS= annual probability of survival

$X_1, X_{2,\ldots,} X_n$ = set of n predictor variables

$\beta_0, \beta_{1,\ldots,} \beta_k$ = unknown regression parameters

ε = independent identically distributed error term with mean zero

exp = base of the natural logarithm

Data from re-measured plots includes multi-year intervals in which tree survival or death is observed. Let 't' be such interval over which to observe tree survival / death. (Hamilton and Edwards, 1976) and (Monserud, 1976) described the following model to estimate the probability of survival assuming that survival time follows uniform distribution over the growth interval.

$$P_j^t = [\, 1 + \exp\{-(\beta_0 + \beta_1 X_{1j} + ... + \beta_k X_{nj})\}]^{-t} + \varepsilon_j \qquad (2)$$

where;

P_j^t = the probability that tree j survives a t year period

t = the number of years in the measurement period, and

X_{ij} = the value of independent variable X_i for tree j

ε_j = independent identically distributed errors with mean zero

The above model predicts the survival probability in different time intervals. When the interval t is zero or at the beginning of growth period the probability that the tree survives is 1 which means the tree is definitely alive. Conversely, as $t \rightarrow +\infty$, the probability of survival decreases and eventually approaches zero (Yao et al., 2001). In this study iteratively reweighted nonlinear regression and nonlinear mixed-effects models were used for the estimation of regression parameters.

2.1.3. Mixed-Effects Model

Models that include both fixed and random effects are mixed models or mixed-effects models (Budhathoki et al., 2008a). When the levels of an effect in our experiment constitute the whole population then the effect is known as fixed effect. Most models in analysis of variance, regression and general linear models (GLM) are fixed effects models. On the other hand, an effect is known as random effect when inferences are made to entire population and the levels in the experiment are only a sample from that population. Through the random-effects modeling, inferences over a wider population can be made which is not possible by general linear models (Hanneman, 2010). Mixed-effects models often perform well and are very flexible in analyzing balanced and unbalanced data. These data arise in several areas of investigation and are characterized by the presence of correlation between observations within the same group; some examples are repeatedly measured data, longitudinal studies, and nested designs (Pinheiro and Bates, 2000). Classical modeling techniques that assume independence of the observations are not suitable for the data which are grouped.

Mixed models consist of more than one error level and are also known as multilevel models, hierarchical model, a nested model, or a random-effects model, depending upon the model types. These mixed models can be extended to nonlinear models. In the nonlinear mixed-effect modeling approach, both random and fixed effects have nonlinear association with response variable (Wolfinger, 1999).

2.2. Literature Review

2.2.1. Tree Survival/Mortality

Survival/Mortality is important factor in predicting growth and yield of trees and forests and in understanding the growth pattern of forest stands. Tree sizes, stand density, species composition, site quality, and competition among the trees are major components that determine mortality (Peet and Christensen, 1987). Yang (2003) emphasized that tree survival/mortality is a major factor in influencing the forest growth and yield, therefore it is very important to get the best estimation of tree survival/mortality. According to Lee (1971) mortality can be classified as regular and irregular mortality. Tree mortality caused by the competition for nutrients, light, and moisture is called regular mortality in which crowded, overtopped and suppressed trees eventually die. Competition among trees within a stand has significant effects on individual tree survival. An increase in basal area per acre/hectare decreases the chance of individual tree survival because of increased competition for moisture, sunlight and nutrients (Teck and Hilt, 1990). Lee further discussed irregular mortality which is caused by insect, disease, and natural disasters such as fire, windfall, and snow. He added that the rate of irregular mortality is very difficult to predict and it is necessary to have large number of sample plots and a long period of data collection to get the best estimation equations for future tree survival/mortality. Hence, either a relatively large number of permanent established sample plots or good quality yield tables have been considered the most reliable sources of data to analyze and estimate the tree survival (Deen, 1933; Krauch, 1930).

Later Vanclay (1994) classified the causes of tree mortality into three major groups: catastrophic, anthropogenic and regular. Catastrophic mortality includes the large scale tree mortality, caused by extraordinary events such as floods, storms, and insect-pests. Anthropogenic mortality is due to human activities such as harvesting, industrialization, and deforestation, and regular mortality is caused by tree age and competition, pests and diseases.

Accurate and reliable information about an individual tree mortality/survival is necessary in any stand growth system (Monserud, 1976). The more accurate the estimation of tree survival, the more accurate the overall growth prediction system will be, therefore the more reliable is the information about tree and stand growth (Allan et al., 2009) . According to Monserud (1976) and Hamilton and Edwards (1976), estimation of probability of individual tree survival from periodically remeasured data is difficult. This probability is often assumed to be constant for the measured interval, which may be several years in length. However during this interval there are changes in stand attributes such as stand age, density, and site quality as well as individual tree attributes such as diameter, height, and crown ratio which may alter the probability of survival.

Murphy (1986) and Clutter et al. (1983) stated that survival/mortality can be modeled at two levels; the stand level and at the individual tree level. In stand level models generally a few variables are used, and the model usually provides information about survival per acre/hectare on a stand-level basis. For example; age, density, site condition might be required in stand level survival model (Murphy, 1986). Stand level mortality or survival models estimate the probability of total number of trees dying or surviving in per unit area. These models utilize stand attributes as predictors (Lee, 1971), whereas individual-tree level survival models estimate the probability of survival of each tree on the basis of individual tree and stand attributes (Monserud, 1976). Additionally Belcher et al. (1982), Burkhart et al. (1987), and Zhang et al. (1997) explained that models which predict the growth and survival of individual trees in a forest stand are called individual-tree models. Specific information about individual tree variables such as tree age, DBH, and crown height can be obtained from individual tree growth and survival models. In individual tree forest growth models, these individual tree variables play a major role in predicting tree growth and survival.

In one approach to modeling forest mortality, mortality models have been developed at stand-level. Using stand-level attributes and estimates for the number of trees per acre at several ages, stand-level mortality functions may be developed. These stand level mortality functions may then be used to predict the number of mortality trees per unit area (acre or hectare) (Clutter et al., 1983; Ek, 1974; Lee, 1971). Improvement in computational techniques and discovery of new methodologies for parameter estimation made it possible to apply mortality models to predict the individual tree mortality beginning in the late 1960's and the 1970's (Yao et al., 2001). In individual tree survival models probability of survival of each individual tree is estimated whereas in stand level survival models, this is not done (Clutter et al., 1983). Hamilton (1974) and Monserud (1976) introduced the logistic function to model the individual tree mortality. Since then, logistic regression has been applied for many tree species in many geographic regions to model the tree survival or mortality (Hamilton, 1986).

Monserud (1976) explained that the techniques which are used in prediction of forest growth and yield are also useful in predicting survival or mortality. Additionally, yield tables can also be a major source of information concerning tree survival, since they provide information concerning trees surviving at various ages. Lee (1971) indicates that linear regression methods where tree age and diameter were used as independent variables were used to estimate the probability of survival/mortality using yield tables of lodgepole pine. Neter and Maynes (1970) recommended weighted non-linear regression or a multivariate maximum likelihood procedure to get the maximum likelihood estimation of the regression parameters of a non-linear function bounded by ‘0’ and ‘1’. These approaches can be applied to estimate the parameters of a logistic function for individual tree survival/mortality. Several other procedures such as neural networks and support vector methods (King et al., 2000), nonparametric classifiers (Dobbertin and Biging, 1998) have been used to estimate survival/mortality. Additionally, various regression models including linear regression (Keister, 1972; Krumland et al., 1977; Lee, 1971), the Weibull distribution function (Somers et al., 1980) and the Richards function (Buford and Hafley, 1985) and logistic regression (Monserud, 1976) have been used to predict the probability of individual tree survival/ mortality. Regardless model type, the survival/mortality model should have the best set of important predictor variables with the best possible parameter estimates and a model evaluation should be performed (Yao et al., 2001).

Compared to other methods such as discriminant analysis, probit analysis, or logit analysis, Monserud (1976) found a generalized logistic function to have the greatest discriminating power for predicting survival or mortality trees. The independent variables used in these comparison analyses were tree diameter and diameter increment, competition index, and length of growth period. Lynch et al. (1998) used a logistic function to develop a survival equation for individual shortleaf pine trees from even-aged natural forests. The data used to develop the model were from a permanently established plots located in eastern Oklahoma and western Arkansas.

2.2.2. Nonlinear Mixed-Effect Approach in Forest Modeling

Lappi and Bailey (1988) found the mixed modeling statistical procedure to be a better alternative method of estimation than the conventional methods for site index. They used a nonlinear mixed-effects growth curve model to predict heights of dominant trees both at the plot level and at the individual tree level. They provided an example of a mixed-effects model that has a multilevel design where random effects for plots and for trees found inside the plots enter linearly into a Chapman-Richards type growth model. Later in 2001 Daniel and Bailey generalized the approach of Lappi and Bailey by allowing multilevel random effects to enter into the model nonlinearly. Gregoire et al. (1995) applied mixed–effects modeling to account for the correlation from grouping in data structures that is generally found in forestry problems. They discussed nonlinear mixed models and pointed out the importance of mixed modeling in the forestry sector. In 1996, Gregoire and Schabenberger (1996a) developed a model for individual-tree cumulative bole volume of sweetgum from east Texas, using a nonlinear mixed-effects approach. Later, they formed a model for cumulative bole volume using the information on spatial correlation between sections of a tree bole (Gregoire and Schabenberger, 1996b). Lappi (1997) used a mixed-effects approach to develop a model that describes the DBH-height relationship. He developed models for plantation forests and natural stands of jack pine.

Wolfinger (1999) explained that statistical models that include both fixed and random effects of parameters which are associated nonlinearly to the response variable in the model have been widely used. He added that these types of nonlinear mixed models have very wide application in nonlinear growth curves and over-dispersed binomial data sets. Although these models can occur in various forms, the most common form is considered to be a conditional distribution for the response variable in which the random effect is given.

Daniel and Bailey (2001) presented the technique for estimating and predicting the parameters for forest growth variables using nonlinear modeling methods. In their research they incorporated estimation techniques for the variables which are subject to nested sources of variability. They considered multilevel nonlinear mixed-effects models that are useful for a variety of forestry applications and found that they have more advantages in growth and yield prediction than the other linearization based methods. Multilevel means the measurements are within trees and trees are nested within sample plots. For a two level case such as when the data are collected from number of trees within a single plot, it is generally easy to do integration using Gaussian quadrature or some other numerical technique to get the maximum likelihood estimate (Daniel and Bailey, 2001). For the two levels case PROC NLMIXED (Wolfinger, 1999) can be used to get an approximation to the maximum likelihood estimate of the parameters for models with tree specific random effects.

According to Lappi (2006) mixed models work better when items in the data sets occur in groups. Grouped datasets may contain longitudinal or repeated measurements or can be defined as multilevel or block designs (Pinheiro and Bates, 2000). Rose et al. (2006) applied a multilevel approach to estimate the probability of survival of individual trees. They mentioned that the data from permanently established plots consist of various sources of heterogeneity due to multilevel structure and repeated measurements. Trincado and Burkart (2006) used mixed models with the tree-level random effects in a loblolly pine individual tree taper model in which they found that violation of the assumption of correlated errors was mitigated by inclusion of random effects.

A DBH-height relationship for shortleaf pine was developed by (Budhathoki et al., 2008b) using a mixed effects model method where plots were used as random-effects. They concluded that mixed-effects models provide improved parameter estimates when partially accounting for spatial and temporal correlation.

2.3. Rationale of the study

In forest management growth and yield models are of great importance. Prediction of the probability of survival of an individual tree is essential when considering growth and yield of a stand. Survival models are important components of forest growth prediction.

In 1970s the first individual tree survival or mortality model which used the logistic function was developed by Hamilton (1974). Using the logistic function as mortality model for thinned and unthinned mixed conifer stands of northern Idaho was developed by (Hamilton, 1986) . Additionally, Avila and Burkhart (1992) developed a survival model for loblolly pine trees in thinned and unthinned plantations throughout the area in the United States where loblolly pine plantation management is in practice. Similarly Yao et al. (2001) developed a mortality model for individual trees such as aspen, white spruce, and lodgepole pine using the data from Alberta mixedwood forests. Temesgen and Mitchell (2005) formulated an individual tree mortality model for complex stands in southeastern British Columbia. They used the generalized logistic model in their study. Murphy and Shelton (1996) developed a survival model for individual loblolly pine (*Pinus taeda L.)* in uneven-aged stands. However no significant research work has been done in developing an individual tree survival/mortality model especially for shortleaf pine in even-aged natural forest system since the work of Lynch et al. (1999a).

Shortleaf pine is economically and commercially a highly valuable species (Lynch et al., 1999b). It is distributed widely in various forest systems within its natural range. Naturally occurring shortleaf pine forests are commonly found on private nonindustrial lands and on industrial forestlands of western Arkansas and eastern Oklahoma (Lynch et al., 1999a). Shortleaf pine forests provide habitat for a variety of wildlife species and are very important in human uses such as in constructing storage boxes, structural material, plywood, and pulpwood. Because of its wide occurrence and importance it is desirable to develop a survival model with the best possible set of independent variables that predicts the probability of survival for individual shortleaf pine trees for even-aged natural forests of Oklahoma and Arkansas.

The growth and yield equation of uneven-aged shortleaf pine stands developed by Murphy and Farrar (1985) only predicts net yields but does not provide specific information concerning the estimation of survival for shortleaf pine (Lynch et al., 1999a). According to Gertner (1989) a survival model is needed in a forest growth system which predicts the survival of trees on individual basis or on a stand basis. He further added that error propagation and budgeting analysis showed that growth prediction depends significantly on the underlying mortality/ survival model.

A complete individual tree growth system includes models for variables such as basal area increment, height increment, and crown ratio. Hamilton (1986) emphasized that the survival or mortality analysis must be undertaken while developing a growth and yield model. Lynch et al. (1998) explained growth and yield information on naturally occurring shortleaf pine stands can be obtained from stand-level tables or equations however these do not provide sufficient information about probability of survival for individual trees. The combination of survival models and individual-tree diameter growth, height growth, and ingrowth allows us to predict forest stand development over time (Avila and Burkhart, 1992; Hamilton, 1990; Teck and Hilt, 1990).

Since the work of Lynch et al. (1998) no models for individual tree survival have been developed in even-aged shortleaf pine natural stands until the initiation of the current project. This research work includes the measured and remeasured data from the plot establishment period (1985-1987) until the fourth remeasurement time (Fall 2000-2001). Because this research work accounts for a large data set from very long period, it is an important source of information to develop a survival model for the shortleaf pine forest system of Arkansas and Oklahoma.

This study also focuses on applying a nonlinear mixed-effects model in developing a survival model. Many forest growth models have been developed using data from permanently established plots. The study also examines the performance of nonlinear mixed effects modeling with a plot-level random effect.

A major goal of this study is to apply the logistic model and estimate the parameter through the iteratively reweighted nonlinear regression modeling approach in developing a survival model of individual trees and to improve the methods for obtaining parameters of individual tree survival equations from periodic measurements. Although the model is developed for predicting individual tree survival of shortleaf pine, the method applied should readily adaptable to other species.

2.4. Objectives

The major objective of the study was to develop an equation that can be used to estimate the survival probability of naturally-occurring, even-aged shortleaf pine trees. Additionally, specific objectives were to find the best set of independent variables for predicting individual tree survival, to apply the logistic function to model the probability of survival, to apply the nonlinear mixed modeling approach to examine its performance on fitting the model and finally to validate the model.

2.5. Materials and Methods

2.5.1. Study Area

Study plots are located in the Ozark and Ouachita National forests in western Arkansas and southeastern Oklahoma as indicated on Figure 2 below.

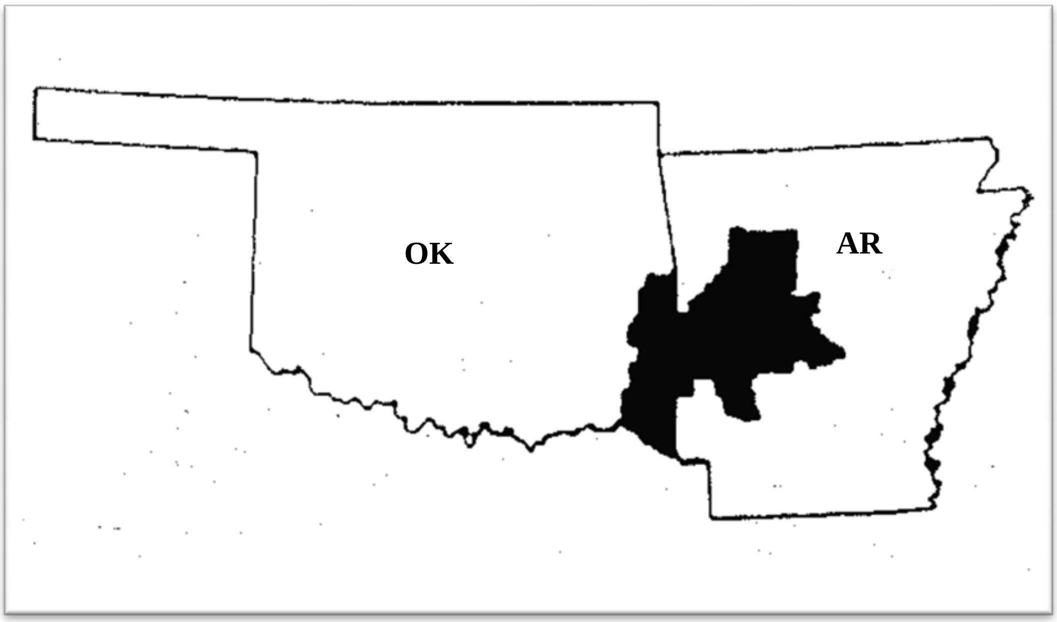

Figure 2. The study area.

Source: An Individual-Tree DBH-Total Height Model with Plot Random-Effects for Shortleaf Pine-PowerPoint presentation – Dr. Thomas B. Lynch

2.5.2. Data

Lynch et al. (1998) explained that until 1985 the major sources of data for the growth and yield of naturally occurring shortleaf pine forests were from fully stocked plots or from unmanaged stands. Considering this, during the period of 1985-1987, the Department of Forestry (now part of the Department of Natural Resource Ecology and Management) at Oklahoma State University and USDA Forest Service Southern Research Station at Monticello, Arkansas collaboratively established growth and yield plots in even-aged natural shortleaf pine stands that were located in the Ozark and Ouachita National Forests. These plots were selected to represent a range of ages, densities and site qualities. The resulting sample of over 200 plots were permanently established in shortleaf pine natural stands located in the Ozark and Ouachita National Forests and were distributed from areas north of Interstate Highway 40 near Russellville in western Arkansas to near Broken Bow in Southeastern Oklahoma. Measurements of individual shortleaf pine tree total height, crown height, and diameter at breast height (DBH) were taken and were used to develop a shortleaf pine survival model. Plots have been remeasured in every 4 to 6 years, and individual tree survival or mortality was recorded at each measurement.

Growth plots were established considering a design criteria that includes basal area (ft^2/acre), Site index (ft at age 50 yrs), and Age (yrs) as design variables. The following stand properties described by Rose (1998) were considered as a guide while establishing the growth plots:

1. Natural forest stand that consists of at least 70% of basal area occupied by Shortleaf pine and trees with at least 0.6 inches DBH.
2. Stands having healthy dominant and codominant trees with a maximum range in age of no more than 10 years.
3. Variation in site index should be less than 10 feet.
4. Even-aged stands with no more than two age classes per plot.
5. No harvesting in the last 5 years.

Table 1 below shows class midpoints and ranges for basal area (ft^2/acre), site index (total height in feet at age 50 years) and age (plot age in years) which were used as design variables for the study. Four classes of basal area, age and site index were established. The original design specified three plots in each combination of age, site index, and basal area, however, this was not accomplished for all combinations. Additional plots from a thinning experiment previously established by Frank Freese of the USDA Forest Service Southern Forest Experiment Station were incorporated into the study.

Table 1. Design variables with class midpoints and ranges for plots located in natural, even-aged shortleaf pine forests in western Arkansas and southeastern Oklahoma (Lynch et. al. 1998).

Design Variable	Class midpoint	Class range
Basal area (ft^2/ac)	30	16-45
	60	46-75
	90	76-105
	120	106-135
Site Index (ft at age 50 yr)	<56	<56
	60	56-65
	70	66-75
	>75	>75
Age (yr)	20	11-30
	40	31-50
	60	51-70
	80	71-90

2.5.3. Growth Plot

Each sample measurement plot is circular with a radius of 57.2 feet and 0.2 acres in size (Figure 3). Each plot is surrounded by a 33-foot isolation boundary. The outer boundary has been painted with white bands on trees just outside the boundary. A buffer strip 33 foot wide bordering the isolation boundary perimeter was constructed to eliminate edge effect. The same silvicultural treatments have been applied to isolation strip as to the measurement plot. The boundary of the buffer strip is marked by blue-painted bands on trees just outside the boundary. This boundary often also surrounds a group of adjoining plots.

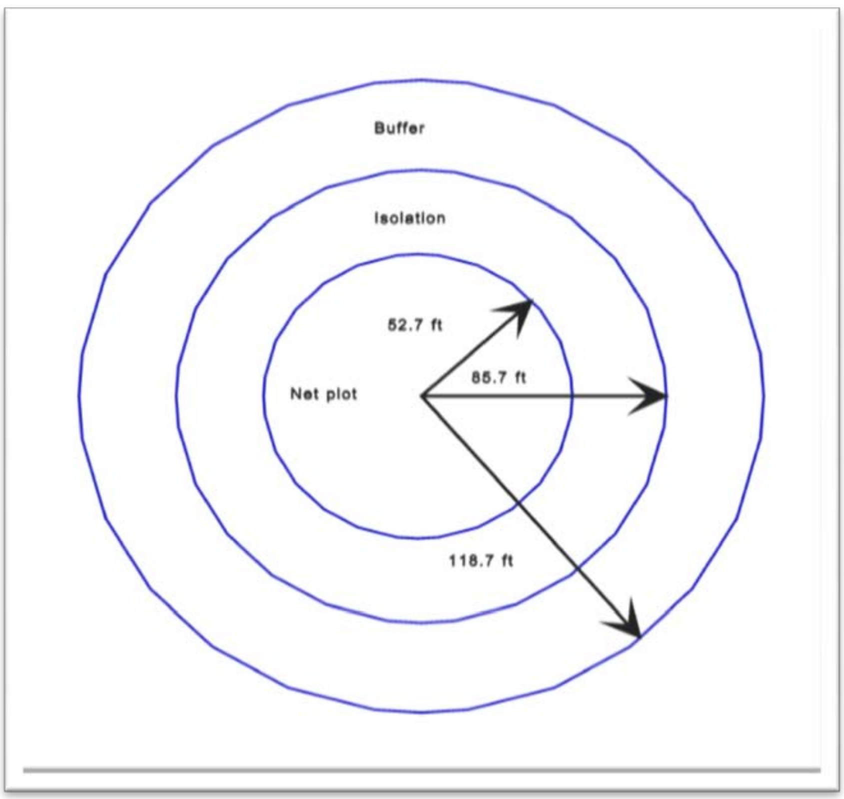

Figure 3. Natural, even-aged short leaf pine growth plot.

Source: An Individual-Tree DBH-Total Height Model with Plot Random-Effects for Shortleaf Pine-PowerPoint presentation – Dr. Thomas B. Lynch

The study plots were chosen using aerial photographs and field reconnaissance. Figure 4 shows the study plot locations within the study area.

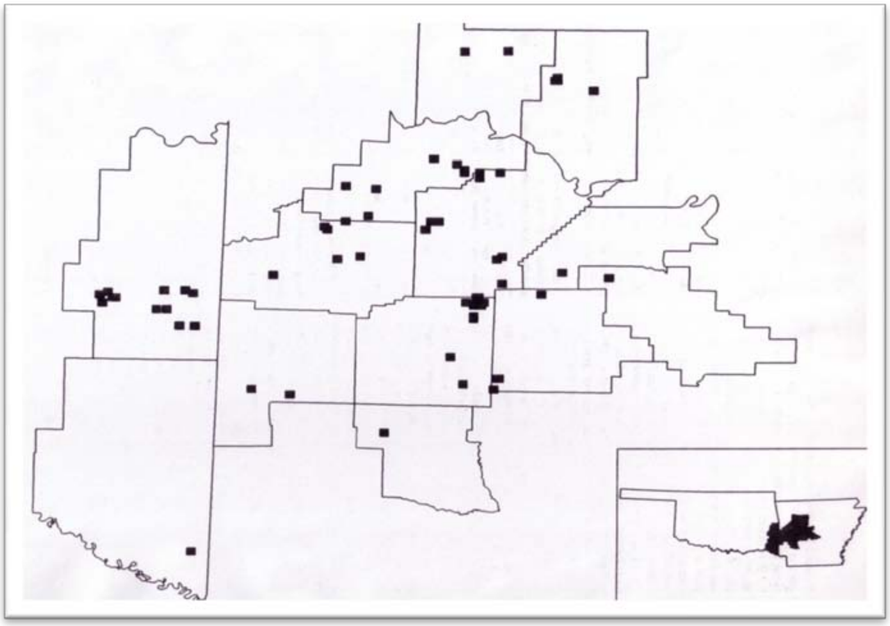

Figure 4. Location of study plots in western Arkansas and southeastern Oklahoma.

Five milacre plots were established within each growth plot for understory measurement. Figure 5 depicts 5-milacre subplots within a net growth plot.

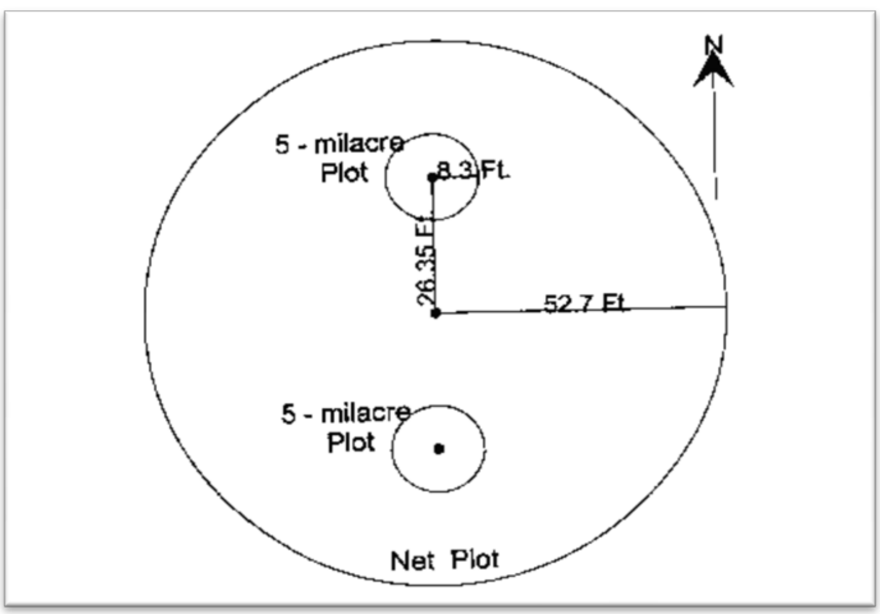

Figure 5. 5-milacre sub plots within a net growth plot.

Each growth plot was thinned from below to a predetermined residual pine basal area at plot establishment and chemical herbicide was applied to control competing vegetation. For every tree in a plot, tree number and breast height (4.5 feet above the ground) were marked and DBH (inches) was measured. Total height (feet) and height to the base of the live crown (feet to bottom live branch) were recorded for a specified sample of trees in each DBH class on the measurement plot. Each of the measured trees were classified as dominant, co-dominant, intermediate, or suppressed. Increment cores were used to Determine the age of trees and a Suunto Clinometer was used to measure height.

2.5.4. Growth plot measurement schedule

The first baseline measurement was obtained from fall 1985 to fall 1987 during the period of plot establishment. The azimuth and distance of each tree from the center of plot were recorded. At the initial measurement, each tree was labeled with a number. DBH was measured, and the DBH measurement point was marked.

From fall 1990 to fall 1992 the first remeasurement was performed. Tree height, DBH, live crown were measured crown class was determined, damage codes were recorded and trees were re-painted and numbered. The second remeasurement was carried out during a period from fall 1995-fall 1997. Again, DBH, height, live crown, were measured and crown class was determined. Also numbers and boundaries were re-painted and damage codes were recorded. At this time many plots were marked for thinning to their original establishment densities, while some were left unthinned. Thinning was performed in the designated plots prior to the next growing season. During fall of 2000, the third remeasurement was performed. There was strong ice-storm in December, 2001 which damaged many trees. For the purposes of this study, only trees showing no ice damage were used. The fourth remeasurement was conducted during the period from fall of 2006 until winter, 2008.

2.5.5. Statistical Analysis

The survival data were modeled by Equation 2 using the SAS/ LOGISTIC procedure (SAS Institute, Inc. 2007). "A usual logistic regression model, proportional odds model and a generalized logit model can be fit for data with dichotomous outcomes, ordinal and nominal outcomes, respectively, by the method of maximum likelihood (Allison, 2001)". A stepwise procedure in PROC LOGISTIC was used to select the best set of predictor variables.

The following SAS statements were used to perform logistic regression:

```
PROC LOGISTIC DATA=LOGISTIC;
MODEL SURV(EVENT='1')=   Mid-BA PERIOD DRINV DBHSQ Mid-BA*DRINV
Mid-BA*DBHSQ DRINV*DBHSQ Mid-BA*DRINV*DBHSQ/
SELECTION=STEPWISE SLENTRY=0.05 SLSTAY=0.05;
RUN;
```

The model statement includes the response variable and the independent (explanatory) variables. SURV is dependent variable and event='1' indicates that the model is for survival trees which are coded as 1, while mortality trees are coded as zero. Mid basal area per acre (Mid-BA), time period (PERIOD), inverse of ratio of quadratic mean diameter to DBH (DRINV), square of DBH (DBHSQ) and their combinations were used as independent variables. The SELECTION statement specifies the variable selection method. SLENTRY and SLSTAY represent respectively, level of significance for entering and removing the variables which was 0.05.

2.5.5.1. Likelihood Function

The likelihood function is a function of the parameters of a statistical model that plays a key role in inferential statistics. It indicates how likely an observed sample is from a particular population. For the logistic model the log -likelihood function is stated as:

Case 1 - Simple Logistic Model:

$$\text{Log}_e\, L\,(\beta_0, \beta_1) = \sum^{n}_{i=1} Y_i(\beta_0 + \beta_1 X_i) - \sum^{n}_{i=1} \log_e [1+\exp(\beta_0 + \beta_1 X_i)] \qquad (3)$$

Case 2- Multiple Logistic Model:

The log-likelihood function for multiple logistic regression is an elaboration of log-likelihood function of simple logistic regression in (3) (Kutner et al., 2004):

$$\text{Log}_e\, L\,(\boldsymbol{\beta}) = \sum^{n}_{i=1} Y_i\,(\mathbf{X}_i'\,\boldsymbol{\beta}) - \sum^{n}_{i=1} \log_e [1+\exp(\mathbf{X}_i'\,\boldsymbol{\beta})] \qquad (4)$$

where, $\mathbf{X_i}'\,\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1}$

$\boldsymbol{\beta}' = [\beta_0, \beta_1 \ldots, \beta_{p-1}]$

$Y_{\mathbf{i}}$ = a binomial (0 or 1) dependent variable

2.5.5.2. Method of Maximum Likelihood Parameter Estimation

Kutner et al. (2004) described the maximum likelihood estimates of the parameters as the values of parameters that maximize the log-likelihood function. The maximum likelihood estimate method is very useful in estimating the parameters when the dependent (response) variable is 0 or 1. Ragavan (2008) stated for a large sample size the distributions of maximum likelihood estimators are approximately normal, and the estimators are asymptotically unbiased estimators. The SAS/LOGISTIC procedure uses a numerical procedure such as Newton Raphson to find the maximum likelihood estimates of parameters for the logistic function (Kutner et al., 2004).

2.5.5.3. Iteratively Reweighted Least Squares

Iteratively reweighted nonlinear regression was used in this study to satisfy the regression assumption homogeneity of variance. Since survival is a binary or Bernoulli random variable, it has variance P(1-P) where P is the probability of success or survival. Therefore, the appropriate weight is defined as the inverse of the variance $P^t(1-P^t)$ where 'P' is annual probability of survival from the logistic model and 't' is the number of years in the measurement period. The model is raised to the power of 't' to obtain the probability of survival for a measurement interval of 't' years for remeasurement plots measured at intervals longer than one year. Maximum likelihood estimates are obtained when the above weight is used.

McCullagh and Nedler (1989) recommended iteratively re-weighted regression as an effective procedure to find the maximum likelihood estimates for the survival model in Equation 2. Monserud and Flewelling (2002) stated that weighted least squares methods minimize the weighted sums of error squared, using weights that are inversely proportional to the estimated variance. It is shown by McCullagh and Nelder (1983) when the weighted sums of squares are minimized, the maximum likelihood estimates are obtained.

The following statement was invoked in SAS/STAT software version, 9.2 (SAS Institute Inc. 2007) to perform iteratively reweighted nonlinear regression. This procedure estimates the parameters in Equation 2 where X_1=Mid-BA, X_2 = DRINV, X_3 = DBHSQ and X_4=Mid-BA*DRINV and n=k=4.

PROC NLIN DATA=NLIN NOHALVE METHOD=MARQUARDT;

MODEL SURV= (1/ (1+exp (-(b_0 + b_1*Mid-BA + b_2*DRINV + b_3*DBHSQ

+ b4*Mid-BA*DRINV))))**PERIOD;

PARMS b0= -1.711 b1= -0.0147 b2=0 b3= -0.0000134 b4=0;

WEIGHT= ((MODEL.SURV)*(1-MODEL.SURV))**(-1);

OUTPUT OUT = CHITEST1 PREDICTED = PRESURV;

where

PARMS : identifies starting values for parameter estimates

MODEL: defines the algebraic relationship between the dependent and independent variables (the mean function).

OUTPUT OUT=CHITEST is the output data set name which contains results of the estimation procedure and other statistics calculated for each observation.

PREDICTED=PRESURV provides the predicted (expected) value for survival probability for each observation.

PROC NLIN is the major SAS procedure for nonlinear (or curvilinear) regression analysis. The NLIN procedure fits nonlinear regression models and estimates the parameters by nonlinear least squares or weighted nonlinear least squares (SAS Institute

Inc. 2007). Estimating parameters in a nonlinear model is an iterative process that commences from some starting values provided. For the NLIN procedure the parameters and some initial values need to be specified. The homoscedasticity assumption of regression can be relaxed by using a weighted residual sum of squares criterion. However, the assumption of uncorrelated errors (independent observations) cannot be relaxed in the NLIN procedure.

2.5.5.4. NLMIXED Procedure

The NLMIXED procedure in SAS (SAS Institute, Inc. 2007) was used to fit the nonlinear mixed models. NLMIXED fits the models using likelihood-based methods. The procedure also takes into account the random effects in the model. Both fixed and random effects are allowed to have a nonlinear relationship to the response variable.

PROC NLMIXED enables one to specify a distribution (binary in this study) for the response variable. PROC NLMIXED maximizes an approximation to the likelihood integrated over the random effects. Although various types of integral approximations are available, Gaussian quadrature and a first-order Taylor series are considered to be the principal approximation methods. A variety of alternative optimization techniques are available to carry out the maximization; the default is a dual quasi-Newton algorithm. It was desired to estimate parameters in the following model, which is similar to Equation 2 except for the inclusion of the plot-level random effect u_k:

$$P_{jk}^{t} = [\, 1 + \exp\{-(\beta_0 + \beta_1 X_{1jk} + \beta_{2j} X_{2jk} + \beta_{3j} X_{3jk} + u_k)\}]^{-t} + \varepsilon_{jk} \qquad (5)$$

where

X_{1j}=Mid-BA for tree j on plot k

X_{2j}=DRINV for tree j on plot k

X_{3j} = Mid-BA*DRINV for tree j on plot k

u_k = plot-level random effect distributed normally with mean 0 and variance σ^2_u

ε_{jk} = error term with mean zero

The statements used to estimate parameters in the nonlinear mixed model are:

DATA NLMIXED;

PROC SORT DATA=NLMIXED out=NLMIXED1;

BY PLOTID;

PROC NLMIXED DATA=NLMIXED1 QPOINTS=20 TECH=NEWRAP;

PARMS b0= 2.1963 b1= -0.0145 b2=4.9246 b3=0 s2u=0.5;

ETA=b0 + b1*Mid-BA + b2*DRINV + b3*Mid-BA*DRINV +u;

m=exp (ETA);

P= ((m/(1+m))**PERIOD);

MODEL SURV~binary(P); RANDOM u~normal(**0**, s2u)subject=PLOTID;

RUN;

where TECH = NEWRAP invokes a Newton-Raphson optimization. The model statement specifies a binary (Bernoulli) distribution with probability P, the RANDOM statement defines u to be the random effect with subjects defined by the PLOT variable, and the subject=PLOTID option models the random effect that changes with plot variable.

2.5.6. Model Evaluation

Once the fitted response function was obtained, a goodness-of-fit test was conducted to check the fitness of the response function. This test is important because a variety of inferences and predictions will be based on the fitted response function.

2.5.6.1. Chi-square Goodness-of-Fit Test

Measures of goodness-of-fit such as the correlation coefficient (r) or coefficient of determination (R^2) are not appropriately applied to binary variables. Instead, a Chi-square test is often used to evaluate the appropriateness of the model (Neter and Maynes, 1970).

After obtaining parameter estimates, comparisons between observed and predicted numbers of live and dead trees were made by DBH classes. Two inch diameter classes were created to classify the tree based on diameter outside bark measured at DBH. The model was evaluated using the χ^2 goodness-of-fit test over the predictor variable using the data set. The χ^2 model evaluation process was formulated as:

Hypothesis formulation for model evaluation

$H_0 : E[Y] = \{ 1 + \exp(-\mathbf{X}_i' \boldsymbol{\beta}) \}^{-1}$ Model fits well

$H_0 : E[Y] \neq \{ 1 + \exp(-\mathbf{X}_i' \boldsymbol{\beta}) \}^{-1}$ Model does not fit well

Test Statistic

$$\chi^2 = \sum_{j=1}^{c} \sum_{k=0}^{1} (O_{jk} - E_{jk})^2 / (E_{jk})$$

where

O_{j1} = Observed number of surviving trees in diameter class j

O_{j0} = Observed number of mortality trees in diameter class j

E_{j1} =Number of surviving trees in diameter class j expected from the model

E_{j0} =Number of mortality trees in diameter class j expected from the model

c= number of diameter classes

Decision Process

If $\chi^2_{calculated} \geq \chi^2_{tabulated\ (1-\alpha,\ c-2)}$ Reject H_0

If $\chi^2_{calculated} \leq \chi^2_{tabulated\ (1-\alpha,\ c-2)}$ Failed to Reject H_0

2.6. Results and Discussion

2.6.1. Logistic Regression and Parameter Estimation using Iteratively Reweighted Nonlinear Regression

Stepwise variable selection procedure in logistic regression using SAS PROC LOGISTIC (SAS Institute, Inc. 2007) was used to obtain the best set of independent variables. The following variables were significant at alpha level 0.05: mid basal area per acre (Mid-BA), time period (PERIOD), inverse of ratio of quadratic mean diameter to DBH (DRINV), square of DBH (DBHSQ) and their interaction terms. All 2 way and 3 way interaction among the variables were found significant. Table 2 provides the information about the variables selected using stepwise logistic regression. Other variables such as DBH, square root of DBH, 1/DBH, square of DBH/basal area, dominant height, plot age, site index and crown ratio were examined , however, they were either found to be insignificant, or they did not contribute to improvement in the chi-square test used for final model evaluation.

Table 2. Significant variables and their parameter estimates using logistic regression with PROC LOGISTIC.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	7.4237	0.9353	63.0014	<.0001
Mid-BA	1	-0.089	0.00639	193.9688	<.0001
PERIOD	1	0.4552	0.0878	26.8969	<.0001
DRINV	1	-4.6564	1.022	20.7568	<.0001
DBHSQ	1	-0.0672	0.00941	50.9055	<.0001
Mid-BA*DRINV	1	0.0812	0.00786	106.7127	<.0001
Mid-BA*DBHSQ	1	0.000704	0.00008	76.4281	<.0001
DRINV*DBHSQ	1	0.055	0.0091	36.5027	<.0001
Mid-BA*DRINV*DBHSQ	1	-0.0006	0.000076	63.4838	<.0001

Table 3 below provides information concerning the variables used for the analysis including number of observations of each variable, their mean, minimum value, maximum value and standard deviation.

Table 3. Summary statistics of the significant variables used to fit parameters to logistic survival model.

Variable	Number	Mean	Minimum	Maximum	Standard deviation
Mid-BA (ft^2/acre)	599	86.02	16.09	186.59	35.75
DRINV	20283	0.97	0.13	2.67	0.24
DBHSQ (inch squared)	20283	82.58	1.21	645.16	76.47

Although logistic regression was used to help suggest a best set of independent variables, parameters estimated through logistic regression cannot be used to estimate annual probabilities of survival because the measurement periods in this study were all longer than one year, and differed in length. Since the remeasurement periods were 4, 5 or 6 years for this study, parameters of significant variables were estimated using iteratively reweighted nonlinear regression in a logistic model raised to the power of period length. These measurement period lengths require the use of Equation 2 in which period length is variable. PROC LOGISTIC does not allow variable period lengths. Estimates of the parameters are shown in the Table 4 below. Parameter estimates given in Table 4 maximize the log likelihood function presented in Equation 4.

Table 4. Parameter estimates for the significant variables using iteratively reweighted nonlinear regression for a logistic model.

Variable	Parameter	Estimate	Standard Error	95% Confidence Limits	
Intercept	β_0	9.791	0.7907	8.2412	11.3409
Mid-BA	β_1	-0.0741	0.00575	-0.0853	-0.0628
DRINV	β_2	-4.0594	0.918	-5.8587	-2.2601
DBHSQ	β_3	-0.0025	0.000947	-0.00431	-0.0006
Mid-BA× DRINV	β_4	0.0736	0.00684	0.0602	0.087

2.6.2. Model formulated using Logistic Regression and Iteratively Reweighted Nonlinear Regression

The final model formulated to estimate the annual probability (t=1) of survival of individual shortleaf pine trees is:

$$APOS= (1/ (1+ (exp-(9.791 - 0.0741 \times \text{Mid-BA} - 4.0594 \times DRINV - 0.0025 \times DBHSQ + 0.0736 \times \text{Mid-BA} \times DRINV)))) \quad (6)$$

where APOS is the annual probability of survival.

The interaction between mid-basal area per acre (Mid-BA) and inverse of ratio of quadratic mean diameter to DBH (DRINV) was found significant in iteratively reweighted nonlinear regression. The interaction effects represent the combine effects of these two variables on the response variable or on the survival probability of each individual tree. The addition of interaction term in the model drastically changes the interpretation of all of the coefficients. Since the interaction term is significant the main effect due to Mid-BA and DRINV cannot be independently interpreted. The impacts on the survival probability of one of the two variables depend on the level (value) of the other variable. The interaction effects indicate the effect of Mid-BA on survival probability is different on different values of DRINV and vice-versa. Hence the unique effects of Mid-BA is not limited to β_1 (-0.0741) but also depends on the values of β_4 and DRINV. The unique effect of Mid-BA is represented by everything that is multiplied by Mid-BA in the model: $\beta_1 + \beta_4$*DRINV i.e. Mid-BA (– 0.0741 + 0.0736 *DRINV). Similarly the unique effect of DRINV is represented by DRINV (-4.0594 + 0.0736 × Mid-BA).

Moreover, the sign of the coefficients of one of the variables, DRINV or Mid-BA, changes depending upon the value of the other independent variable. The coefficient of Mid-BA depends upon the value of DRINV. The following formulation can be used to determine when DRINV implies a positive or negative sign associated with Mid-BA:

Mid-BA (– 0.0741 + 0.0736 *DRINV) =0

DRINV$= \frac{0.0741}{0.0736} \approx 1$. When DRINV is greater than 1 then the coefficient of Mid-BA is positive and when the value of DRINV is smaller than 1 then the coefficient of Mid-BA is negative.

Furthermore, the following formulation determines when a value of Mid-BA would imply a positive or negative coefficient of DRINV:

DRINV (- 4.0594 + 0.0736 × Mid-BA) = 0

Mid-BA $= \frac{4.0594}{0.0736} \approx 55$ (ft^2/acre). Hence, the coefficient of DRINV is positive when Mid-BA is greater than 55 and is negative when Mid-BA is less than 55.

Higher density forest stands with the mid-basal area per acre greater than 55 ft^2 result in a positive coefficient of DRINV. When the coefficient of DRINV is positive, the survival probability of each individual tree is increased with increase in DRINV. According to Monserud and Sterba (1999) the DBH of a tree provides significant information about the tree's size. They explained that large DBH trees have high survival rates because this size increases the strength of the tree for competing for sunlight, nutrients and other requirements of growth. From the final model in this study DBH has the same role in predicting an individual tree survival when basal area is above 55 ft^2/acre. Under these circumstances when the DBH of a tree is smaller that makes the inverse of ratio of quadratic mean diameter to DBH is smaller than 1 which in turn decreases the probability of survival for that tree.

On the other hand, when DRINV is less than 1 the coefficient of Mid-BA is negative. And, when the coefficient Mid-BA is negative then increase in basal area in a stand causes the probability of survival of individual shortleaf pine trees to decrease. The sign of the DBHSQ indicates that the survival probability is reduced with increasing DBHSQ, if other variables are held constant.

2.6.3. Correlation Matrix

Table 5 below shows strong correlation between b_1 (coefficient of Mid-BA) and b_2 (coefficient of DRINV), b_1 and b_4 (coefficient of Mid-BA*DRINV) and b_2 and b_4 . The highest correlations are highlighted below. This shows a possible multicollinearity problem in the independent variables. Despite the high correlation the interaction variable

Mid-BA*DRINV contributed to a substantial reduction in the chi-square statistic used to evaluate model fit, therefore it was decided to retain this interaction variable even though correlations with other variables in the model were high.

Table 5. Correlation matrix for coefficients in a shortleaf pine survival model.

	b0	**b1**	**b2**	**b3**	**b4**
b0	1	-0.9742	-0.9547	0.16964	0.91415
b1	-0.9742	1	0.93471	-0.101	-0.9576
b2	-0.9547	0.93471	1	-0.3455	-0.957
b3	0.16964	-0.101	-0.3455	1	0.20451
b4	0.91415	-0.9576	-0.957	0.20451	1

2.6.4. Model Evaluation using chi-square goodness-of-fit test

The test was used to evaluate how well the logistic model fits observed data. The test involves comparison of expected versus observed number of survival and mortality trees.

Hypothesis:

H_0: E [Y] = (1/ (1+ (exp-(9.791 – 0.0741 × Mid-BA - 4.0594 × DRINV – 0.0025 × DBHSQ + 0.0736 × Mid-BA × DRINV)))) The logistic model fits well for the observed survival and mortality data

H_1: E [Y] ≠ (1/ (1+ (exp-(9.791 – 0.0741 × Mid-BA - 4.0594 × DRINV – 0.0025 × DBHSQ + 0.0736 × Mid-BA × DRINV)))) The logistic model does not fit well for the observed survival and mortality data

Test Statistics:

$\chi^2_{calculated} = 23.41$

$\chi^2_{tabulated\ (1-\alpha,\ c-2)} = \chi^2_{0.95,\ 11} = 19.68$ where $\alpha = 0.05$

p-value is 0.0154 which is smaller than $\alpha = 0.05$.

Decision: $\chi^2_{calculated} \geq \chi^2_{tabulated\ (1-\alpha,\ c-2)}$, Reject the Null Hypothesis.

Conclusion: It was found that the logistic model formulated in Equation (6) does not fit well for the observed survival and mortality data at 5% level of significance.

However, at alpha level of 0.01 $\chi^2_{tabulated\ (1-\alpha,\ c-2)} = \chi^2_{0.99,\ 11}$ =24.73. In this case the calculated value is less than the tabulated value hence failed to reject the null hypothesis. Also the p-value is bigger than $\alpha = 0.01$ hence failed to reject the null hypothesis that the model fits well.

The chi-square values for survival for each diameter class are very low according to Table 6 below. However this is not the case for mortality. Therefore, it can be observed that the model for individual tree survival rates fits better than the model for mortality. Total survival chi-square value for over all diameter classes is 1.2711 which can be considered as low.

On the other hand chi-square values for mortality for some diameter classes such as 2, 6, 8, 14 etc. seem to be very high. For example, chi-square values for diameter classes 6 and 8 are 3.48181 and 6.29232 respectively which are quite high (Table 6). Total chi-square value for mortality for over all diameter class is 22.13761 which is very large compared with the total chi-square value from survival model. The model is rejected by the goodness-of-fit test at the significant level of 0.05 because of these high mortality results in some of those diameter classes.

Examination of the table below and Figure 6 demonstrates that the number of surviving trees are in increasing order from diameter class 2 to diameter class 6 and are gradually decreasing in higher diameter classes and much less in 18, 20 and so on. The chi-square values in Table 6 represent survival and mortality chi-square value for 4, 5 or 6 years since the data are from measurement intervals of 4, 5 or 6 years.

Table 6. Observed survival and mortality and expected survival and mortality by 2-inch DBH class with chi-square values for the plot remeasurement period from a logistic model fitted by iteratively reweighted nonlinear regression.

Diameter Class	Total Trees	Survived			Mortality		
		Observed No.	Expected No.	Chi. Square	Observed No.	Expected No.	Chi. Square
2	1451	1109	1138.74	0.77683	342	312.258	2.83295
4	3869	3582	3583.63	0.00074	287	285.374	0.00927
6	3886	3791	3770.99	0.10619	95	115.011	3.48181
8	3320	3265	3242.99	0.14943	55	77.013	6.29232
10	2801	2740	2735.32	0.00799	61	65.675	0.33282
12	2271	2211	2221.01	0.04516	60	49.985	2.00657
14	1541	1496	1506.72	0.07622	45	34.283	3.3499
16	706	686	689.3	0.01579	20	16.7	0.65189
18	299	294	291.46	0.02208	5	7.537	0.85376
20	110	109	106.87	0.04247	1	3.131	1.44996
22	24	24	23.23	0.02584	0	0.775	0.77472
24	4	4	3.9	0.00232	0	0.095	0.09511
26	1	1	0.99	0.00004	0	0.007	0.00653
Total	**20283**	**19312**	**19315.15**	**1.2711**	**971**	**967.844**	**22.13761**

Figure 6 reveals that the observed numbers of live trees are very close to the expected (predicted) numbers of live trees for each diameter class. Conversely, Figure 7 shows substantial differences in observed and expected numbers of mortality trees. Observed numbers of mortality trees are higher than the expected numbers in some DBH classes such as in 2, 12 and 14 diameter classes. And, observed numbers of mortality trees are less than the expected numbers in DBH classes such as 6, 8 and 10. Since there is not much fluctuation in observed and expected number of survival trees (Figure 6), it can be concluded that the model for survival behaved better than the model for mortality.

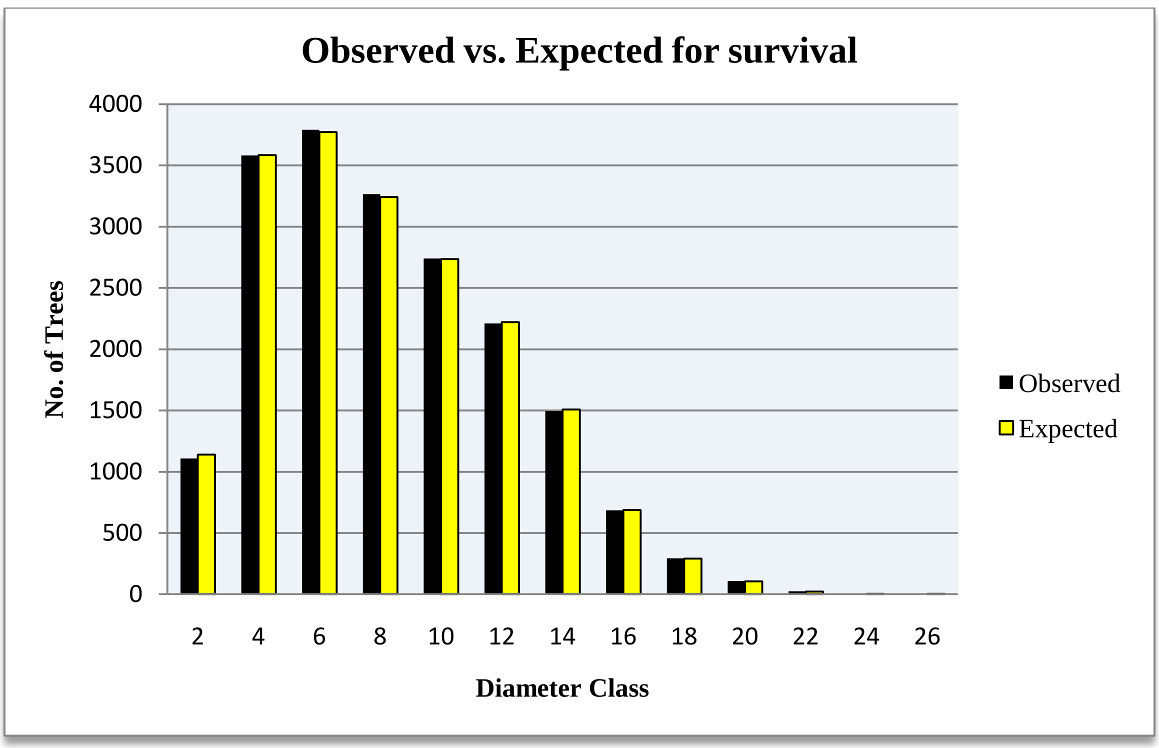

Figure 6. Observed and expected number of surviving shortleaf pine trees for each 2-inch diameter class for the remeasurement interval from a logistic model fitted by iteratively reweighted nonlinear regression.

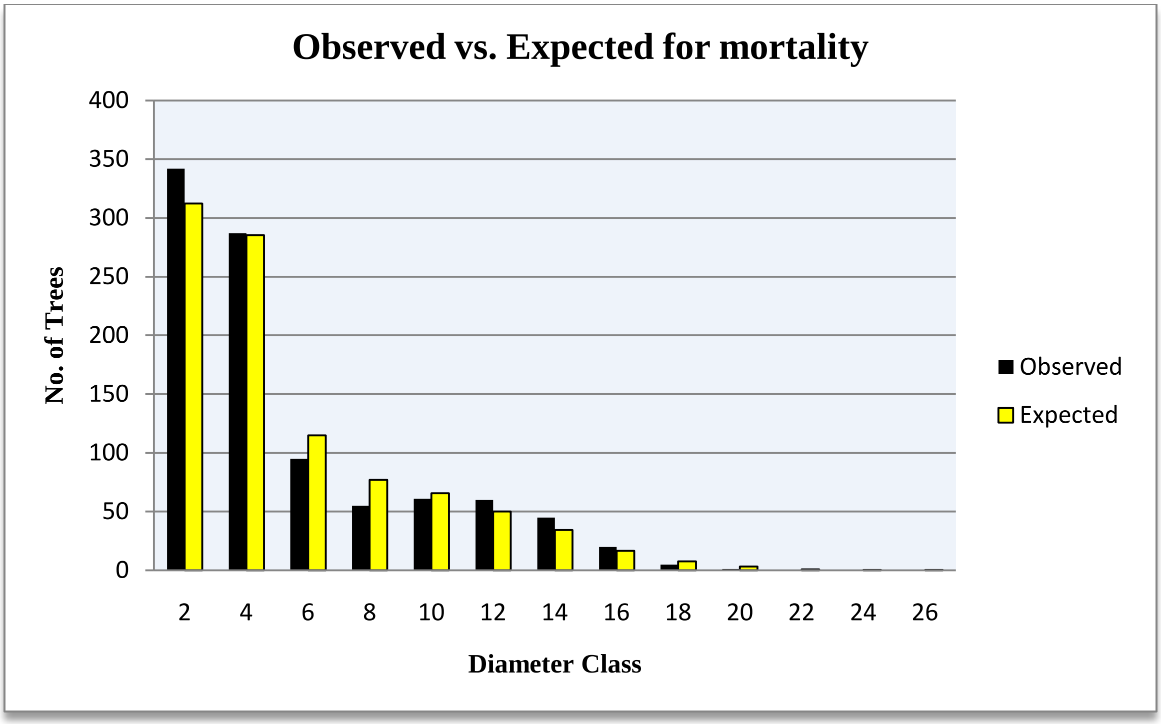

Figure 7. Observed and expected number of shortleaf pine mortality trees for each 2-inch diameter class during the remeasurement interval from a logistic model fitted by iteratively reweighted nonlinear regression.

2.6.5. Logistic Regression and Parameter Estimation using a Nonlinear Mixed Model

Independent variables used in the nonlinear mixed model were Mid-BA, DRINV and the interaction of Mid-BA and DRINV. The combination of these two variables and their interaction term provided smaller chi-square values compared with the other variables. The variable DBHSQ used in PROC NLIN was not used in PROC NLMIXED because it resulted a high chi-square value which made the model more insignificant. The "Parameter estimates" Table 7 below indicates high significance of the two fixed-effects parameters, their interaction effect (also fixed) and one random parameter associated with the plot effect. The sign of the coefficients have to be interpreted in carefully in light of the fact that the interaction effect was significant, as in the iteratively re-weighted nonlinear regression model discussed previously. Furthermore, the variance of the random parameter (σ^2_u) has a high t-value and very a low p-value indicating the significant plot level effect.

Table 7. Parameters, their maximum likelihood estimates, standard errors and inferential statistics - nonlinear mixed procedure.

Variable name	Parameter	Estimate	Standard Error	DF	t-Value	Pr > |t|	Alpha
Intercept	β_0	11.2753	0.7044	207	16.01	<.0001	0.05
Mid-BA	β_1	-0.0792	0.005229	207	-15.15	<.0001	0.05
DRINV	β_2	-5.5501	0.7504	207	-7.4	<.0001	0.05
Mid-BA×DRINV	β_3	0.08083	0.005812	207	13.91	<.0001	0.05
Variance component	σ^2_u	1.1755	0.2233	207	5.26	<.0001	0.05

2.6.6. Model formulated using logistic regression and the nonlinear mixed procedure

$$APOS= (1/ (1+ (\exp-(11.2753 - 0.0792 \times \text{Mid-BA} -5.5501 \times \text{DRINV}+ 0.08083\times \text{Mid-BA}\times \text{DRINV} + u)))) \quad (7)$$

where;

$\hat{\beta}_0$=11.2753, $\hat{\beta}_1$= -0.0792, $\hat{\beta}_2$= -5.5501 and $\hat{\beta}_3$= 0.08083 are fixed-effects parameter estimates

σ^2_u= variance component which describes the spread of the random coefficient

u= the random parameter where $u \sim N(0, \sigma^2_u)$

Table 8 shows the results from the chi-square goodness-of-fit test. The test was applied using the parameters estimated through the nonlinear mixed model procedure. The total number of trees used for this analysis was 20,283 of which 19,312 were observed survival trees and 971 were observed mortality trees. The total chi-square calculated value is 253.209 which is far more than the chi-square tabulated value 19.68 at the alpha level of 0.05 hence rejecting the hypothesis that the model fits well.

Overall the chi-square value for survival which is 17.41 is very much smaller than the value for mortality i.e. 235.80. Because of high chi-square values for mortality in certain diameter classes such as in classes 2, 4, 10 through 16, the model did not fit well.

Table 8. Observed survival and mortality and expected survival and mortality from a logistic model by 2-inch DBH class with chi-square values for the plot remeasurement period- nonlinear mixed procedure.

Diameter Class	Total Trees	Survival			Mortality		
		Observed No.	Expected No.	Chi. Square	Observed No.	Expected No.	Chi. Square
2	1451	1109	1237.76	13.3943	342	213.241	77.7476
4	3869	3582	3680.68	2.6454	287	188.325	51.7025
6	3886	3791	3809.52	0.0901	95	76.478	4.486
8	3320	3265	3270.36	0.0088	55	49.637	0.5794
10	2801	2740	2761.15	0.1621	61	39.846	11.2304
12	2271	2211	2242.18	0.4336	60	28.821	33.7285
14	1541	1496	1522.47	0.4604	45	18.526	37.8338
16	706	686	697.85	0.2014	20	8.146	17.25
18	299	294	295.71	0.0099	5	3.288	0.892
20	110	109	108.8	0.0004	1	1.197	0.0324
22	24	24	23.72	0.0034	0	0.284	0.2839
24	4	4	3.97	0.0002	0	0.031	0.0311
26	1	1	1	0	0	0.001	0.0012
Total	**20283**	**19312**	**19655.17**	**17.41**	**971**	**627.821**	**235.7988**

2.7. Conclusions and Recommendations

Logistic regression was used to investigate the important variables in estimating the probability of survival of individual shortleaf pine trees. Various independent variables such as crown class, square root of DBH, square of DBH / mid-basal area, tree dominant height, site index, mid-plot age, interaction between Mid-BA and DBHSQ, DBHSQ and DRINV etc. were tested in addition to the variables finally selected for use in the model. Although some these independent variables were found significant in logistic regression, they were found nonessential in nonlinear regression and in estimating annual survival of individual shortleaf pine trees. Both iteratively reweighted nonlinear regression and a nonlinear mixed model were used in the parameter estimation.

With the parameters estimated through iteratively reweighted nonlinear regression, the model evaluation suggested mid-basal area per acre, inverse of ratio of quadratic mean diameter to DBH, square of DBH and the interaction between mid-basal area per acre and inverse of ratio of quadratic mean diameter to DBH as the best set of independent variables. This set of these variables provided less chi-square value compared with other sets of the variables. These independent variables were found to be very significant in predicting annual probability of survival of a tree.

Nonlinear mixed modeling was applied to investigate effects due to a random parameter at the plot level and to evaluate the model performance. Mid-basal area per acre, square of DBH and interaction between them were important variables using the nonlinear mixed procedure. The set of these independent variables provided better chi-square estimates than the other sets of the variables. However the total chi-square value was still high. The nonlinear mixed modeling suggested that there is significant effect on the model due to random parameters for the study.

Although a nonlinear mixed modeling approach was attempted to improve the performance of model, the model obtained with logistic regression and the parameter estimated from iteratively reweighted nonlinear model was found better in estimating survival probability. Both methods of parameter estimation yielded high chi-square statistics rejecting the null hypothesis that the model fits well, but the chi-square values from the nonlinear mixed model were substantially higher than those from the iteratively reweighted regression model. Elevated chi-square values for both models might be due to high mortality chi-square values in some diameter classes.

The model below which utilized the parameters estimated through iteratively reweighted nonlinear regression was chosen as the final model for this study.

$$APOS = (1/(1+(\exp(-(9.791 - 0.0741 \times \text{Mid-BA} - 4.0594 \times DRINV - 0.0025 \times DBHSQ + 0.0736 \times \text{Mid-BA} \times DRINV))))$$

The interaction effect of Mid-BA and DRINV was found significant therefore the interpretation of main effect due to Mid-BA and DRINV separately is affected by the interaction term. The total coefficient of Mid-BA (that is, the sum of the coefficient for the interaction term multiplied by a constant DRINV and the coefficient of the linear term) is positive when DRINV is greater than 1 and is negative when DRINV is less than

1. On the other hand, the total coefficient of DRINV is positive when Mid-BA is greater than 55 ft^2/ acre and is negative when Mid-BA is less than 55 ft^2/acre.

Based on the results from the final model, the observed frequencies of individual tree survival have no severe deviation from the expected frequencies however; this was not the case with the frequencies for mortality trees. Also the chi-square calculated values for survival are much less than those for mortality indicating the model fit much better for survival.

Further study is recommended using other models with different combinations of independent variables that may improve the prediction probability of survival of individual shortleaf pine trees. The final model above is considered as a better alternative than a constant survival rate and could be selected for use in the Shortleaf Pine Stand Simulator (SLPSS) (Huebschmann et al. 1998) which is a distance-independent individual tree growth simulator for naturally-occurring shortleaf pine. The model provides important information about individual tree survival for the shortleaf pine trees found in the eastern Oklahoma and western Arkansas region.

LITERATURE CITED

ALLAN, S., K. ANDRES, H. MARIS, L. DIANA, AND K.V. GADOW. 2009. Estimating tree survival: a study based on the Estonian forest research plots network. *Annales Botanici Fennici* 46:336-352.

ALLISON, P.D. 2001. *Logistic Regression using the SAS System: theory and applications*, 2nd ed. SAS® publishing, books by users. press.288p.

AVILA, O.B., AND H.E. BURKHART. 1992. Modeling survival of loblolly-pine trees in thinned and unthinned plantations. *Can. J. For. Res.* 22:1878-1882.

BELCHER, D.M., M.R. HOLDAWAY, AND G.J. BRAND. 1982. *A discription of STEMS-the stand and tree evaluation and modeling system*. USDA For. Serv. Gen. Tech. Rep. NC-49.

BUDHATHOKI, C.B., T.B. LYNCH, AND J.M. GULDIN. 2008a. Nonlinear mixed modeling of basal area growth for shortleaf pine. *For. Ecol.Manag.* 255:3440-3446.

BUDHATHOKI, C.B., T.B. LYNCH, AND J.M. GULDIN. 2008b. A mixed-effects model for the dbh-height relationship of shortleaf pine (*Pinus echinata* Mill.). *South J. Appl. For.* 32(1):5-11.

BUFORD, M.A., AND W.L. HAFLEY. 1985. Modeling the probability of individual tree mortality. *For. Sci.* 31:331-341.

BURKHART, H.E., K.D. FARRAR, R.L. AMATEIS, AND R.F. DANIELS. 1987. *Simulation of individual tree growth and stand development in loblolly pine plantations on cutover, site-prepared areas.* Sch. For. Wildl. Res., Va. Polytech. Inst. & State Univ., Publ. No. FWS-1-87. 47p

CAO, V. 2000. Prediction of annual diameter growth and survival for individual trees from periodic measurements. *For. Sci.* 46:127-131.

CLUTTER, J.L., J.C. FORTSON, L.V. PIENAAR, G.H. BRISTER, AND R.L. BAILEY. 1983. *Timber management: a quantitative approach*. John Wiley & Sons, New York. 333p.

DANIEL, B.H., R.L. BAILEY. 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *For. Sci.* 47(3):311-321.

DEEN, J.L. 1933. A survival table for even-aged stands of northern white pine. *Journal of Forestry.* 31:42-44.

DOBBERTIN, M., AND G.S. BIGING.1998. Using the non-parametric classifier CART to model forest tree mortality. *For. Sci.* 44(4):507-516.

EK, A.R. 1974. Nonlinear models for stand table projection in northern hardwood stands. *Can. J. For. Res.* 4:23-27.

GERTNER, G.Z. 1989. The need to improve models for individual tree mortality. P. 59-61 in *Proc. 7th Central Hardwood Conf.* General Technical report. USDA Forest Service, North Central Forest Experiment Station, NC-132.

GREGOIRE, T.G., AND O. SCHABNBERGER. 1996a. A non-linear mixed-efects model to predict cumulative bole volume of standing trees. *J. Appl. Stat.* 23:257-271.

GREGOIRE, T.G., AND O. SCHABENBERGER. 1996b. Nonlinear mixed-effects modeling of cumulative bole volume with spatially correlated within-tree data. *J. Agric. Biol. Environ. Stat.* 1:107-119.

GREGOIRE, T.G., O. SCHABENBERGER, AND J. BARRETT. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25: 137-156.

HAMILTON, D.A., AND B.M. EDWARDS. 1976. *Modeling the probability of individual tree mortality.* USDA For. Serv. Res. Pap. INT-185., Ogden, UT. 22p.

HAMILTON, D.A. 1974. *Event probabilities estimated by regression.* USDA For. Serv. Intermt. Res. Pap. INT-152. 18p.

HAMILTON, D.A. 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of northern Idaho. *For. Sci.* 32(4):989-1000.

HAMILTON, D.A.1990. Extending the range of applicability of an individual tree mortality model. *Can. J. For. Res.* 20:1212-1218.

HANNEMAN, R. Linear Models. 2010. Available online at http://faculty.ucr.edu/~hanneman/linear_models/c4.html#4.1%20%20Introduction . Last accessed 7/2/2010.

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. M.S. thesis. Oklahoma State Univ., Stillwater, OK. 78 p.

HOSMER, D.W., S. LEMESHOW, W. JOHN, AND I. Songs. 2000. *Applied logistic regression*, Wiley, New York. 373p.

HUEBSCHMANN, M.M., T.B. LYNCH, AND P.A. MURPHY. 1998. *Shortleaf pine stand simulator: An even-aged natural shortleaf pine growth and yield model user's manual*. Res. Rep. P-967, Oklahoma State Univ. Agric. Exp. Stn. 25 p.

KEISTER, T.D. 1972. Predicting individual tree mortality in simulated southern pine plantations. *For. Sci.* 18:213-217.

KING, S.L., K. P. BENNETT, AND S. LIST. 2000. Modeling non-catastrophic individual tree mortality using logistic regression, neural networks, and support vector methods. *Computers and Electronics in Agriculture* 27:401-406.

KRAUCH, H. 1930. Mortality in cut-over stands of western yellow pine. *Journal of Forestry* 28:1085-1097.

KRUMLAND, B., J. DYE, AND L.C. WENSEL. 1977. *Individual tree mortality models for the coast region of California*. University of California, Co-op Redwood Yield Research Project, Berkely, California. Research Note 6.7 p.

KUTNER, M.H., C.J. NACHTSHEIM, AND J. NETER. 2004. Applied linear regression models, 4th Edition. McGraw-Hill/ Irwin.701p.

LAPPI, J. 1997. A lognituding analysis of height/diameter curves. *For. Sci.* 43:555-570.

LAPPI, J. 2006. A multivariate, non parametric stem-curve prediction method. *Can. J. of For. Res.* 36:291-301.

LAPPI, J. , AND R.L. BAILEY. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. *For. Sci.* 34:907-927.

LAWSON, E.R. 1990. Shortleaf pine, Pinus echinata Mill. P. 316-326 *in* Silvics of North America: conifers. Agric. Handb. 654. Burns, R. M., B. H. Honkala. Tech. comps. Washington, DC: U.S. Department of Agriculture, Forest Service: Vol. 1.

LEE, Y.J. 1971. Predicting mortality for even-aged stands of lodgepole pine. *Forestry Chronicle* 47:29-32.

LYNCH, T.B., M.M. HUEBSCHMANN, AND P.A. MURPHY. 1998. A survival model for individual shortleaf pine trees in even-aged natural stands. P. 14-16 in *Proceedings of an international conference on integrated resource inventories.* Boise, ID.

LYNCH, T.B., L.R. GERING, M.M. HUEBSCHMANN, AND P.A. MURPHY. 1999a. A survival model for shortleaf pine trees growing in uneven-aged stands. Paper presented at the tenth biennial southern silvicultural research conference. 531-535. Shreveport, LA February 16-18.

LYNCH, T.B., K.L. HITCH, M.M. HUEBSCHMANN, AND P.A. MURPHY. 1999b. An individual-tree growth and yield prediction system for even-aged natural shortleaf pine forests. *South J. App. For.* 23(4):203-211.

LYNCH, T.B., AND P.A. MURPHY.1995. A compatible height prediction and projection system for individual trees in natural, even-aged shortleaf pine stands. *For. Sci.* 41:194-209.

MCCULLAGH, P., AND J.A. NELDER. 1983. *Generalized Linear Models.*Chapman and Hall, New York. 261p.

MCCULLAGH, P., AND J.A. NELDER. 1989. *Generalized Linear Models.* 2nd Edition. Chapman and Hall, New York. 511p.

MONSERUD, R.A. 1976. Simulation of forest tree mortality. *For. Sci.* 22:438-444.

MONSERUD, R.A., J.W. FLEWELLING. 2002. Comparing methods for modeling tree mortality. P. 168-177 In USDA Forest Service Proceedings RMRS-P25.

MONSERUD, R.A., AND H. STERBA. 1999. Modeling individual tree mortality for Austrian forest species. *For. Ecol. Manag.*113(2-3):109-123.

MURPHY, P.A.1986. Growth and yield of shortleaf pine. P. 159-177 In *Proc. of symp. on the Shortleaf pine ecosystem,* Murphy, P.A. (ed.). March 31-April2, 1986, Little Rock, Ar. Arkansas Coop. Ext. Serv., Univ. of Arkansas, Monticello, AR 272 p.

MURPHY, P.A., AND R.M. FARRAR. 1985. *Growth and yield of uneven-aged shortleaf pine stands in Interior Highlands*. USDA For. Serv. Res. Pap.SO-218. New Orleans, LA:U.S Department of Agriculture, Forest Service, Southern Forest Experiment Station. 11p.

MURPHY, P.A., AND M.G. SHELTON. 1996. An individual-tree based area growth model for Loblolly pine stands. *Can. J. For. Res.* 26:327-331.

NETER, J., AND E.S. MAYNES. 1970. On the appropriateness of the correlation coefficient with a 0, 1 dependent variable. *Journal of the American Statistical Association* 65:501-509.

ODAFS. 2000. Forest Trees of Oklahoma, Oklahoma Department of Agricultural Forestry Services, Oklahoma City. pp. 205 p.

PEET, R.K., AND N.L. CHRISTENSEN. 1987. Competition and tree death. *BioScience* 37:586-595.

PINHEIRO, J. C., AND D.M. BATES. 2000. *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, New York. 528p.

RAGAVAN, A. 2008. Data mining application of non-linear mixed modeling in water quality analysis. In *Proceedings of the data mining and predictive modeling section, SAS Global Forum,* San Antonio, TX. 14p.

ROSE, C.E., JR. 1998. Parameter estimation for shortleaf pine (*Pinus echinata* Mill.) basal area growth model. M.S. Thesis. Oklahoma State University,Stillwater, OK. 100p.

ROSE, C.E. JR., D.B. HALL, B.D. SHIVER, M.L. CLUTTER, AND B. BORDERS. 2006. A multilevel approach to individual tree survival prediction. *For. Sci.* 52 (1):31-43.

SAS Institute, Inc. 2007. SAS online help documentation. SAS Institute, Inc. Cary, NC. http://support.sas.com/documentation/onlinedoc/guide/

SMITH,K.L. 1986. Abstract to the symposium on the shortleaf pine ecosystem. P. 1-8 in *Proceedings of the symposium on the shortleaf pine* ecosystem, Monticello, AR. University of Arkansas, Arkansas Cooperaative Extension Service

SOMERS, G.L., R.C. ODERWALD, W.R. HARIS, AND O.G. LANGDON. 1980. Predicting mortality witha Weibull funciton. *For. Sci.* 26:291-300.

TECK, R.M., AND D.E. HILT. 1990. *Individual tree probability of survival model for the Northeastern United States*. USDA Forest Service. Northeastern For Exp Stn Res Pap, NE-642, 10p.

TEMESGEN, H., AND S.J. MITCHELL. 2005. An individual-tree mortality model for complex stands of southeastern British Columbia. *West. J. Appl. For.*20(2):101-109.

TRINCADO, G., AND H.E. BURKHART. 2006. A generalized approach for modeling and localizing stem profile curves. *For. Sci.* 52.

VANCLAY, J.K. 1994. M*odeling forest growth and yield. Applications to mixed tropical forest*. CAB International, Wallingford, UK. 312p.

WILLETT, R.L. 1986. Forward to the symposium on the shortleaf pine ecosystem. P. iii-iv. In *Proceedings of the symposium on the shortleaf pine ecosystem.*Monticello, AR: University of Arkansas, Arkansas Cooperaative Extension Service:. 272p.

WOLFINGER, R.D. 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. SAS User's Group Internet. Pap#287.

YANG, Y., S.J. TITUS, AND S. HUANG. 2003. Modeling individual tree mortality for white spruce in Alberta. Ecological Modeling 163:209-222.

YAO, X., S. J. TITUS, AND S.E. MACDONALD. 2001. A generalized logistic model of individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta mixedwood forests. *Can. J. For. Res.* 31(2):283-291.

ZHANG, S., R.L. AMATEIS, AND H.E. BURKHART. 1997. Constraining Individual Tree Diameter Increment and Survival Models for Loblolly Pine Plantations. For. Sci. 43:414-423.

VITA

SUDIP SHRESTHA

Candidate for the Degree of

Master of Science

Thesis: ESTIMATING THE PROBABILITY OF SURVIVAL OF INDIVIDUAL SHORTLEAF PINE *(Pinus Echinata* Mill.*)* TREES

Major Field: Natural Resource Ecology and Management- Forest Resources

Minor Field: Statistics

Education:

Completed the requirements for the Master of Science in Forest Resources specialization in Forest Biometrics at Oklahoma State University, Stillwater, Oklahoma in July, 2010.

Received Bachelor of Science in Forestry from Institute of Forestry at Tribhuvan University, Pokhara, Kaski, Nepal in 2006.

Experience:

Worked as a consultant forest technician in Nepal-Swiss Community Forestry Project, Nepal, in 2001

Professional Memberships:

Xi Sigma Pi National Honor Society
American Statistical Association (ASA)
Institute of Mathematical Statistics (IMS)
Biometrics-Eastern North American Region (ENAR)

Name: Sudip Shrestha Date of Degree: July, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: ESTIMATING THE PROBABILITY OF SURVIVAL OF INDIVIDUAL SHORTLEAF PINE *(Pinus Echinata* Mill.*)* TREES

Pages in Study: 59 Candidate for the Degree of Master of Science

Major Field: Forestry

Minor Field: Statistics

Scope and Method of Study:
Logistic regression was used to obtain the best set of independent variables for prediction of individual shortleaf pine tree survival in even-aged natural stands. Logistic survival model parameters were estimated using iteratively reweighted nonlinear regression. A nonlinear mixed model was also applied to investigate inclusion of a random plot level effect in the survival model. The final logistic individual tree survival model can be used to predict the annual survival rate of individual trees of even-aged shortleaf pine forests located in Ozark and Ouachita National Forests and in the surrounding regions. The logistic model could be selected for use in the Shortleaf Pine Stand Simulator, which is a distance independent forest growth simulator for naturally-occurring shortleaf pine trees.

Findings and Conclusions:
Mid-basal area per acre, inverse of ratio of quadratic mean diameter to diameter at breast height, their interaction and square of diameter at breast height were found to be significant variables in predicting the survival of individual shortleaf pine trees. Iteratively reweighted nonlinear regression was used to estimate logistic model parameters. Also a nonlinear mixed model was developed and a plot-level random effect was found significant. However the model from logistic regression with parameters estimated through iteratively reweighted nonlinear regression was considered the best final model by comparing the test statistics from the chi-square goodness-of-fit test. The goodness-of-fit test rejected the hypothesis that the model fits well. This might be because of high mortality chi-square values in some diameter classes. The contribution to chi-square from survival is much lower than for mortality indicating that he model fits better for survival than for mortality.

ADVISER'S APPROVAL: Dr. Thomas B. Lynch