DYNAMICS OF UNDERTORY SHORTLEAF PINE (*Pinus echinata* Mill.) AND HARDWOOD AFTER THINNING SHORTLEAF PINE FORESTS IN THE SOUTHEASTERN UNITED STATES

By

ANUP KC

Bachelor of Science in Forestry
Institute of Forestry, TU
Pokhara, Nepal
2007

Master of Science in Aquaculture and Fisheries
University of Arkansas at Pine Bluff
Pine Bluff, Arkansas
2011

Submitted to the Faculty of the

Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2016

DYNAMICS OF UNDERTORY SHORTLEAF PINE (*Pinus echinata* Mill.) AND HARDWOOD AFTER THINNING SHORTLEAF PINE FORESTS IN THE SOUTHEASTERN UNITED STATES

Dissertation Approved:

Thomas B. Lynch

Dissertation Advisor

Lan Zhu

Rod Will

Stephen Hallgren

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my major advisor, Dr. Thomas B. Lynch, for his guidance and support during the last four years of my research. I am thankful to Dr. Rodney Will, Dr. Lan Zhu, and Dr. Stephen Hallgren for serving on my dissertation committee. I am also thankful to the Department of Natural Resource Ecology and Management for funding my research and education for four years.

I am especially thankful to my family. I would like to thank my father and mother for everything that they have done for me in this life. Thank you to my lovely wife, Ambika Shrestha, for her hard work, patience, and the encouragement throughout this journey. I thank my daughter Ansu for being the most adorable little girl I have ever encountered. I also want to express my sincere thanks to my uncles and aunts for their unwavering love and support.

Finally, I would like to say 'THANK YOU' to all the other known and unknown mentors and friends who helped me to become a better person in life.

Name: ANUP KC

Date of Degree: JULY, 2016

Title of Study: DYNAMICS OF UNDERTORY SHORTLEAF PINE (*Pinus echinata* Mill.) AND HARDWOOD AFTER THINNING SHORTLEAF PINE FORESTS IN THE SOUTHEASTERN UNITED STATES

Major Field: NATURAL RESOURCE ECOLOGY AND MANAGEMENT

Abstract: The shortleaf pine (*Pinus echinata* Mill.) population is consistently declining in southeastern United States. Shortleaf pine forests are thinned frequently to improve the growth and development of residual stands. But, the effect of thinning on growth and development of understory woody-plants in long term has not been extensively studied. We assessed the effects of thinning, overstory shortleaf pine characteristics, climatic, and topographic factors on shortleaf pine regeneration applying various predictive modeling techniques. We applied decision tree models to predict shortleaf pine regeneration. We also developed, evaluated, and compared the performance of three other predictive models to predict shortleaf pine regeneration. We used understory shortleaf pine data that were collected from shortleaf pine forests of Arkansas and Oklahoma spanning a period of 25 years following thinning and hardwood control treatments. The shortleaf pine densities have declined in every subsequent measurement since the first measurement of understory trees in 1996. Thinning treatments played a significant role on the understory shortleaf pine density. The decision tree model using the Gini criteria as the splitting rule to predict the shortleaf pine regeneration had a low misclassification rate of 7.6 percent. The decision tree model can be an efficient tool to make shortleaf pine stand management decisions. The best performing logistic regression model showed precipitation, plot age, site index, and overstory thinning were the significant inputs affecting shortleaf pine regeneration with validation misclassification rate of 8 percent. The best performing artificial neural network model predicted the shortleaf pine regeneration with validation misclassification rate of 7.6 percent, and cumulative lift of 5, 2.5 and 1.66 at depth of 20, 40 and 60 respectively. An artificial neural network model performed best to predict the shortleaf pine regeneration. Poor shortleaf pine regeneration performance over decades in study sites suggests the future of shortleaf pine dominated forests is questionable unless further regular silvicultural treatments are applied. We recommend continual hardwood removal every 10-15 years to obtain the satisfactory understory shortleaf pine regeneration in shortleaf pine forests of Arkansas and Oklahoma.

TABLE OF CONTENTS

Chapter Page

I. Introduction 1

References...8

II. Long-term response of understory shortleaf pine (*Pinus echinata* Mill.) and hardwoods to thinning in natural shortleaf pine-oak stands in Arkansas and Oklahoma 13

Abstract 13
Introduction..15
Materials and Methods...18
Results..22
Discussion 26
Conclusions..31
References..33

III. Predicting shortleaf pine regeneration (*Pinus echinata* Mill.) after thinning the overstory in Ozark and Ouachita mountain Forests: A Decision Tree Model Approach 52

Abstract 52
Introduction..54
Materials and Methods...56
Data Analysis 64
Results..65
Discussion 68
Conclusions..72
References..73

Chapter Page

IV. Predicting shortleaf pine regeneration after thinning in southeastern United States: A comparison of logistic regression, artificial neural network, and support vector machine methods 93

Abstract 93
Introduction 94
Materials and Methods 96
Results 101
Discussion 102
Conclusions 105
References 106

V. Conclusion 124

LIST OF TABLES

Table Page

Table II-1 Overstory basal area (m^2 ha^{-1}) levels of permanent plots after thinning in 1985-1986 in Ozark-St. Francis National Forest and Ouachita National Forest 39

Table II-2 Characteristics of regenerated woody-vegetation at various measurement periods 40

Table II-3 Density of understory woody-plants with their standard error (SE) per hectare at four measurement periods in Ozark-St. Francis and Ouachita National Forests 41

Table II-4 Percent of plots with high (≥1730 stems ha^{-1}) shortleaf pine density m^2 ha^{-1} at each thinning level in Ozark-St. Francis National Forest (OZNF) and Ouachita national forests (OUNF) 44

Table III-1 General description of inputs showing assigned variable names, their scale and maximum and minimum values 82

Table III-2 Input variables of interval and ratio scale with imputed values and other descriptive summary statistics 83

Table III-3 Fit statistics, model significance and significant variables for three logistic regression (LR) models 84

Table III-4 Summary statistics of the best performing logistic regression model (LR2) 85

Table III-5 Fit statistics of the three Decision Tree (DT) models..................................86

Table III-6 Prediction results from the application of the default Decision Tree (DT1) model on validation data 87

Table III-7 Decision rules derived from the default Decision Tree (DT) model (DT1) 88

Table III-8 Percentage distribution of response variable (SLPN) in four thinning levels 89

Table IV-1 Description of explanatory variables (inputs) and their assigned variable names, scale, maximum, and minimum values..121

Table IV-2 Descriptive summary of interval inputs 122

Table IV-3 Logistic regression models and their fit statistics 123

Table IV-4 Parameter estimates and odds ratio of inputs from LR2 model 124

Table IV-5 Performance of ANN models on predicting shortleaf pine regeneration 125

Table IV-6 Performance of SVM models on predicting shortleaf pine regeneration 126

Table IV-7 Event classification for selected LR, ANN, and SVM models for training and validation data...127

LIST OF FIGURES

Figure …Page

Fig. II-1 Mean residual overstory basal area (m^2 ha^{-1}) of shortleaf pine over various measurement periods at four thinning treatment levels. First thinning was conducted in 1985 to create plots with four distinct thinning levels. Plots were thinned second time after 1996 measurement period to maintain the overstory basal area level similar at 1985. A represents Ozark St-Francis National Forest (OZNF) and B represents Ouachita National Forest (OUNF) 45

Fig. II-2 Regeneration density of shortleaf pine at four thinning treatment levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Four thinning treatment levels are created based on the overstory shortleaf pine basal area (m^2 ha^{-1}) after thinning 46

Fig. II-3 Regeneration density of oaks at four thinning levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B) 47

Fig. II-4 Regeneration density of red maple at four thinning levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B) 48

Fig. II-5 Density of shortleaf pine regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents shortleaf pine density (ha^{-1}) at particular dbh for that measurement year 49

Fig. II-6 Density of oak regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents oak density (ha^{-1}) at particular dbh for that measurement year 50

Fig. II-7 Density of red maple regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents red maple density (ha^{-1}) at particular dbh for that measurement year 51

Fig. III-1 Schematic illustration of a hypothetical Decision Tree. Two hypothetical inputs (Thinning and Precipitation) are used to split the Decision Tree. Any node that is not further splitting is leaf node 90

Fig. III-2 Schematic presentations of the predictive model building process....................91

Fig. III-3 Flow diagram for model development and comparison on SAS Enterprise Miner 12.3. SLP is the name of a dataset that we used to build all the models. Arrows make the connections between nodes, and the data analysis process moves a step forward. Graph Explore and Stat Explore were used to conduct the descriptive analysis of the data before building the Decision Tree and Logistic Regression models 92

Fig. III-4 Default Decision Tree (DT1) model to predict the shortleaf pine regeneration. Class A and B are number of plots with low (<1730 stems ha^{-1}) and high (≥1730 stems ha^{-1}) regeneration respectively. Results from training data (Train %) on DT1 has been validated using validation data (Valid%)...93

Fig. III-5 An example of an interactive DT model where the model split three times using three input variables (Site index, Overstory basal area and Precipitation). Because it is an interactive DT model, the values of input variables were selected by the authors………94

Fig. III-6 ROC (Receiver Operating Characteristics) curves comparing the performances of default Decision Tree (DT1) model with the best performing Logistic Regression (LR2) model. Sensitivity and the specificity are described in methods section 95

Fig. III-7 Percent of plots with high shortleaf pine regeneration (B>1730 stems ha^{-1}) over four measurement periods at four thinning levels (A<10.332 m^2 ha^{-1}, 10.332 m^2 ha^{-1} ≤B< 17.22 m^2 ha^{-1}, 17.22 m^2 ha^{-1} ≤C< 24.108 m^2 ha^{-1}, and D≥ 24.108 m^2 ha^{-1}). These four thinning levels are created using the shortleaf pine residual basal area. Results are extracted using the interactive Decision Tree (DT2) model 96

Fig. IV-1 Architecture of the MLP neural network for SLPN estimates.........................128

Fig. IV-2 Process flow map of the model building processes, in SAS EM software version 12.3 interface, and their comparison among the group of similar models and the selected models. “SLP Data” is a dataset that is cleaned and ready to be analyzed. Then, data has been imputed for the missing values and transformed as required. “Data Partition” node randomly separated data into two sets as training and validation data sets. “Drop” node dropped all the variables from the dataset that are not used by ANN and SVM models. “Model Comparison” node compared the performances of the models...129

Fig. IV-3 Receiver operating characteristics (ROC) curves to compare the performances of selected LR (LR2), ANN (ANN3), and SVM (SVM2) models for training and validation data...130

Fig. IV-4 Cumulative lift charts for ANN3 and SVM2 models with baseline 131

CHAPTER I

INTRODUCTION

Historically, shortleaf pine (*Pinus echinata* Mill.) has been one of the most common forest types in the southeastern United States (McWilliams et al., 1986; Kabrick et al., 2010), and it is second only to loblolly pine (*Pinus taeda L.*) among southern pines in standing volume (Budhathoki, et al., 2008). It grows in 22 states over more than 1,139,600 km^2, ranging from southeastern New York to eastern Texas (Willet, 1986), covering the broadest range among the southern pines (Williston and Balmer, 1980). Indeed, shortleaf pine is one of the most important tree species in Arkansas and eastern Oklahoma (Zhang et al., 2012). The Ouachita Mountains cover 6.6 million acres of area, and nearly 40 percent of total forested areas in Ouachita Mountains are shortleaf pine dominated forests (Guldin et al., 2004). Despite its wide distribution in the region, shortleaf pine is the least understood species among the four major pines (Guldin, 2007). Shortleaf pine grows well in areas having the mean annual temperature from 9 ºC to 21 ºC, with minimums of -30 ºC and maximums of 39 ºC (Williston and Balmer, 1980).

Shortleaf pine has been desirable in the region in terms of timber production for southern pine lumber which is typically used in building and home construction. It is also a source of southern pine pulpwood for the pulp and paper industry. Shortleaf pine is relatively more resistant to surface fire, and capable of re-sprouting than hardwoods or loblolly pine (*Pinus taeda* L.) after the fire incidents. This resistance to fire makes shortleaf pine desirable for restoration efforts that feature controlled burning such as restoration to the shortleaf pine-bluestem grass ecosystem. Shortleaf pine stands are particularly desirable for red-cockaded woodpecker habitats (Zhang et al., 2012) from the wildlife management perspective. The esthetic values of shortleaf pine are also important for tourism and recreation (Lawson and Kitchens, 1983). In spite of these beneficial aspects of shortleaf pine forests, shortleaf pine populations have been declining in recent years (Moser et al., 2006; KC et al., 2015). KC et al. (2015) suggested that the current rate of regeneration of shortleaf pine seedlings is not adequate to maintain the shortleaf pine dominated forests in the long-term in Ozark-St. Francis and Ouachita Mountains of Arkansas and Oklahoma.

Most previous studies of naturally-regenerated shortleaf pine forests have focused on the growth and development of overstory shortleaf pine stands after thinning. Studies conducted by Budhathoki et al. (2006, 2008 a, 2008 b) are some examples. In some circumstances it is possible that the total cost for thinning could be higher than the value of the resulting benefits. In such cases, the entire thinning process becomes economically unrealistic. However, in many cases, thinning can be profitable (Larson and Mirth, 2004). Many wildlife and game species prefer shortleaf pine-bluestem habitat over shortleaf pine-hardwood habitats. Many studies in recent past focused on assessing the effect of shortleaf pine-bluestem restoration for red-cockaded woodpeckers (see, Masters et al., 1998; Conner et al., 2002; Thill et al., 2004). Thinning to appropriate levels of shortleaf overstory and control of understory hardwoods using fire are essential features to restore the shortleaf-pine bluestem ecosystem. These studies suggest there has been a renewed interest on restoring the shortleaf pine-bluestem grass habitat (Kabrick et al. 2011).

Thinning is a common practice in shortleaf pine forest to maximize productivity of residual stocks. Pre-commercial thinning of natural stands is sometimes beneficial and its importance is well documented in previous studies. Thinning increases residual individual tree volume and reduces competition with other hardwoods (Jhang et al., 2012). Though thinning can be used to maximize the amount of volume that a stand produces, it is not economically feasible to do so in all cases. Overstory growth and development is not always the sole purpose of thinning. Thinning promotes understory shortleaf pine and hardwood regeneration and offers better habitat for wildlife such as red-cockaded woodpecker, bobwhite quail (*Colinus virginianus*), Bachman's sparrow (*Aimophila aestivalis*), and eastern wild turkey (*Meleagris gallopavo*) (Bukenhofer and Hedrick 1997; Guldin et al., 2004). Thinning also increases the amount and palatability of wildlife food plants in the thinned stands (Lawson and Kitchens, 1983).

Lawson and Kitchens (1983) reported shortleaf pine stands can be managed using single tree selection silvicultural systems. The selection system is especially attractive for the managers of small tracts, and the selection harvesting system also supports the regeneration. Guldin et al. (2004) recommended that reducing the overstory basal area to 18.36-17.21 m^2 ha^{-1} (75-80 ft^2 $acre^{-1}$) creates better habitat for wildlife. Lynch et al (2003) and Nkouka (1999) studied effects of multiple overstory factors on shortleaf pine regeneration and reported that higher levels of overstory basal area affect the shortleaf pine regeneration negatively. Lawson (1986), Nkouka et al. (1999), and Lynch et al. (2003) assessed the effect of overstory shortleaf pine and reported that higher site indices affect the shortleaf pine regeneration negatively.

Shortleaf pine is a shade-intolerant (Baker et al., 1996) species, and hardwoods are the climax vegetation in many areas of the southeastern United States. Baker (1992) indicated that young shortleaf pine seedlings tolerate shade relatively well; however, it becomes more intolerant as the stand gets older. When a shade intolerant or moderately tolerant species like shortleaf pine fails in response to the intense competition and rapid height growth of competing trees to remain in top canopy, they lag behind and succumb to hardwood competition (Baker et al., 1996). When a dense hardwood understory is expected to hinder natural pine regeneration, eliminating hardwoods in combination with pine thinning is an excellent management practice (Rogers and Brinkman, 1965; Stevenson et al., 2010). Controlling hardwoods along with thinning increases the productivity of residual shortleaf pine (Lowery, 1986). This practice also increases shortleaf pine seed production (Phares and Rogers, 1962). Competition control measures should be implemented when the competition for water and light becomes critical to newly established seedlings. This allows extra space and resources for adequate natural shortleaf pine regeneration. Single-stem injection, foliar spray, or soil application of herbicides are effective measures to eliminate hardwoods; especially when the hardwood tree sizes are small (Loyd et al., 1978). Mechanical methods, such as hand cutting and shearing also temporarily reduce hardwood competition, but may cause problems with sprouting. Maple (1965) found that brush cutting provided higher survival percentages of shortleaf pine seedlings (2.9) and stocking levels than chemical treatment (1.3) and burning (0.4). Crow and Shilling (1980) reported beginning a burning program several years before the harvest/regeneration cut reduces hardwood competition for newly established seedlings. Rapid regrowth of most hardwoods is possible after conducting the mechanical control method (Lowery, 1986).

Lilly et al. (2012) reported that the shortleaf pine population is declining in southeastern United States. Recent studies (Nkouka 1999; Lynch et al., 2003, KC et al., 2015; KC et al., 2016) show the shortleaf pine regeneration in Arkansas and Oklahoma is not satisfactory on the study sites that they examined. Those findings reveal both understory and overstory shortleaf pine populations are not as prolific as might be desired. Because many studies conducted in the past focused on overstory shortleaf pine, there is a gap of knowledge concerning the understory of shortleaf pine stands. In addition to the response of the residual overstory post thinning, it is equally important to assess how the understory of shortleaf pine stands responds to overstory or understory treatments. Certainly, there is not just one single factor responsible for the decline of shortleaf pine in the region. But, selective removal of shortleaf pine, intense hardwood competition, short fire intervals (<3 years) before and after logging, no surface fire treatments, and global climate change are major factors making the situation more adverse for shortleaf pine abundance and regeneration. This study assesses how the understory shortleaf pine stands response the overstory thinning treatments in long-term. Here, we answer questions that are related to the understory regeneration, growth and development.

The general objective of this study is to study the development of understory shortleaf pine density and associated hardwood understory development in Arkansas and southeastern Oklahoma. We try to shed some light on factors that affect the understory shortleaf pine survival and development and on what measures can be applied to promote the understory shortleaf pine regeneration in the region. To achieve these goals, we conducted three studies with specific objectives, which are described below.

Firstly, we assess understory shortleaf pine and hardwoods densities and their development in long-term. We assessed how the understory shortleaf pine and hardwoods interact to each other after conducting the thinning treatments to the overstory shortleaf pine at four thinning levels. Specifically, we (1) quantified the species richness, dominance, and diversity of the woody plants. We (2) also assessed densities and relative frequencies of understory woody plants and shortleaf pine for approximately 25 years. We (3) assessed the effect of four thinning treatment levels on understory shortleaf pine density. And, we (4) assessed whether the shortleaf pine, oaks, and red maple densities at various dbh levels differ in long-term.

Secondly, in light of the fact that shortleaf pine regeneration is low in the region (KC et al., 2016), this study evaluates the effects of overstory stand level variables (site index, plot age, overstory basal area hectare^{-1}) and other climatic (precipitation) and topographic (slope, aspect, altitude) factors on shortleaf pine regeneration in Ozark and Ouachita National Forests in Arkansas and Oklahoma. We predict the chance of shortleaf pine regeneration at satisfactory levels using several decision tree (DT) models representing various limitation or growth conditions. We also compared the predictive performance of the selected DT models to logistic regression (LR) models. Specifically, we apply the decision tree model to assess the shortleaf pine regeneration response to overstory thinning in the long term. Additionally, we illustrate an interactive DT where the forest managers can interactively change the inputs to achieve the desired number of shortleaf pine regeneration stems in their forests. Furthermore, we examine the association between shortleaf pine regeneration and thinning level over the period of 25 years. We expect this study to be helpful to manage the shortleaf pine forests not only in Arkansas and Oklahoma but also in the entire southeastern United States. Most importantly, this study will establish a precedent that the predictive models are helpful in forest management related research which supports stakeholder decision making.

Thirdly, we compared multiple forms of logistic regression (LR), artificial neural network (ANN) and support vector machine (SVM) models to predict the shortleaf pine regeneration in Ozark and Ouachita national forests in Arkansas and Oklahoma. To the best of our knowledge, this study is first in kind to use and compare predictive modeling techniques to assess shortleaf pine regeneration in the southeastern United States. It is important to develop the efficient statistical/ predictive models that assess the major factors influencing shortleaf pine regeneration. Our data are the widest ranging study of shortleaf pine response to thinning with the longest monitoring period for understory tree development of which we are aware. We expect this study will help to better understand the present and future status of shortleaf pine forests in Arkansas and Oklahoma and to develop efficient management programs in shortleaf pine forests.

References

Baker, J.B., Cain, M.D., Guldin, J.M., Murphy, P.A., Shelton, M.G., 1996. Uneven-aged silviculture for the loblolly and shortleaf pine forest cover types. USDA Forest Service, Southern Research Station. General Technical Report SO-118. P. 1.

Budhathoki, C.B., 2006. Mixed-effects modeling of shortleaf pine (*Pinus echinata* Mill.) growth data, Oklahoma State University.

Budhathoki, C.B., Lynch, T.B., Guldin, J.M., 2008. A Mixed-Effects Model for the dbh–Height Relationship of Shortleaf Pine (*Pinus echinata* Mill.). Southern J. Appl. Forestry 32(1): 5-11.

Bukenhofer, G.A., Hedrick, L.D., 1997. Shortleaf pine/bluestem grass ecosystem renewal in the Ouachita Mountains. Trans. North Am. Wildl. Nat. Resour. Conf. 62: 509-515.

Conner, R.N., Shackelford, C.E., Schaefer, R.R., Saenz, D., Rudolph, C.D., 2002. Avian community response to southern pine ecosystem restoration for red-cockaded woodpeckers. The Wilson Bulletin, 114(3), pp.324-332.

Crow, A.B., Shilling, C.L., 1980. Use of prescribed burning to enhance southern pine timber production. Southern J. Appl. Forestry. 4(1): 15-18.

Guldin, J.M., 2007. Restoration and management of shortleaf pine in pure and mixed stands - science, empirical observation, and the wishful application of generalities. In: Kabrick, J.M., Dey, D.C., Gwaze, D. (Eds.), Shortleaf pine restoration and ecology in the Ozarks: proceedings of a symposium. General Technical Report NRS-P-15. USDA Forest Service, Northern Research Station, Newton Square, PA, pp. 47–58.

Guldin, J.M., Strom, J., Montague, W., Hedrick, L.D., 2004. Shortleaf pine-bluestem habitat restoration in the interior highlands: implications for stand growth and regeneration. In: Silviculture in special places: Proceedings of the National Silviculture Workshop, Shepperd, WD., Eskew, LG., (ed.). USDA, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 182-190.

Kabrick, J.M., Dey, D.C., Shifley, S.R., Villwock, J.L., 2010. Early survival and growth of planted shortleaf pine seedlings as a function of initial size and overstory stocking. P. 277-286 *in* 17th central hardwood forest conference. U.S. Department of Agriculture, Forest Service, Northern Research Station, Lexington, KY.

Kabrick, J.M., Dey, D.C., Shifley, S.R., Villwock, J.L., 2011. Early survival and growth of planted shortleaf pine seedlings as a function of initial size and overstory stocking. In: Proceedings of the 17th central hardwood forest conference. USDA, Forest Service, Northern Research Station, Lexington, KY, pp. 277-286.

KC, A., Lynch, T.B., Stevenson, D., Wilson, D., Guldin, J.M., Heinermann, B., Holeman, R., Anderson, K., 2015. Development of understory tree vegetation after thinning naturally occurring shortleaf pine forests. In: Holley, A. Gordon; Connor, Kristina F.; Haywood, James D., (ed.). Proceedings of the 17th biennial southern silvicultural research conference. E-Gen. Tech. Rep. SRS-203. Ahseville, NC: USDA, Forest Service, Southern Research Station: 99.

Larson, D.S., Mirth, R.A., 2004. Case study on the economics of thinning in the wildland urban interface. Western J. Appl. Forestry. 19(1): 60-5.

Lawson, E.R., 1986. National regeneration of shortleaf pine. In: Murphy, P.A. (Ed.), Proceedings of a Symposium on the Shortleaf Pine Ecosystem. Univ. Arkansas Coop. Ext. Serv., Monicello, AR, 53-63.

Lawson, E.R., Kitchens, R.N., 1983. Shortleaf pine. Silvicultural systems for the major forest types of the United States. USDA Forest Service, Agriculture Handbook 445: 157.

Lilly, C.J., Will, R.E., Tauer, C.G., Guldin, J.M., Spetich, M.A., 2012. Factors affecting the sprouting of shortleaf pine rootstock following prescribed fire. For. Ecol. Manage. 265, 13-19.

Lowery, R.F., 1986. Woody Competition Control. In: Murphy, P.A. (Ed.), Symposium on the Shortleaf Pine Ecosystem. Cooperative Extension Servie, Monticello, AR., Little Rock, AR, pp. 147-158.

Loyd, R.A., Thayer, A.G., Lowry, G.L., 1978. Pine growth and regeneration following three hardwood control treatments. South. J. Appl. For. 2(1): 25-27.

Lynch, T.B., Nkouka, J., Huebschmann, M.M., Guldin, J.M., 2003. Maximum likelihood estimation for predicting the probability of obtaining variable shortleaf pine regeneration densities. For. Sci. 49(4), 577-584.

Maple, W.R., 1965. Shortleaf pine stands five years after seedfall on prepared sites. USDA, For. Serv., South. For. Exp. Stn., Res. Note 20-2 7.

Masters, R.E., Lochmiller, R.L., McMurry, S.T., Bukenhofer, G.A., 1998. Small mammal response to pine-grassland restoration for red-cockaded woodpeckers. Wildlife Society Bulletin, pp.148-158.

McWilliams, W.H., Sheffield, R.M., Hansen, M.H., Birch, T.W., 1986. The shortleaf resource. In: Proceedings of the symposium on the shortleaf pine ecosystem; 1986 March 31 – April 2, Little Rock, AR. Arkansas Cooperative Extension Service: 9-24.

Moser, W.K., Hansen, M., McWilliams, W.H., Sheffield, R.M., 2006. Shortleaf pine composition and structure in the United States. P. 19-27 *in* Shortleaf pine restoration and ecology in the Ozarks. U.S. Department of Agriculture, Forest Service, Northern Research Station, Springfield, MO.

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.S. thesis, Oklahoma State University, Stillwater, OK.

Phares, R.E., Rogers, N.F., 1962. Improving shortleaf pine seed production in Missouri. J. For. 60, 3.

Roberts, D.W., 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66, 123-31.

Rogers, N.F., Brinkman, K.A., 1965. Shortleaf pine in Missouri; understory hardwoods retard growth. Central States Forest Experiment Station. U.S. Forest Service Res. Pap. CS-15, pp. 9.

Stevenson, D.J., Lynch, T.B., Guldin, J.M., 2010. Hardwood regeneration related to overstory shortleaf pine (*Pinus echinata* Mill.) basal area, site index, and time since cutting in Arkansas and eastern Oklahoma. In: 14th biennial southern silvicultural research conference, Stanturf, J.A. (ed.). USDA, Forest Service, Southern Research Station, Athens, GA. pp. 587-588.

Thill, R.E., Rudolph, D.C., Koerth, N.E., 2004. Shortleaf pine-bluestem restoration for Red-cockaded Woodpeckers in the Ouachita Mountains: Implications for other taxa.

Willet, R.L., 1986. Foreword. In: Murphy, P.A. (Ed.), Proc. Symp. on the Shortleaf Pine Ecosystem. Arkansas Coop. Ext. Ser., University of Arkansas, AR pp. iii–iv.

Williston, H.L., William, E.B., 1980. Shortleaf pine management. Southeastern Area, State and Private Forestry, (6).

Zhang, D., Huebschmann, M.M., Lynch, T.B., Guldin, J.M., 2012. Growth projection and valuation of restoration of the shortleaf pine–bluestem grass ecosystem. Forest Policy Econ 20: 10-15. doi:10.1016/j.forpol.2012.01.014

CHAPTER II

Long-term response of understory shortleaf pine (*Pinus echinata* Mill.) and hardwoods to thinning in natural shortleaf pine-oak stands in Arkansas and Oklahoma

Abstract

Shortleaf pine (*Pinus echinata* Mill.) is considered one of the most important tree species in Arkansas and eastern Oklahoma, and it has been used for southern pine lumber. Recent studies suggest that shortleaf pine population is consistently declining in the region, but there have been relatively few studies conducted in past to assess the long-term understory response of shortleaf pine-oak stands in southeastern United States. This study assessed the effects of thinning on understory woody-plant density and regeneration dynamics. It also assessed the trend of density change at five dbh classes for understory shortleaf pine, oaks and red maple. The understory regeneration data were collected since 1996. The study plots were located in Arkansas and Oklahoma. Ecological metrics including species richness, evenness, relative densities, and relative frequencies were calculated. We found understory shortleaf pine densities are declining in every subsequent measurement since the first measurement in 1996. The thinning treatment played a significant role on the understory shortleaf pine density ($P<0.001$). Plots with the heaviest thinning treatment (overstory residual basal area <10 m^2 ha^{-1}) had significantly high

numbers of understory shortleaf pine densities in all measurement years ($P<0.001$). Thinning did not show a significant effect on understory oak densities ($P>0.05$). The results revealed overstory thinning can have a positive impact on understory shortleaf pine growth and development. But, heavy thinning (overstory residual basal area <10 m^2 ha^{-1} in present case) is required to establish the desired level of understory shortleaf pine densities. We concluded that one time thinning is not sufficient to maintain the desired level of understory shortleaf pine densities in the absence of hardwood control measures. The continual control of hardwood every 10-15 years interval is recommended to ascertain the satisfactory understory shortleaf pine densities in shortleaf pine forests of Arkansas and Oklahoma.

1. Introduction

Shortleaf pine-oak forests are a common forest type in southeastern United States (McWilliams et al., 1986; Kabrick et al., 2011), and thinning is a common practice in those stands to promote the growth and development of residual shortleaf pine (*Pinus echinata* Mill.) trees (Wittwer et al., 1996; Sabatia et al., 2009, 2010). Thinning is not primarily focused on understory regeneration; however, it helps to promote the understory regeneration and development (Shelton and Cain, 2000; Elliot and Vose, 2005) by increasing light, nutrient and water availability. Overstory thinning also frequently enhances habitat for wildlife such as red-cockaded woodpecker, bobwhite quail (*Colinus virginianus*), Bachman's sparrow (*Aimophila aestivalis*), and eastern wild turkey (*Meleagris gallopavo*) (Bukenhofer and Hedrick, 1997; Guldin et al., 2004). Thinning increases the amount and palatability of wildlife food plants (Lawson and Kitchens, 1983). Thinning improves the esthetical value of shortleaf pine for the visitors and recreationists (Lawson and Kitchens, 1983). Therefore, many forest land owners in Arkansas and Oklahoma prefer to grow shortleaf pine or shortleaf-hardwood mixed stands in their private lands.

Shortleaf pine is considered one of the most important tree species in Arkansas and eastern Oklahoma (Zhang et al., 2012). It can be used for southern pine lumber. These products are often used in housing industry. In southeast Oklahoma, paper mills use shortleaf pine pulpwood for paper production. Shortleaf pine is more valuable in this region than low-quality hardwood timber. Shortleaf pine is important ecologically as a dominant component of the "pre-settlement" forest in the Ouachita region (Guldin et al., 2004). Shortleaf pine is also particularly desirable for red cockaded woodpecker habitat (Zhang et al., 2012). Poor regeneration performance of shortleaf pine in the region (KC et al., 2015) is one of the major factors for a sharp decline in the shortleaf pine stocking. Consistent shortleaf pine regeneration is essential to establish a sound shortleaf pine-oak forest; Lynch et al (2003) and Guldin et al. (2004) have described in detail what level of shortleaf pine regeneration is considered sufficient for stand establishment.

Although the high woody plant regeneration following thinning is a common phenomenon (Nagai and Yoshida, 2006; Royo and Carson, 2006), not all woody plants, including shortleaf pine, that regenerate following thinning survive in the long-term (KC et al., 2015). At present, we lack information, specifically for shortleaf pine-oak forest, on what percentage of early regeneration gets established. It is important to assess whether an excellent response of understory shortleaf pine following thinning (KC et al., 2015) helps to establish a long-term shortleaf pine dominated forest. The regenerated woody-vegetation community that is initiated by thinning is often composed of species groups that compete with each other for resources. This competition affects the establishment of understory shortleaf pine in later years (Kuehne and Puettmann, 2008). Shade tolerant understory species already present in the canopy hinder the development of shade intolerant species like shortleaf pine even after the thinning (Alaback and Herman, 1988). The long-term effects of thinning on forests, in particular the shortleaf pine-oak forest in this case, need to be assessed with large-scale data driven studies to fully understand the shifting dynamic of woody understory plants (Vallauri et al., 2002; Larsen 2006; Ares et al., 2010). Ecological metrics such as species richness, evenness, relative density, relative frequency, and diversity have been utilized in past (Sagar and Singh, 2006) to assess how vegetation dynamics changes over time. These metrics should provide the useful information for understory shortleaf pine-oak forests dynamics on long term data.

Recent studies suggest that shortleaf pine population is consistently declining in the region in recent years (Moser et al., 2006; KC et al., 2015). These considerations underline the importance of study on status of understory shortleaf pine in the region. However, relatively few studies have been conducted in past to assess the long-term understory response of shortleaf pine-oak stands in southeastern United States. Therefore, we assessed the long-term response of understory shortleaf pine and hardwood to overstory thinning. The findings will help the forest managers to manage and improve their shortleaf pine stands by achieving satisfactory understory shortleaf pine densities in future. In this study, (1) we quantified the species richness, dominance, and diversity of the woody plants. We expect this objective to show how the understory woody-plant combinations and dominance change in the long term after thinning treatment. We assessed (2) densities and relative frequencies of understory woody plants and shortleaf pine for approximately 25 years. Thirdly (3), we assessed the effect of four thinning treatment levels on understory shortleaf pine density. The null hypothesis was that the thinning treatment levels have no significant effect on understory shortleaf pine density. This will also provide insights into how much should we thin and at what time interval to achieve certain understory shortleaf pine densities. We assessed (4) whether the shortleaf pine, oak, and red maple densities at various dbh levels differ in long-term. We expect to find the major understory species density distribution at multiple dbh levels and to examine their transitions among dbh classes. And, at last, we assessed (5) how the overstory thinning, overstory shortleaf pine characteristics, climatic, and topographic factors affect the shortleaf pine density and at what level. Here, we expect to investigate which factors that have a major effect on understory shortleaf pine density and at what level.

2. Material and Methods

2.1. Study area

The USDA Forest Service Southern Research Station and the Department of Forestry, Oklahoma State University (now part of the Department of Natural Resource Ecology and Management) jointly established 180 permanent study plots in Ozark-St. Francis National Forest (OZNF) (then named the Ozark National Forest) and Ouachita National Forests (OUNF) during the period from 1985 to 1987. Study plot locations ranged from OZNF near Russellville, Arkansas (latitude 35.3º N, longitude 93.1º W) to areas on the OUNF near Broken Bow (latitude 34.0º N, longitude 94.7º W) in southeastern Oklahoma (Lynch et al., 2003). Out of 180 plots, 133 plots were from OUNF and 47 plots were from OZNF. Plots were circular, 809.37 m^2 in area and 16.06 m radius. A 10.05 m isolation buffer was created outside each plot. Study plots and the buffer area were thinned from below at the time of plot establishment in 1985-86 to create four distinct overstory basal area levels: <10, 10–17, 17–24 and ≥24 m^2 ha^{-1} (Table II-1). Most overstory shortleaf pine plots were thinned for second time after third overstory measurement in 1996. The purpose of second thinning was to return overstory shortleaf pine basal area to levels similar to those after the first thinning in 1985. The overstory basal area details of the four thinning treatment levels at over four measurement periods are shown in figure (II-1). Hardwoods greater than 2.54 cm in diameter at ground level were removed from study plots and buffer areas using herbicides at the time of plot establishment but there was no hardwood control at the second thinning during the 1996 measurement.

2.2. Overstory, climatic and topographic data

Overstory shortleaf pine trees were measured for diameter at breast height (dbh) immediately after establishment of the plots in 1985-86 and thereafter, at 4-5 year intervals. All shortleaf pine trees and saplings present in 1985 after creating the thinning levels were considered the overstory. Subsamples of shortleaf pine tree heights and crown lengths were also obtained on each plot. Dominant shortleaf pine trees were measured for height and age to determine site index base age 50 years. Mean annual precipitation (mm), elevation (m), slope and aspect were determined using the GPS locations of the study plots. Aspect values were transformed into northness and eastness using trigonometric functions as described by Roberts (1986).

2.3. Understory data

In 1996, two 20.23 m^2 subplots were created inside the 809.37 m^2 plot to measure understory trees and shrubs. Both plots were on a line crossing the plot center and equidistant between plot center and plot boundary. Two more subplots of same size were created in 2001. These were on a line perpendicular to the line joining the other two understory plots and also equidistant between plot center and plot radius. Understory trees and shrubs were counted, and the dbhs were measured four times: 1996, 2001, 2006 and 2013. Measured dbhs were divided into five dbh classes: <1.27 cm, 1.27–3.81cm, 3.81–6.35 cm, 6.35–8.89 cm, and 8.89–25.4 cm. All shortleaf pine and hardwoods that regenerated after the chemical treatment of 1985 were considered the understory. The maximum dbh of understory shortleaf pine measured in 1996, 2001, 2006, and 2013 were 7.62, 12.7, 12.7, and 17.78 cm respectively. At the same time, the maximum dbh for oaks were 7.62, 10.16, 12.7, and 17.78 cm.

2.4. Data analysis

2.4.1. Ecological metrics

We calculated species richness, the Shannon-Wiener index, dominance and evenness of understory woody plants. The understory woody-plants were measured and counted only in two subplots in first understory measurement (1996) and in four subplots thereafter. Species richness is the number of species per unit area. Shannon-Weiner index is used to characterize species diversity in a community (McArthur 1965). We used a two sample t-test to assess the mean difference in species richness over measurement periods between two sites. Species diversity was calculated applying the Shannon-Weiner diversity index ($H^{'}$) as described by Steen et al (2010):

$$H^{'} = \sum p_i \ln p_i \tag{1}$$

where p_i = proportion of individuals found in species *i*. The maximum possible diversity ($H_{\max}$) was calculated as described by Boyce (2005):

$$H_{\max} = \ln(S) \tag{2}$$

where S = Species richness. The Shannon-Weiner index was used to calculate the evenness of species distributions (*E)* in two sites according to Pielou (1966):

$$E = \frac{H^{'}}{H_{\max}} \tag{3}$$

Evenness assumes values between 0, implying completely heterogeneous, and 1, implying completely homogenous. Relative density provided the numerical strength of a species in relation to the total number of individuals of all the species:

$$\text{Relative Density} = \frac{\text{Number of individuals of the species}}{\text{Number of individuals of all the species}} \times 100\% \qquad (4)$$

At the same time, relative frequency provided the degree of dispersion of individual species in an area in relation to the number of all the species occurred:

$$\text{Relative Frequency} = \frac{\text{Number of occurrences of the species}}{\text{Number of occurrences of all the species}} \times 100\% \qquad (5)$$

2.4.2. Two-way and repeated measured ANOVAs

The goodness of fit test (Kolmogorov-Smirnov test) was applied to check the normality of all the covariates. If the continuous predictor variables were not normal, we applied the proper transformations to reduce the skewness and the kurtosis of the variables. The correlations between the covariates were calculated to assess whether there were high correlations between the covariates. Two-way ANOVAs were used to assess if the overstory thinning levels and the sites (OZNF vs OUNF) had significant effects on mean understory shortleaf pine and oak basal area at each measurement levels. Further, Tukey's post hoc tests were applied whenever needed. The repeated measured ANOVA was used to assess if the mean shortleaf pine counts were different over all measurement periods. *P*-values of Wilks' lambda tests were used to determine the level of significance. Similar tests were conducted to assess whether the understory oak densities were significantly different among measurement periods. Observations with at least one missing value were dropped by SAS PROC GLM as a standard procedure of analyzing repeated measured ANOVA. The *P*-value of 0.05 was considered as the cutoff point for significance level throughout all tests.

2.4.3. Logistic Regression model

We applied binomial logistic regression model to predict the probability of satisfactory understory shortleaf pine density. Previous studies (Lynch et al., 2003) suggested that the shortleaf pine density of 1730 stem ha^{-1} was satisfactory density to establish a shortleaf pine dominated forest. Therefore, we categorized the understory shortleaf pine densities (stems ha^{-1}) into two classes; less than 1730 stems ha^{-1} and 1730 or more stems ha^{-1}. Splitting the densities exactly at 1730 stem ha^{-1} kept present study consistent with earlier research (see, Lynch et al., 2003). The stepwise selection method was used to select the best fitted logistic regression model. The Hosmer and Lemeshow goodness of fit test (see, Hosmer et al., 2013) was used to check the adequacy of the fitted model for the data set used. Again, the *P*-value of 0.05 was considered as the cutoff point for significance level.

3. Results

3.1. Understory species richness, dominance and diversity

A total of 68 understory tree and shrubs species were recorded in all study sites and times. The lowest numbers of understory woody plant species, 33 in OZNF and 37 in OUNF, were recorded in 1996. The highest number of understory woody plant species, 43 in OZNF and 57 in OUNF, were recorded in 2006. No significant difference was found between sites in terms of species richness ($P<0.05$). The Shannon-Wiener diversity index (H) consistently increased over measurement periods for OUNF; however, there was not a similar increase in the index on the OZNF (Table II-2). The evenness index decreased over measurement periods for both sites (Table II-2). Individually, shortleaf pine was the most dominant understory species, in terms of densities, in 1996. However, the dominance of shortleaf pine decreased during the following measurement periods. Red maple (*Acer rubrum*) and oaks, mostly southern red oak (*Quercus falcata*), white oak (*Quercus alba*) and northern red oak (*Quercus rubra*), were consistently dominant in both sites in following years (Table II-3). Red maple and oaks were the most dominant species in OZNF. However, oaks only were the most dominant species group in OUNF.   Shortleaf pine declined from 1996 to 2013 and the oaks and maple increased.

3.2. Relative frequency, stem counts and densities

Hickory species (*Carya spp.*) had the highest relative frequency (13.18±2.86) percent in OZNF in 1996. Red maple dominated the understory woody-vegetation (15.46±3.79 in 2001, 17.97±2.92 in 2006 and 22.68±3.79 in 2013) in terms of relative frequency in rest of the measurement periods. Similarly, blackjack oak (*Quercus marilandica*) was the most dominant species, in terms of relative frequency, (12.12±1.95 in 1996, 9.28±1.34 in 2006, 10.40±1.68 and 10.72±1.60 in 2013) in OUNF region throughout the measurement periods. Relative frequency of red maple has increased consistently over the past 25 years in both sites. The relative frequency of shortleaf pine has declined consistently in both sites since the first measurement, and it is closely followed by the flowering dogwood (*Cornus florida*). Shortleaf pine had the highest average density of any particular species in OZNF with an average of 1205±471 stem ha^{-1} in 1996. After that, red maple had the highest average density in 2001, 2006 and 2013 respectively (Table II-3). Similarly, in the case of OUNF, shortleaf pine had the highest average density of any particular species in 1996 and 2001 respectively. Red maple and blackjack oak had the highest average densities in 2006 and 2013. White oak, southern red oak, winged elm (*Ulmus alata*), sparkleberry (*Vaccinium arboreum*) etc. were some other major species in terms of stem density. Further details for all species densities are displayed in table (II-3).

The mean numbers of understory shortleaf pine stems have changed significantly over four measurement time periods ($P<0.001$). The changes in numbers of shortleaf pine stems at different measurement time periods were site (OZNF and OUNF) dependent ($P<0.05$). The measurement time periods and the thinning levels together affected the overall shortleaf pine densities ($P<0.001$). Similarly, the mean understory oak stems were significantly different over four measurement periods ($P<0.001$), but not between two sites ($P>0.05$).

3.3. Response of understory shortleaf pine, oaks, and red maple at four thinning treatment levels

Treatment levels resulting from thinning from below had a significant effect on understory shortleaf pine density ($P<0.001$). Thinning treatment plots with basal area less than 10 m^2 ha^{-1} had significantly higher number of understory shortleaf pine density ($P<0.001$) compared to other three treatment levels in all measurement years (Fig. II-2). Thinning treatment plots with basal area between 10 m^2 ha^{-1} and 17 m^2 ha^{-1} had the significantly higher number of understory shortleaf pine density ($P<0.001$) compared to thinning treatment levels that had basal area greater than 27 m^2 ha^{-1} . However, thinning treatment levels with basal area of 10 m^2 ha^{-1} to 17 m^2 ha^{-1} had no significant difference with levels 17 m^2 ha^{-1} to 24 m^2 ha^{-1}, and level 17 m^2 ha^{-1} to 24 m^2 ha^{-1} had no significantly different density with greater than 27 m^2 ha^{-1} ($P>0.05$). The total shortleaf pine density in thinning treatment level with basal area less than 10 m^2 ha^{-1} has declined in subsequent measurement periods (Fig. II-2). Understory shortleaf pine density was significantly higher in 1996 ($P>0.05$) compared to 2006 and 2013. But, the density difference was not significant in 1996 and 2001.

The results were little different for understory oak density. Firstly, the oak density was significantly different between two sites ($P=0.002$). None of the thinning treatment levels were significantly different from each other (P>0.05) except for the thinning treatment level of less than 10 m^2 ha^{-1}" with greater than 24 m^2 ha^{-1}. Figure II-3 shows the understory densities of oaks over four measurement periods for both sites at four thinning treatment levels. Analyzing the oak densities at different measurement years, the understory oak density at 2006 was significantly higher than the densities from any other measurement years. There was no significant difference in oak densities at other measurement years; for example, 1996 vs 2001, 1996 vs 2013, and 2001 vs 2013. The understory densities of the red maple at four thinning levels are shown in Figure II-5.

3.4. Understory shortleaf pine, oaks, and red maple density distribution at five dbh levels

The understory shortleaf pine densities declined with increasing dbh classes in 1996 in both sites. The trend was similar in 2001 measurement. However, the shortleaf pine densities were higher in dbh class "1.27-3.81 cm" than in dbh class "<1.27 cm" for measurement years 2006 and 2013 for both sites (Fig. II-5). Newly regenerated shortleaf pine stem densities were highest in dbh class "<1.27" in 2001 for both sites (Fig. II-5). The densities in dbh class "<1.27" for measurement years 2006 and 2013 are significantly lower (P<0.05) than in 1996 and 2001 (Fig. II-4). On the other hand, understory oaks are distributed well in all dbh classes. Densities are low in higher dbh classes, but oaks are regenerating well even in recent measurement years (Fig. II-5). Red maple densities are lower in higher dbh classes except for dbh class "1.27-3.81" in 2013 (Fig. II-6).

3.5. Effect of overstory characteristics, climatic, and topographic factors on understory shortleaf pine

Plots thinned to low basal area levels at establishment period consistently showed better understory shortleaf pine density over all four measurements. Basal area levels less than 10 m^2 ha^{-1} have the highest percentage of plots with high densities (>1730 stems ha^{-1}) for both sites. Table (II-4) lists detailed information about plots with high and low regeneration percentages over four measurement periods. A logistic regression model showed that the average annual precipitation, overstory residual shortleaf pine basal area and site index were the significant independent variables with negative effects on understory shortleaf pine density ($P<0.001$). Plot age (Average ages of the residual shortleaf pine trees) was the only significant variable that affected understory shortleaf pine density positively ($P<0.05$). The odds ratios for average annual precipitation, plot age, overstory basal area and site index were 0.988, 1.018, 0.896, and 0.703 respectively.

4. Discussion

After thinning shortleaf pine, studies are often focused on overstory residual growth and development and ignore the understory vegetation dynamics. This is because the primary objective of overstory shortleaf pine thinning is to improve the growth and development of residual trees. But, consistent understory shortleaf pine density at satisfactory level is key to developing and maintaining a long-term, sustainable, naturally regenerated shortleaf pine forest. In the southeastern USA, the land area in Shortleaf pine forests has been declining in recent years (Moser et al., 2006; Lilly et al., 2012; KC et al., 2015; KC et al., 2016). Industrial and private non-industrial land owners increasingly prefer fast growing loblolly pine over shortleaf pine. But, some landowners prefer shortleaf pine because it maintains natural forest aesthetics and offer less expensive establishment cost (Shortleaf pine: Land Manager's Guide, 2014). Shortleaf pine forests also have reduced the risk from climate change and are associated with native trees and habitats (Shortleaf pine: Land Manager's Guide, 2014). Despite several benefits of shortleaf pine, lack of understory shortleaf response to overstory thinning treatment can damage the future sustainability of naturally regenerating shortleaf pine stands.

Species richness, diversity, evenness, species densities, and relative frequencies provide detailed insight regarding current understory density and regeneration trends using longitudinal data. Statistically insignificant changes in species richness over four measurements in 25 years reveal no boom or bust pattern of woody-plants. The numbers of woody-plants have not changed substantially over time in either of the sites. However, a consistent increment in Shannon-Weiner diversity index in OUNF shows species composition is changing over time (Table II-2). Single species dominance or the concentration (density) has decreased over time, and the woody-plant species distributions are more balanced today than 25 years ago. Obviously, shortleaf pine is the species of a major interest that has sharply declined in recent years (Fig. II-2), and other woody-plants became more dominant in recent years. The Shannon-Weiner diversity index does not show increasing or decreasing trends of woody-plant abundance concentration in OZNF (Table II-2) suggesting that species composition is not changing as much as in OUNF. Single species dominance or the concentration (density) has not decreased at the same rate as in OUNF. The woody-plant species distributions are not consistently getting more balanced over time. Therefore, understory shortleaf pine population has not plummeted in OZNF as sharply as in OUNF. Figure (II-2) shows understory shortleaf pine density is better in OZNF. This study reveals although study sites in the two national forests were treated similarly, the understory growth and development dynamics after thinning treatment are different in OZNF and OUNF.

Shortleaf pine regeneration remained consistently relatively good throughout all measurement periods on plots that were thinned to lowest overstory basal area levels (<10 m^2 ha^{-1}) (Fig. II-2). Significantly better density of shortleaf pine in the highest thinned plots (<10 m^2 ha^{-1}) and lowest density in the lowest thinned plots (≥24 m^2 ha^{-1}) suggests that, indeed, thinning has a strong effect on shortleaf pine regeneration. Maintaining the overstory shortleaf pine basal area below certain level, below 17 m^2 ha^{-1} in this case, helps shortleaf pine forest to regenerate well and maintain a healthy understory shortleaf pine population for the long term. This study suggests thinning is an important option to consider for maintaining the future sustainability of natural shortleaf pine forests in the region by obtaining relatively better regeneration. In case of oaks, they regenerated well and maintained a healthy density irrespective of thinning levels (Fig. II-3). These results suggest that the oaks will regenerate well in these areas even if the silvicultural treatments are not applied in the shortleaf pine overstory. In fact, high understory oak density in all thinning treatment levels is not a surprising result. Even the lowest density of oaks is a lot higher than the shortleaf pine density on both sites. Ice storms that occurred in winter of 2000 (Stevenson et al., 2016) severely damaged some of the study plots in OUNF (Stevenson et al., 2016). Ice storms caused some of shortleaf pine overstory mortality, and that opened up the overstory a little on the affected plots. Bragg et al. (2003) suggested the severe winter storms such as ice and snow are some of the most important causes of forest disturbance. The highest number of woody plant counts in 2006 measurement period is the reflection of the overstory damage caused by ice storm in 2000. We presume both understory hardwoods and shortleaf pine were damaged to some extent by the ice storms in 2000. Hardwoods, especially oaks, recovered well in later years. However, shortleaf pine density did not increase in later years too. Shortleaf pine density has decreased in 2006 at all thinning treatment levels indicating that ice storms definitely hit hard at least to shortleaf pine seedlings and saplings. Shortleaf pine seedlings and saplings which regenerated after thinning had difficulty surviving in plots with thinning treatments that had high residual basal area after thinning. Moser et al. (2006) stated that they observed the shortleaf pine regeneration in many states, except for Arkansas and Oklahoma, in smaller quantities, and suggested that the shortleaf pine regeneration in southeastern U.S. is declining in recent years. This study covered the sites from Arkansas and Oklahoma, and showed the similar result. Indeed, the shortleaf pine regeneration is critically low in Arkansas and Oklahoma. Here, the reported understory shortleaf pine densities were only from the thinned plots. We presume the status can be even worse where the thinning has not been performed for a long time.

Assessing understory shortleaf pine and oaks densities at multiple dbh classes (<1.27 cm, 1.27-3.81 cm, 3.81-6.35 cm, 6.35-8.89 cm, and ≥8.89 cm) for four measurements reveals some interesting regeneration and then establishment patterns. Firstly, relatively high shortleaf pine densities at dbh level "<1.27 cm" and "1.27-3.81 cm" in 1996 for both sites shows that shortleaf pine responded well to overstory thinning in early years (Fig. II-5). The shortleaf pine densities in higher dbh classes (3.81-6.35 cm, 6.35-8.89 cm, and ≥8.89 cm) in 1996 are very low. This is because all the understory woody plants were eliminated using herbicide from the understory in 1985. Disturbance of the litter layer due to logging at the time of thinning probably enhanced conditions for shortleaf regeneration. Exposure of bare mineral soil is favorable to shortleaf pine regeneration (Clabo and Clatterbuck, 2005). The newly regenerated shortleaf pine cohorts after thinning in 1985 were not large enough to move into higher dbh levels in 1996. The data measured in 2001 followed trends similar to those in 1996. The density in dbh class "<1.27 cm" is less in 2001 revealing that the shortleaf pine did not regenerate well and it might also have died. The shortleaf pine density at dbh level “<1.27 cm” in 2006 and 2013 are critically low (Fig. II-5) in both sites. High understory shortleaf pine density in dbh class “1.27-3.81 cm” and “3.81-6.35 cm” in measurement year 2006 and 2013 shows that some of the newly regenerated shortleaf pine are being established in the forest, and they are shifting to the bigger dbh classes. The only problem is shortleaf pine regeneration is critically low in recent measurements, and at the same time, densities in higher dbh classes are also far below the satisfactory level. These results suggest that shortleaf pine is not regenerating well on these study sites. The majority of newly regenerated shortleaf pine saplings have died after few years instead of transitioning to the higher dbh classes. On the other hand, oaks are regenerating well in both sites (Fig II-6). They are also transitioning from smaller dbh classes to the higher dbh classes in a good numbers overall. Successful transitioning of understory oaks to higher dbh levels, and mortality of understory shortleaf pine instead of movement to higher dbh level indicates that forest will not have the sufficient shortleaf pine trees in future. The shortleaf dominant stands of today on these sites will shift to the oak dominant forests in future. These trends indicate that additional control of hardwood understory competition though controlled burning or herbicides would be needed on these sites to enhance survival and growth of the shortleaf pine understory.

Increased hardwood regeneration hinders the growth and development of understory shortleaf pine (Phares and Rogers, 1962). The results of the present study where we used long-term data to examine these trends suggests a similar conclusion. Here, oaks and red maple are the major species hindering the growth and development of shortleaf pine. Shortleaf pine is shade intolerant (Lilly et al., 2012) and cannot compete with hardwoods in terms of regeneration (Lowery, 1986). This study finds that thinning from below does promote the shortleaf pine regeneration. However, the intense competition of shortleaf pine with hardwoods for resource utilization inhibits the development and establishment of shortleaf pine regeneration (Rogers and Brinkman, 1965; Stevenson et al., 2010). Relatively high shortleaf pine regeneration densities in heavily thinned plots indicate similar results (Table II-4). Interestingly, shortleaf pine densities are more favorable in plots with poor site indices. This occurs because the hardwoods regenerate aggressively in sites with better site indices, and shortleaf pine regeneration and development is hindered by this competition. Lawson (1986) and Lynch et al (2003) described similar results concerning the effect of site index on shortleaf regeneration. Overstory basal area and plot age are other major factors that can be used predict the success of shortleaf pine regeneration (Lynch et. al. 2003). This study is in agreement with those results. Additionally, average annual precipitation also plays an important role in understory shortleaf pine density; but negatively. However, the effect is not very strong (odds ratio=0.988 based on the logistic regression analysis). An odds ratio value of less than 1 suggests the negative effect of the independent factor and a value greater than 1 suggests a positive effect, in this case of obtaining adequate shortleaf pine regeneration. This negative effect of precipitation in understory shortleaf pine density may occur because understory hardwoods regenerate and develop more aggressively when there is ample precipitation, and understory shortleaf pine is further stressed by this competition. This study show that shortleaf pine only regenerates adequately on sites where understory hardwoods don't proliferate. This study clearly demonstrates the importance of controlling hardwoods to obtain the satisfactory shortleaf pine regeneration. Maintaining the overstory basal area $10m^2ha^{-1}$ or less by using a heavy thinning from below is also a key factor for the success of understory shortleaf pine establishment. Even maintaining overstory shortleaf pine basal area below $17m^2ha^{-1}$ provides positive results for maintaining understory shortleaf pine population at certain level. Otherwise, significant number of understory shortleaf pine saplings will die, even before attaining a dbh of 2-3 cm, due to intense competition posed by the understory hardwoods.

5. Conclusion

The decline of shortleaf pine regeneration raises a serious concern for the future of shortleaf pine-oak mixed forests in Arkansas and Oklahoma unless forest managers and landowners actively control the hardwood understory in shortleaf pine stands. Single tree selection thinning of shortleaf pine and oaks in shortleaf pine-hardwood mixed stands at an interval of around 10-15 years would strongly improve the understory shortleaf pine density. But, absence of silvicultural treatments to control understory hardwoods in shortleaf pine stands for 25 years or long nullifies the benefits that we would receive from the first thinning in terms of regeneration. This study suggests the continual intervention is mandatory to achieve healthy shortleaf pine regeneration naturally. Here, understory regeneration dynamics change significantly in later years. Therefore, short term understory count data may provide misleading results as we report good shortleaf pine regeneration in 1996. However, the status of understory shortleaf pine in 2013 is in a critically poor condition. Therefore, we recommend further silvicultural interventions to stimulate and strengthen the understory shortleaf pine regeneration. Treatments including thinning from below, controlled burning, selective understory hardwood clearance could be the possibilitieswhere these are economical and feasible. We conclude thinning from below at sufficient levels every 10-15 years to keep the overstory basal area below17 m^2ha^{-1} would provide sufficient understory shortleaf pine in the long-term if combined with measures to control understory hardwoods.

References

Alaback, P.B., Herman, F., 1988. Long-term response of understory vegetation to stand density in Picea-Tsuga forests. Can. J. Forest. Res. 18, 1522-1530.

Ares, A., Neill, A.R., Puettmann, K.J., 2010. Understory abundance, species diversity and functional attribute response to thinning in coniferous stands. For. Ecol. Manage. 260, 1104-1113.

Boyce, R.L., 2005. Life under your feet: Measuring soil invertebrate diversity. Teaching Issues and Experiments in Ecology. Ecological Society of America 3.

Bragg, D.C., Shelton, M.G., Zeide, B., 2003. Impacts and management implications of ice storms on forests in the southern United States. For. Ecol. Manage. 186, 99-123.

Bukenhofer, G.A., Hedrick, L.D., 1997. Shortleaf pine/bluestem grass ecosystem renewal in the Ouachita Mountains. Trans. North Am. Wildl. Nat. Resour. Conf. 62: 509-515.

Clabo, D., Clatterbuck, W.K., 2005. Identification, ecology and silviculture of shortleaf pine. The University of Tennessee extension, Institute of Agriculture. P. 9.

Elliott, K.J., Vose, J.M., 2005. Effects of understory prescribed burning on shortleaf pine (*Pinus echinata* Mill.)/mixed-hardwood forests. J. Torrey Bot. Soc. 132, 236-251.

Guldin, J.M., 2004. Ouachita and Ozark Mountains symposium: ecosystem management research. Gen. Tech. Rep. SRS–74. USDA, Forest Service, Southern Research Station, Asheville, NC, pp. 321.

Guldin, J.M., Strom, J., Montague, W., Hedrick, L.D., 2004. Shortleaf pine-bluestem habitat restoration in the interior highlands: implications for stand growth and regeneration. In: Silviculture in special places: Proceedings of the National Silviculture Workshop, Shepperd, WD., Eskew, LG., (ed.). USDA, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 182-190.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. John Wiley & Sons.

Kabrick, J.M., Dey, D.C., Shifley, S.R., Villwock, J.L., 2011. Early survival and growth of planted shortleaf pine seedlings as a function of initial size and overstory stocking. In: Proceedings of the 17th central hardwood forest conference. USDA, Forest Service, Northern Research Station, Lexington, KY, pp. 277-286.

KC, A., Lynch, T.B., Guldin, G.M., 2016. Shortleaf pine (*Pinus echinata* Mill.) and hardwood regeneration after thinning natural shortleaf pine forests in southern United States. In: Schweitzer, C.J., Clatterbuck, W.K., Oswalt, C.M., eds. Proceedings of the 18th biennial southern silvicultural research conference. e–Gen. Tech. Rep. SRS–212. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 614

KC, A., Lynch, T.B., Stevenson, D., Wilson, D., Guldin, J.M., Heinermann, B., Holeman, R., Anderson, K., 2015. Development of understory tree vegetation after thinning naturally occurring shortleaf pine forests. In: Holley, A. Gordon; Connor, Kristina F.; Haywood, James D., (ed.). Proceedings of the 17th biennial southern silvicultural research conference. E-Gen. Tech. Rep. SRS-203. Ahseville, NC: USDA, Forest Service, Southern Research Station: 99.

Kuehne, C., Puettmann, K.J., 2008. Natural regeneration in thinned Douglas-fir stands in western Oregon. Journal of Sustainable Forestry 27, 246-274.

Larsen, D.R., 2006. Forest stand dynamics of shortleaf pine in the Ozarks. In: Kabrick, J.M., Dey, D.C., Gwaze, D. (Eds.), Shortleaf pine restoration and ecology in the Ozarks. USDA, Forest Service, Northern Research Station, Springfield, MO, pp. 33-34.

Lawson, E.R., 1986. National regeneration of shortleaf pine. In: Murphy, P.A. (Ed.), Proceedings of a Symposium on the Shortleaf Pine Ecosystem. Univ. Arkansas Coop. Ext. Serv., Monicello, AR, pp. 53-63.

Lawson, E.R., Kitchens, R.N., 1983. Shortleaf pine. Silvicultural systems for the major forest types of the United States. USDA Forest Service, Agriculture Handbook 445: 157.

Lilly, C.J., Will, R.E., Tauer, C.G., Guldin, J.M., Spetich, M.A., 2012. Factors affecting the sprouting of shortleaf pine rootstock following prescribed fire. For. Ecol. Manage. 265, 13-19.

Lowery, R.F., 1986. Woody Competition Control. In: Murphy, P.A. (Ed.), Symposium on the Shortleaf Pine Ecosystem. Cooperative Extension Servie, Monticello, AR., Little Rock, AR, pp. 147-158.

Lynch, T.B., Nkouka, J., Huebschmann, M.M., Guldin, J.M., 2003. Maximum likelihood estimation for predicting the probability of obtaining variable shortleaf pine regeneration densities. For. Sci. 49(4), 577-584.

MacArthur, R.H., 1965. Patterns of species diversity. Biol. Rev 40(4), 510-533.

McWilliams, W.H., Sheffield, R.M., Hansen, M.H., Birch, T.W., 1986. The shortleaf resource. In: Proceedings of the symposium on the shortleaf pine ecosystem; 1986 March 31 – April 2, Little Rock, AR. Arkansas Cooperative Extension Service: 9-24.

Moser, W.K., Hansen, M., McWilliams, W.H., Sheffield, R.M., 2006. Shortleaf pine composition and structure in the United States. In: Kabrick, J.M., Dey, D.C., Gwaze, D. (Eds.), Shortleaf pine restoration and ecology in the Ozarks. USDA, Forest Service, Northern Research Station, Springfield, MO, pp. 19-27.

Nagai, M., Yoshida, T., 2006. Variation in understory structure and plant species diversity influenced by silvicultural treatments among 21-to 26-year-old Picea glehnii plantations. J. For. Res. 11, 1-10.

Phares, R.E., Rogers, N.F., 1962. Improving shortleaf pine seed production in Missouri. J. For. 60, 3.

Pielou, E.C., 1966. Species-diversity and pattern-diversity in the study of ecological succession. Theor. Biol. 10(2): 370-83.

Roberts, D.W., 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66, 123-31.

Rogers, N.F., Brinkman, K.A., 1965. Shortleaf pine in Missouri; understory hardwoods retard growth. Central States Forest Experiment Station. U.S. Forest Service Res. Pap. CS-15, pp. 9.

Royo, A.A., Carson, W.P., 2006. On the formation of dense understory layers in forests worldwide: consequences and implications for forest dynamics, biodiversity, and succession. Can. J. For. Res. 36, 1345-1362.

Sabatia, C.O., Will, R.E., Lynch, T.B., 2009. Effect of Thinning on Aboveground Biomass Accumulation and Distribution in Naturally Regenerated Shortleaf Pine. South. J. Appl. For. 33, 5.

Sabatia, C.O., Will, R.E., Lynch, T.B., 2010. Effect of thinning on partitioning of aboveground biomass in naturally regenerated shortleaf pine (*Pinus echinata* Mill.). In: Stanturf, J.A. (Ed.), 14th biennial southern silvicultural research conference. USDA, Forest Service, Southern Research Station, pp. 577-578.

Sagar, R., Singh, J.S., 2006. Tree density, basal area and species diversity in a disturbed dry tropical forest of northern India: implications for conservation. Environ. Conserv. 33(03), 256-262.

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

Shortleaf pine: natural regeneration in Oklahoma. 2014. Oklahoma Forestry Service, Department of Agriculture, Food, and Forestry, Oklahoma City, Oklahoma.

Steen, A., Ziervogel, K., Arnosti, C., 2010. Comparison of multivariate microbial datasets with the Shannon index: An example using enzyme activity from diverse marine environments. Org. Geochem. 41(9), 1019-1021.

Stevenson, D.J., Lynch, T.B., Guldin, J.M., 2010. Hardwood regeneration related to overstory shortleaf pine (*Pinus echinata* Mill.) basal area, site index, and time since cutting in Arkansas and eastern Oklahoma. In: 14th biennial southern silvicultural research conference, Stanturf, J.A.

(ed.). USDA, Forest Service, Southern Research Station, Athens, GA. pp. 587-588.

Stevenson, D.J., Lynch, T.B., Saud, P., Heineman, R., Holeman, R., Wilson, D., Anderson, K., Cerny, C., Guldin, J.M., 2016. Winter in the Ouchitas- A severe winter storm signature in Pinus echinata in the Ouachita mountains of Oklahoma and Arkansas, USA. Dendrochronologia.

Vallauri, D.R., Aronson, J., Barbero, M., 2002. An analysis of forest restoration 120 years after reforestation on badlands in the Southwestern Alps. Restor. Ecol. 10(1), 16-26.

Wittwer, R.F., Lynch, T.B., Huebschmann, M.M., 1996. Thinning Improves Growth of Crop Trees in Natural Shortleaf Pine Stands. South. J. Appl. For. 20, 6.

Zhang, D., Huebschmann, M.M., Lynch, T.B., Guldin, J.M., 2012. Growth projection and valuation of restoration of the shortleaf pine–bluestem grass ecosystem. For. Policy Econ. 20, 10-15.

Table II-1 Overstory basal area ($m^2\ ha^{-1}$) levels of permanent plots after thinning in 1985-1986 in Ozark-St. Francis National Forest and Ouachita National Forests

Ozark-St. Francis National Forest (OZNF)				
Basal Area Class	Total Plot	Mean BA± SD	Skewness	AOTC ha^{-1}
Less than 10 (A)	10	7.08±0.27	0.21	145
Between 10 and 17 (B)	8	14.49±1.01	2.43	312
Between 17 and 24 (C)	10	20.90±0.81	1.14	981
Greater than 24 (D)	12	27.27±0.73	-0.58	1080
Ouachita National Forest (OUNF)				
Less than 10 (A)	36	7.15±0.50	0.88	201
Between 10 and 17 (B)	39	14.16±0.63	1.19	391
Between 17 and 24 (C)	39	21.08±1.50	0.09	446
Greater than 24 (D)	29	27.40±0.87	0.13	971

Note: Mean BA=Mean overstory basal area ($m^2\ ha^{-1}$) after conducting the single tree selection thinning in 1985, AOTC ha^{-1}=Average overstory shortleaf pine tree counts ha^{-1} after thinning in 1985-1986

Table II-2 Characteristics of regenerated woody-vegetation at various measurement periods.

Site	OUNF				OZNF			
Year	1996	2001	2006	2013	1996	2001	2006	2013
Species richness	37	41	57	46	33	33	43	34
Shannon-Wiener diversity index (H)	3.01	3.10	3.26	3.80	2.81	2.75	2.92	2.74
Evenness (E)	0.84	0.84	0.80	0.80	0.80	0.79	0.78	0.78

Table II-3 Density of understory woody-plants with their standard error (SE) per hectare at four measurement periods in Ozark-St. Francis and Ouachita National Forests

	Ozark-St. Francis National Forest				Ouachita National Forest			
Year ------>	1996	2001	2006	2013	1996	2001	2006	2013
Scientific name (Species)								
Acer rubrum (Red maple)	1169±410*	1318±319*	1616±296*	1425±270*	565±85*	424±62*	665±96*	514±67*
Amelanchier arborea (Serviceberry)	58±52	27±22	52±45	-	58±21	303±76	39±12	19±8
Carpinus caroliniana (American hornbeam)	-	2±2	5±5	28±21	56±25	62±24	56±28	48±25
Carya spp (Hickory species)	808±141*	724±114*	388±62	714±142*	409±52	423±44*	373±64	305±50
Celtis laevigata (Hackberry)	5±5	-	5±3	3±3	50±24	4±2	12±11	7±5
Celtis laevigata (Sugarberry)	-	-	-	18±12	-	2±1	2±1	-
Cercis canadensis (Eastern redbud)	-	2±2	10±10	10±6	45±19	20±8	17±7	9±3
Cornus florida (Flowering dogwood)	813±245*	371±73	371±80	224±45	1156±193*	491±71*	481±64*	291±42
Crataegus spp. (Hawthorn)	137±48	116±34	84±22	45±15	-	36±16	14±6	7±3
Diospyros virginiana (Common persimmon)	268±103	155±39	111±50	48±39	41±13	17±5	8±3	1±1
Fraxinus americana (White ash)	200±83	82±45	86±22	53±17	186±45	49±16	92±19	101±31
Fraxinus pennsylvanica (Green ash)	-	15±6	67±40	71±50	-	81±17	74±22	17±6
Ilex opaca (American holly)	-	15±8	27±27	3±3	30±15	44±25	81±56	54±33
Juniperus virginiana (Eastern redcedar)	16±9	56±22	59±23	53±20	95±29	95±21	149±34	131±32
Liquidambar styraciflua (Sweetgum)	131±51	162±55	143±49	126±49	256±61	259±60	261±64	264±73
Morus alba (White mulberry)	21±12	10±6	-	-	11±6	11±5	-	-
Nyssa sylvatica (Black tupelo)	615±192*	494±158*	319±77	270±67	379±55	258±41	360±58	266±48
Ostrya virginiana (Eastern hophornbeam)	152±94	143±60	180±85	139±63	260±80	87±21	202±44	190±49
Pinus echinata (Shortleaf pine)	1205±471*	707±323*	316±127	262±112	689±163*	534±118*	390±81	319±68
Prunus americana (Wild plum)	5±5	53±30	52±25	98±54	35±22	2±1	30±12	38±12
Prunus serotina (Black cherry)	89±28	107±24	86±20	43±11	110±22	91±16	106±17	49±10
Quercus alba (White oak)	294±72	191±50	472±93*	419±80*	217±39	184±32	272±38	284±48
Quercus falcata (Southern red oak)	358±94	-	539±138*	590±140*	152±36	-	194±37	124±34
Quercus marilandica (Blackjack oak)	452±72	148±35	111±32	245±49	468±59*	357±41*	489±62*	536±75*
Quercus nigra (Water oak)	-	44±18	5±3	13±7	-	99±27	223±56	121±35

Quercus rubra (Northern red oak)	284±88	22±17	343±76	416±100*	113±24	53±12	130±23	108±35
Quercus stellate (Post oak)	358±88	162±41	133±34	91±29	253±42	210±32	340±54	227±36
Quercus velutina (Black oak)	47±24	392±75	339±57	35±14	32±11	180±25	145±29	82±20
Qurecus phellos (Willow oak)	16±11	19±13	-	-	59±29	11±4	60±33	1±1
Rhamnus spp (Buckthorn)	5±5	-	62±20	-	9±8	74±26	19±11	2±2
Rhus coriaria (Sumac)	321±100	414±83	408±91*	184±94	440±66	266±40	362±59	130±35
Sassafras albidum (Sassafras)	163±70	44±14	109±32	18±8	30±13	18±6	38±14	7±4
Ulmus alata (Winged elm)	279±111	216±38	395±128	411±130	494±79*	347±58	538±83*	396±66*
Ulmus americana (American elm)	-	12±7	-	-	4±4	40±12	3±2	-
Ulmus rubra (Slippery elm)	26±21	-	2±2	-	45±20	1±1	-	7±7
Vaccinium arboreum (Sparkleberry)	-	-	168±48	177±71	-	60±17	488±78*	344±64*
Vaccinium spp (Blueberry)	116±42	10±8	-	-	299±75	29±15	-	-
Viburnum spp (Viburnum)	11±7	-	20±10	3±3	37±23	1±1	36±15	7±5

Note:

* represents that it is one of the five most common understory vegetation of that site at that measurement time

- represents that the species did not present at any plot at that measurement year

± sign separates the standard deviation (SE) with mean values

Note: Species which had missing records for entire measurement periods either in Ozark-St. Francis National Forest or in Ouachita National Forests or the species which had density less than 5 stem ha^{-1} for all measurement periods are not included in the table. Those species are *Platanus occidentalis* (American sycamore), *Robinia pseudoacacia* (Black locust), *Juglans nigra* (Black walnut), *Rubus fruticosus* (Blackberry), *Ceanthus cuneatus* (Buckbrush), *Quercus prinus* (Chestnut oak), *Quercus muehlenbergii* (Chinkapin oak), *Bumelia lanuginosa* (Gum bumelia), *Gleditsia triacanthos* (Honey locust), *Lonicera caprifolium* (Honeysuckle), *Quercus rugosa* (Netleaf oak), *Maclura pomifera* (Osage orange), *Castanea ozarkensis* (Ozark chinquapin), *Asiminia triloba* (Paw paw), *Toxicodendron radicans* (Poison ivy), *Ilex decidua* (Possumhaw), *Zanthoxylum americanum* (Prickly ash), *Morus rubra* (Red mulberry), *Betula nigra* (River birch), *Quercus shumardii* (Shumard oak), *Ulmus pumila* (Siberian elm), *Oxydendrum arboretum* (Sourwood), *Parthenocissus quinquefolia* (Virginia creeper), and *Salix alba* (White willow)

Table II-4 Percent of plots with high (≥1730 stems ha^{-1}) shortleaf pine density m^2 ha^{-1} at each thinning level in Ozark-St. Francis National Forest (OZNF) and Ouachita national forests (OUNF)

	Percent of plots with adequate shortleaf pine regeneration			
TL* (OZNF)	1996	2001	2006	2013
<10	40	30	40	30
Between 10 and 17	20	0	0	0
Between 17 and 24	10	0	0	0
≥24	0	0	0	0
TL* (OUNF)	1996	2001	2006	2013
<10	15.04	3.76	2.26	2.92
Between 10 and 17	15.04	2.26	2.26	1.46
Between 17 and 24	3	1.50	0.75	1.46
≥24	6	1.50	0	0

Note: TL=Overstory basal area thinning treatment level. These are the four levels designed to study the shortleaf pine regeneration performance at multiple overstory thinning levels

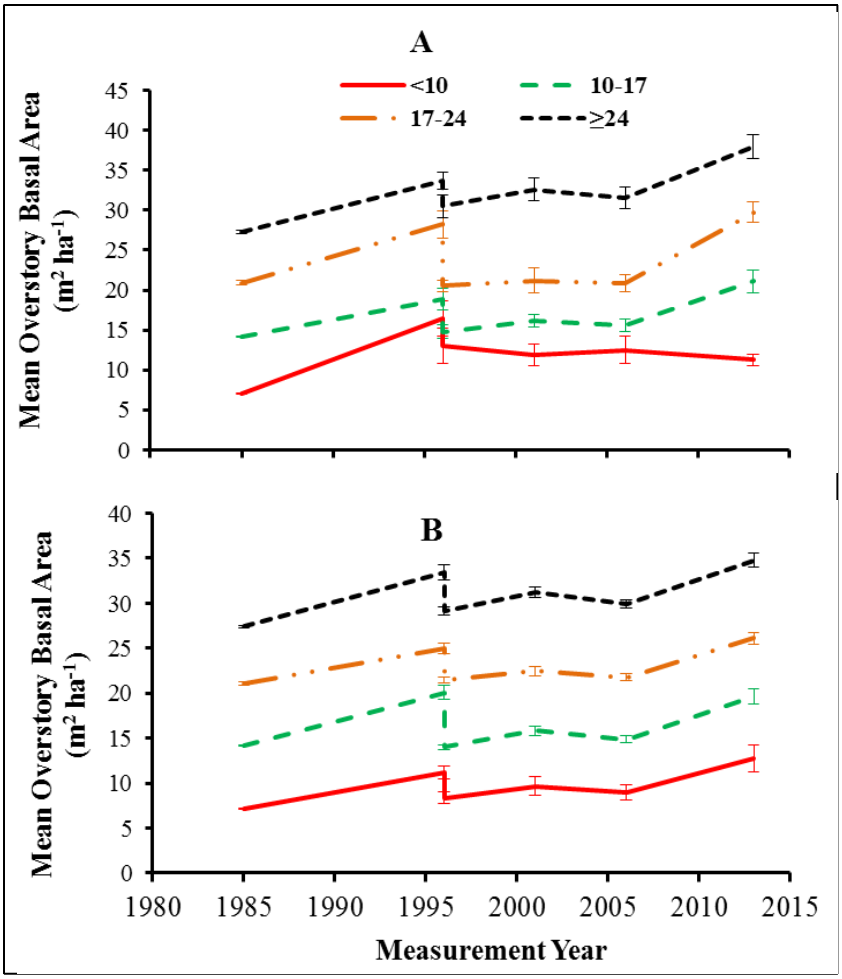

Fig. II-1 Mean residual overstory basal area (m^2 ha^{-1}) of shortleaf pine over various measurement periods at four thinning treatment levels. First thinning was conducted in 1985 to create plots with four distinct thinning levels. Plots were thinned second time after 1996 measurement period to maintain the overstory basal area level similar at 1985. A represents Ozark St-Francis National Forest (OZNF) and B represents Ouachita National Forest (OUNF)

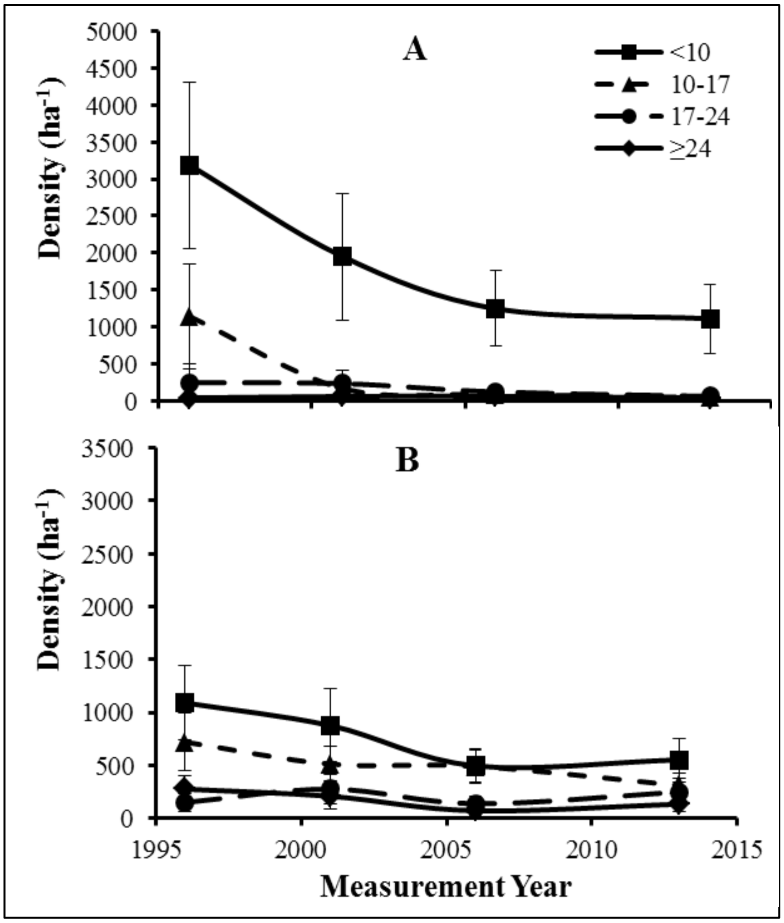

Fig. II-2 Regeneration density of shortleaf pine at four thinning treatment levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Four thinning treatment levels are created based on the overstory shortleaf pine basal area (m^2 ha^{-1}) after thinning.

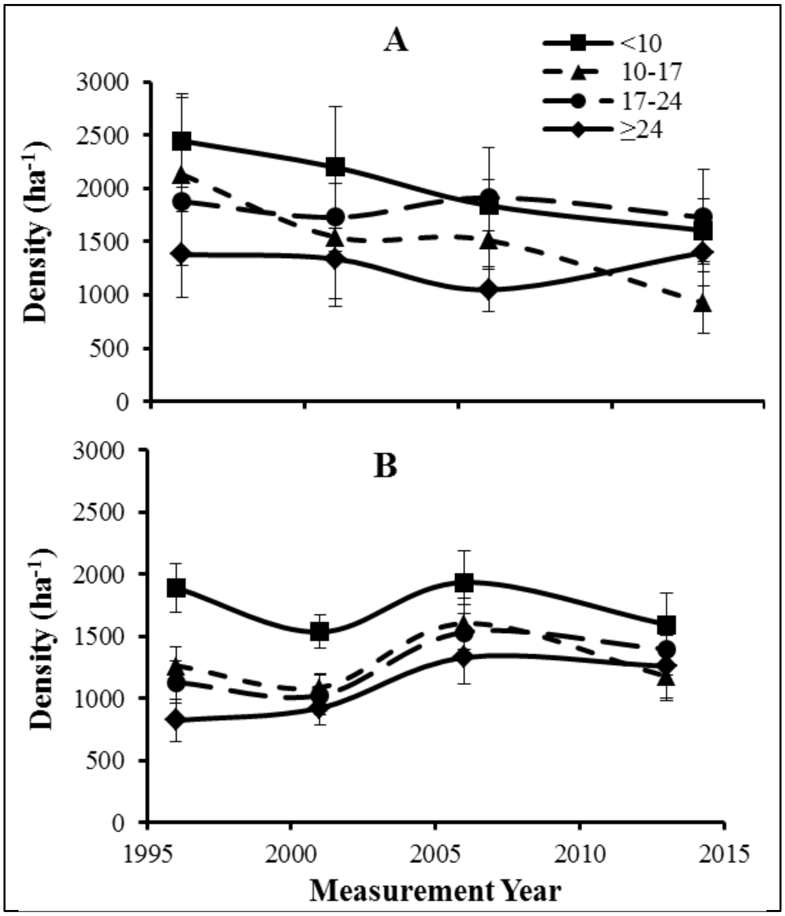

Fig. II-3 Regeneration density of oaks at four thinning levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B).

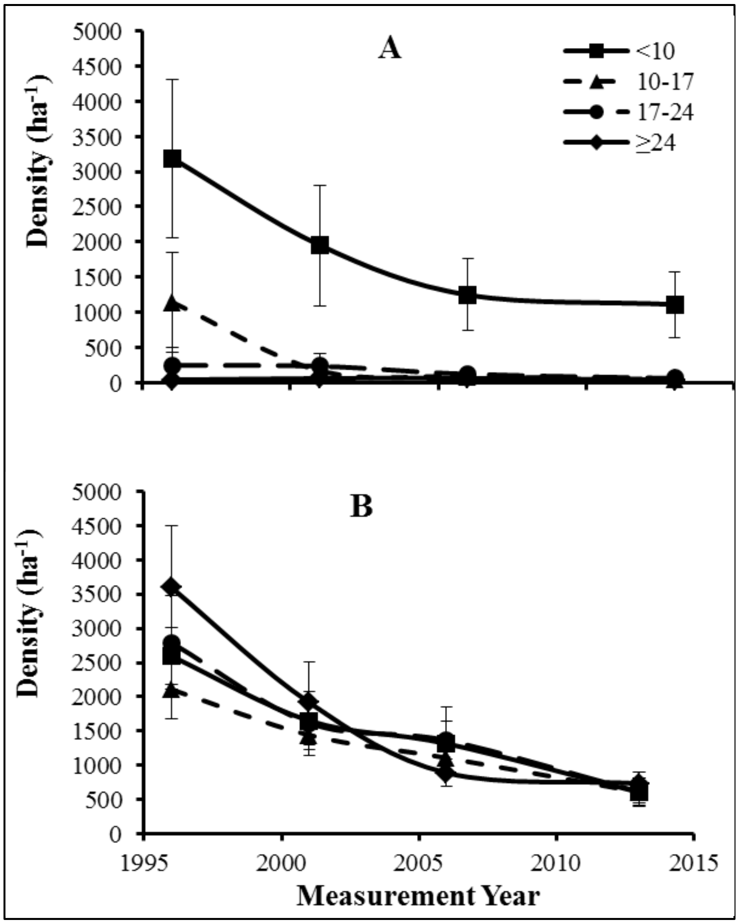

Fig. II-4 Regeneration density of red maple at four thinning levels over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B).

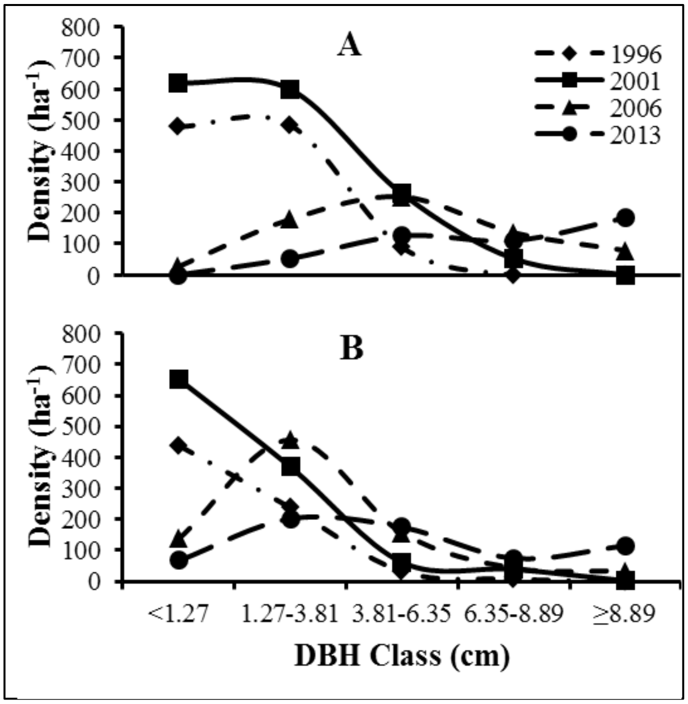

Fig. II-5 Density of shortleaf pine regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents shortleaf pine density (ha^{-1}) at particular dbh for that measurement year.

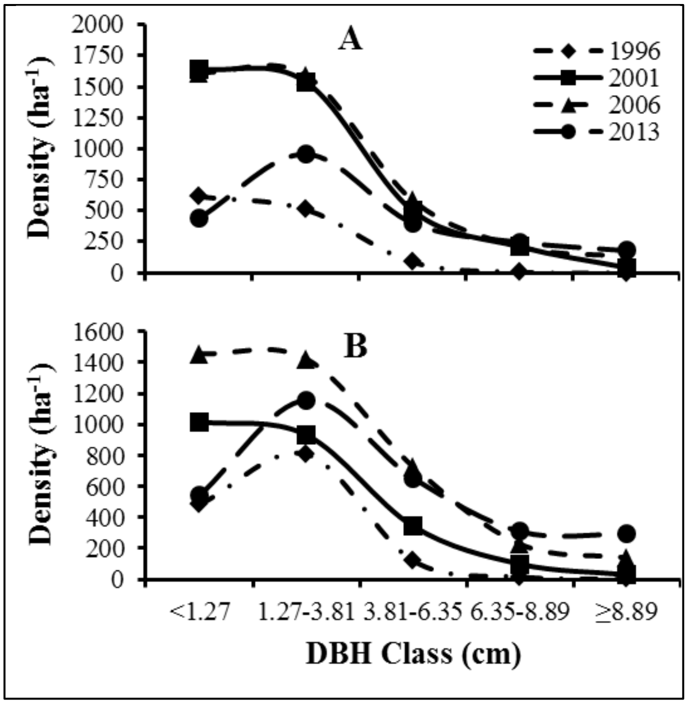

Fig. II-6 Density of oak regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents oak density (ha^{-1}) at particular dbh for that measurement year.

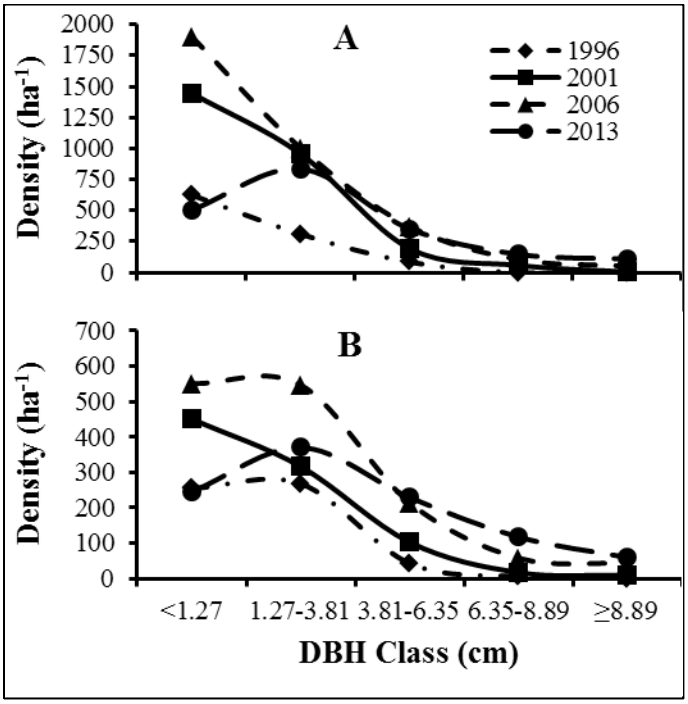

Fig. II-7 Density of red maple regeneration at various dbh classes over four measurement periods in Ozark-St. Francis National Forest (A) and Ouachita National Forests (B). Horizontal axis represents five dbh classes (cm) and the vertical axis represents red maple density (ha^{-1}) at particular dbh for that measurement year.

CHAPTER III

Predicting shortleaf pine regeneration (*Pinus echinata* Mill.) after thinning the overstory in Ozark and Ouachita mountain Forests: A Decision Tree Model Approach

Abstract

We propose the decision tree and logistic regression models to predict the shortleaf pine (*Pinus echinata* Mill.) regeneration in Ozark and Ouachita mountain forests of Arkansas and Oklahoma, and compare their performances using various fit statistics. We apply 3 forms of logistic regression (LR) and decision tree (DT) models to assess the effects of overstory shortleaf pine characteristics in association of climatic and topographic factors on shortleaf pine regeneration. We use shortleaf pine regeneration count data collected from the natural shortleaf pine forests of Arkansas and Oklahoma and spanning a period of 25 years after overstory forest plot establishment. Fit statistics such as misclassification rate (MR) and average square error (ASE) are used to select the best performing model that predicts the shortleaf pine regeneration. The overstory thinning levels, precipitation, site index, and age are the significant factors affecting shortleaf pine regeneration. The DT model using the Gini criteria as the splitting rule performed better than the LR models to predict the shortleaf pine regeneration with the lowest MR of 7.6 percent. The satisfactory shortleaf pine regeneration density (>1730 stems ha^{-1}) was considerably high in the plots (20.47%) with high thinning level than the plots (1.64%) with low thinning levels. Though the primary purpose of thinning is not to improve the understory regeneration, the

present study suggests that thinning has a strong positive impact on shortleaf pine regeneration. Poor shortleaf pine regeneration performance over decades in study sites suggests the future of shortleaf pine dominated forest is questionable unless further regular silvicultural treatments are applied. The DT model can be a simple, efficient and accurate method to assess the effect of multitude of factors on shortleaf pine regeneration and to make the best possible shortleaf pine stands management decisions.

Keywords: decision tree, logistic regression, shortleaf pine regeneration, misclassification rate, thinning

1. Introduction

In recent years, predictive modeling/ machine learning techniques have emerged as alternatives to traditional regression modeling approaches because of their flexibility, speed and accuracy (Aquino et al., 2008). These techniques use several artificial intelligence (AI) algorithms, such as classification and regression trees (CART), artificial neural networks (ANN), support vector machines (SVM), ensemble models, and others to obtain the better fits (Aquino et al., 2008). Predictive modeling techniques also facilitate the data collection, management and cleaning process (Piramuthu, 2004). Large-size, longitudinal data collection and model development in a limited time and with limited resources is a significant challenge to researchers. In many studies with relatively larger data sizes, researchers use around 80% of their time and resources on data cleaning and preprocessing (Piramuthu, 2004; Tirelli and Pessani, 2011). The application of predictive modeling techniques not only offers better fit but also provides simple and precise methods that solve complex data management and modeling issues. These models have substantial future promise in assessment and interpretation of non-linear patterns that we often encounter in forest measurements data.

Decision tree (DT) models as a predictive modelling approach have been successfully applied for various purposes such as predicting plant ecological properties (e.g. Lees and Rittman, 1991), soil abiotic properties (e.g. Bui et al., 2006; Kim and Park, 2009; Kim et al., 2011) and rainfall runoff studies (e.g. Valipour et al., 2013; Valipour, 2015). These models offer an advantage of splitting the complex data into groups. To the best of our knowledge, DT models have not been used in any kind of shortleaf pine (*Pinus echinata* Mill.) regeneration prediction studies despite their simplicity and advantages. Therefore, we introduce the DT model to assess the effects of overstory shortleaf pine characteristics and other climatic and topographic factors in shortleaf pine regeneration in Ozark and Ouachita mountain forests located in Arkansas and Oklahoma. We hypothesized that the DT model will provide a better fit and simpler approach to predict the shortleaf pine regeneration in the region.

Historically, shortleaf pine forests have been one of the most common forest types in southeastern United States (Kabrick et al., 2010; McWilliams et al., 1986). Shortleaf pine is considered one of the most important tree species in Arkansas and eastern Oklahoma (Zhang et al., 2012). Shortleaf pine has been desirable in the region in terms of timber production for southern pine lumber, which is primarily used in the housing industry. Shortleaf pine is also particularly desirable for red cockaded woodpecker habitat (Zhang et al., 2012) from the wildlife management perspective. Despite its importance, shortleaf pine populations have been declining in recent years (Moser et al., 2006; Lilly et al., 2012; KC et al., 2015; KC et al., 2016). KC et al. (2015) suggested that the current rate of shortleaf pine regeneration is not adequate to maintain the shortleaf pine dominated forests in long-term in Ozark and Ouachita Mountains of Arkansas and Oklahoma. In many shortleaf pine-hardwood mixed natural forests, hardwood regeneration is dominant compared to that of shortleaf pine. Low levels of shortleaf pine regeneration for the long term and, meantime, dominance of hardwood tree species as the understory vegetation greatly affects the sustainability of shortleaf pine dominated forests.

In light of the fact that shortleaf pine regeneration is low in the region, this study evaluates the effects of overstory stand level variables (site index, plot age, overstory basal area per hectare) and other climatic (precipitation) and topographic (slope, aspect, altitude) factors on shortleaf pine regeneration in Ozark and Ouachita National Forests in Arkansas and Oklahoma. We predict the chance of shortleaf pine regeneration at satisfactory levels using several decision tree models representing various circumstances.We also compared the predictive performance of the selected DT models to LR models. Specifically, we apply the decision tree model to assess the shortleaf pine regeneration response to overstory thinning in the long term. Additionally, we illustrate an interactive DT where the forest managers can interactively change the inputs to achieve the desired number of shortleaf pine regeneration stems in their forests. Furthermore, we examine the association between shortleaf pine regeneration and thinning level over a period of 20 years. We expect this study to be helpful to manage the shortleaf pine forests not only in Arkansas and Oklahoma but also in the entire southeastern United States. Most importantly, this study will establish a precedent that the predictive models are helpful in forest management related research which supports stakeholder decision making.

2. Materials and Methods

2.1. Study area and data collection

The USDA Forest Service Southern Research Station and the Department of Forestry (now part of the Department of Natural Resource Ecology and Management) at Oklahoma State University jointly established permanent study plots in the Ozark and Ouachita National Forests during 1985 to 1987. Study plot locations range from the Ozark National Forest near Russellville, Arkansas (latitude 35.3º N, longitude 93.1º W) to areas on the Ouachita National Forest near Broken Bow (latitude 34.0º N, longitude 94.7º W) in southeastern Oklahoma (Lynch et al., 2003). This study was established to assess the effect of thinning on the growth and development of overstory and understory shortleaf pine forests. Results based on the overstory characteristics only (Budhathoki et al., 2006; Budhathoki et al., 2008; Budhathoki and Lynch, 2008; Budhathoki et al., 2010) have been published in past. To date, one study has utilized the understory data

(Lynch et al., 2003) to predict shortleaf pine regeneration that used two measurement periods data only.

Two 20.23 m^2 subplots were established in 1996 within each 0.081 ha overstory measurement plot to measure the understory woody-vegetation. Hardwoods and shortleaf pine regeneration located inside the subplots and taller than 1.37 m in height were measured. Only two subplots were measured in 1996 however during all subsequent measurements a total of four subplots were measured within each 0.081 ha overstory measurement plot. Hereafter, the 1996 measurement of understory is termed the first measurement, 2001 as the second, 2006 as the third and 2013 as the fourth measurement. We used 182 permanent plots for this study which include 133 plots from Ouachita National Forest and the 47 plots from Ozark National Forest. We eliminated two plot records from the dataset because of the missing overstory information. In the winter of 2000, ice storms heavily damaged the shortleaf pine study plots in Ouachita National Forests (Stevenson et al., 2016). Therefore, we eliminated 22 study plots from the subsequent measurements that had the overstory shortleaf pine damage greater than 40 percent as described by Saud et al. (2016).

Overstory measurement plots were circular and 0.081 ha in area with a 16.06 m radius. While establishing the plots, the understory hardwoods greater than 2.54 cm in diameter at ground level were eliminated using chemical herbicide. The measurement plots were isolated with 10.06 m buffer area. The isolation buffers had the same thinning and herbicide treatments as the measurement plots. This was done so that the entire interior measurement plots experienced similar levels of competition. Understory woody-vegetation including shortleaf pine started regenerating a few years after the plot establishment. At establishment period, each plot was thinned from below to specified residual basal areas ranging from 3.97 m^2 ha^{-1} to 48.68 m^2 ha^{-1}.

Most plots were thinned for second time after third overstory measurement in 1996. The purpose of second thinning was to reduce overstory shortleaf pine basal area levels to levels similar to those after the first thinning in 1985. Overstory shortleaf pine characteristics including diameter at breast height (DBH), age, site index were measured at approximately five year intervals from the time of plot establishment to the most recent measurement ending in 2013. We used the geographical positioning system (GPS) location of each plot to extract the topographic information such as altitude, slope and aspect. Similarly, GPS locations were used to access climatic information including precipitation amount for each plot. Table III-1 provides a list of all variables used for modeling the shortleaf pine regeneration along with scale and their range of values.

2.2. Model building process

The main objective of the present study is to build an accurate predictive model for shortleaf pine regeneration. In the predictive modeling literature, there are several alternative models which we can use to achieve this objective. A common feature among all predictive modeling techniques is that they try to find the best fitting rules for predicting the values of one or more variables in a data set, usually called outputs, from the values of other variables in the same data set, commonly referred to as inputs. This study focuses on the usage of two well-known predictive modeling techniques, namely, Logistic Regression (LR) and Decision Tree (DT) models. In the following two subsections, we provide a general description for each of the two modeling techniques and summarize the steps of model generation.

2.2.1. Logistic Regression (LR) Models

Regression analysis is one of the most popular techniques used for predictive modeling. When the output (also called; target or response) variable is categorical, the LR model is often used. The theory of both binomial (for binary targets) and multinomial (for categorical targets with more than two categories) LR models is well-established and used in ecological, medical, business studies, and in many other research studies. Since the target variable for the present study is binary (Low/High) as it will described latter in coming sections, we focus on binomial LR models. A LR model with more than one input variable (also called explanatory or independent variable) has the following form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots.. + \beta_k x_{ki} \tag{1}$$

Where, $p_i = P(Y_i = 1 \mid X_{1i} = x_{1i}, \ldots\ldots.., X_{ki} = x_{ki})$, Y is the binary target variable taking values 0 or 1, the X 's are input variables, β 's are model parameters to be estimated. Using the parameter estimates of the model for any given set of values for input variables, we can estimate the probability (p_i) that the target is 1 (Y) and hence we can classify new observations into one of two categories. The estimation of parameters in the LR model is performed using the maximum likelihood method. Several variable selection techniques, including stepwise, forward or backward selection, can be used to determine which inputs should be retained in the final LR model.

2.2.2. Decision Tree (DT) Model

A DT maps observations (inputs) about an item to conclusions about the item's target value. There are several types of DT models, including classification and regression tree (CART), chi-squared automatic interaction detector (CHAID), C4.5 and MARS. Both classification trees and

C4.5 are mainly used to model categorical target variables while regression trees are structured to model numerical targets (Dhar, 2011). CHAID models perform multi-level splits while generating the decision tree and hence they are more complicated than CART models. And, MARS models were developed to generate more accurate decision trees for numerical targets (Dhar, 2011).

In general, a DT is a flowchart-like structure consisting of nodes and directed edges. A simple hypothetical DT is displayed in Fig. III-1. There are three types of nodes in the chart, namely root node, internal node and leaf (terminal) node. The root node has no incoming edges and it has zero or more outgoing edges. An internal node has exactly one incoming age and it has two or more outgoing edges. Leaf (terminal) node has exactly one incoming edge with no outgoing edges, and it represents class label; assuming the target variable is coded as classes. Each internal node represents a test on one of the inputs whereas each directed edge (branch) represents the outcome of the test. The path from the root node to a leaf node represents classification rules.

"There are exponentially many DTs that can be constructed from a given set of attributes. While some of the trees are more accurate than others, finding the optimal tree is computationally infeasible because of the exponential size of the search space" (Tan et al., 2006). However, efficient DT models can be developed using well-established algorithms. Almost all existing DT models use the split-search algorithm (also called Hunt's algorithm) to grow DTs **(**Tan et al., 2006). This algorithm cultivates DTs by performing two steps repeatedly. Letting D_t denote the set of training observations that reach at node t and $(c_1, c_2, \ldots\ldots, c_m)$ denote the class labels: i = 1 to m, the algorithm can be summarized in the following:

Step 1. If all records in D_t belong to one class, c_i, then t is a leaf node assigned label c_i. Another case where we stop splitting, so that the node is considered to be a leaf node occurs if D_t has a small number of records.

Step 2. If the records in D_t belong to more than one class, a test on one of the inputs is applied to partition the records into subsets. Steps 1 & 2 are done in every generated node until all nodes are leaf nodes or the maximum number of splits is reached.

In order to grow an efficient DT model in a reasonable amount of time, one should find a way to determine the best split of the records in each node. There are many measures to identify the best split. For categorical targets, Gini, entropy and chi-squared logworth are three commonly used measures for evaluating split worth (Tan et al., 2006). Variance and ProbF logworth are designed for interval targets (Tan et al., 2006). In this section, we will describe the first three measures for developing DT models because our target variable is binary.

The Gini and entropy are based on the degree of impurity of splits and hence are defined in terms of the class distribution of the records before and after splitting. The more skewed the class distribution, the smaller the degree of impurity. For instance, a node with class distribution (0, 1) has zero impurity whereas a node with class distribution (0.5, 0.5) has the highest impurity. Since each criterion uses a different philosophy to determine the best split, each grows a different style of tree.

The Gini measure attempts to separate classes by focusing on one class at a time. Once the first split is made, Gini continues attempting to split the data that require further segmentation. Since Gini is so often the best splitting rule, it is the default rule in CART. Let $p(c_i \mid t)$ denote the proportion (p) of records in the data belonging to class c_i at node t. For binary targets (0, 1), we have only two classes (c_1, c_2) at any node and we can use the notation $p_0 = p(c_1 \mid t)$ and $p_1 = p(c_2 \mid t)$ to represent the fraction of records belonging to each class. Using this notation, for a target variable with m classes, Gini is calculated as follows:

$$\text{Gini} = 1 - \sum_{i=1}^{m} [p(c_i \mid t)]^2 \qquad (2)$$

where, m is 2 due to binary classification, Gini varies between 0 to 0.5; where, Gini equals 0 if the node is perfectly pure and equals 0.5 if the class distribution is uniform $(p_0 = 1 - p_1 = 0.5)$.

On the other hand, the philosophy of entropy is different. Rather than initially pulling out a single class, entropy first segments the classes into two groups, attempting to find groups that together add up to 50 percent of the data. Entropy then searches for a split to separate the two subgroups. Entropy can take any value in the range $(0,1)$ where the smaller the value of entropy, the smaller the impurity of the split. Entropy can be calculated using the following formula:

$$\text{Entropy} = -\sum_{i=0}^{m-1} p(i \mid t) \log_2 p(i \mid t) \qquad (3)$$

where, $0 \log_2 0 = 0$; other notations are as described above.

Finally, the goodness of each test condition should be evaluated using some objective measure. A natural way to determine how well a test condition performs is to compute the difference between the degree of impurity of the parent node and the degree of impurity of the child nodes. This difference is called the gain. Good test conditions are expected to have higher gain value. Let t be the parent node under splitting, $(t_1, t_2, \ldots.., t_k)$ be the resulting child nodes, $N(t)$ the number of records in the parent node and $N(t_i)$ is the number of records in the i^{th} child node. The gain, Δ, can be calculated using the following formula:

$$\Delta = I(t) - \sum_{i=1}^{k} w_i I\,(t_i) \tag{4}$$

Where, $I(t)$ is the impurity measure (i.e. Gini or Entropy) of a given node and $\mathrm{w}_i = N(t_i)/N(t)$ is the weight of the i^{th} child node.

2.2.3. Model comparison

There are many criteria to compare the performance of competing models, including Misclassification rate (MR) and average squared error (ASE). The MR is defined as the proportion of disagreement between the predicted outcome and the actual outcome, i.e. the number of misclassified records divided by the total number of records, while the ASE is given by

$$\mathrm{ASE} = \frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2 \tag{5}$$

Where, y_j is the actual j^{th} value of the target output y, $\hat{y}_j$ is the predicted j^{th} value for the target output y and n is the total number of records of the target output in the data. Smaller values of MR or ASE provide better model performance. Both measures are used for model selection in section 4. Additionally, the MR is utilized to detect overfitting when growing DT models. Overfitting happens when the MR of validation data exceeds the MR of training data. Splitting of the DT must be stopped before the overfitting starts.

Since our target output has a binary response, we used the *receiver-operating-characteristics* (ROC) curve to investigate the relative performance of different candidate models. The ROC graph is a two-dimensional plot with (1-specificity) on the x-axis and sensitivity on the y-axis and the area under the cure (AUC) measures the model discrimination ability. Sensitivity measures the ability of the model to correctly classify subjects with positive target output as positive whereas specificity measures the resistance of the model against misclassifying subjects with negative target output as positive. Denoting *true positive* (*true negative*) by TP (TN) and *false positive* (*false negative*) by FP (FN), sensitivity=TP/(TP+FN) and specificity=TN/(TN+FP). The trade-off between model sensitivity and specificity is represented by the ROC curve. The AUC is used as a measure of model accuracy in many applications (Swets 1988).The closer the curve to the top left corner of the ROC space, the higher the accuracy of the model.

2.3. Data preparation

We used SAS Enterprise Miner version 12.3, SAS EM hereafter, for data preparation and model development. We assigned the regenerated shortleaf pine stem density ha^{-1} as a target variable. Other variables such as plot age, overstory basal area, measurement years, thinning class, site class, annual precipitation (mm) , altitude (m) and aspect were assigned as the input variables (Table III-1). Site, thinning class and year of measurement were assigned as categorical inputs. Originally, the target variable (SLP) contained the shortleaf pine regeneration density ranged from 0 to 13,344 stems ha^{-1} (Table III-1). Later, we assigned densities into two classes so that response variable can fit into binary LR and DT models. Regeneration densities of 1730 stems ha^{-1} or less were assigned “low regeneration” and the densities greater than 1730 stems ha^{-1} were assigned as “high regeneration”, and the new binary variable is denoted as SLPN. A previous study has suggested that the regeneration of shortleaf pine greater than 1730 stems ha^{-1} is a satisfactory regeneration, and anything less than 1730 represents a poor regeneration (Lynch et al., 2003) for naturally-occurring shortleaf pine. Aspect was transformed into NORTHNESS and EASTNESS using trigonometric functions (Roberts, 1986) where NORTHNESS is cosine and EASTNESS is sine of aspect. We assigned four thinning levels (A, B, C, D): less than 10.332 m^2 ha^{-1} as A, between 10.332 and 17.22 m^2 ha^{-1} as B, between 17.22 and 24.108 m^2 ha^{-1} as C, and greater than 24.108 m^2 ha^{-1} as D respectively based on the residual overstory shortleaf pine basal area at the time of plot establishment period 1985-1987 (see, Lynch et al., 1999). Eight of the plots had missing information on climatic and topographic variables. Those values were imputed using the mean value. Table III-2 introduces some descriptive measures of interval variables in the data.

Additionally, the symmetry assumption has been checked using measures of skewness and kurtosis. The symmetry for the interval inputs was not violated except for the variable slope which has the skewness of -2.943 (Table III-2). Therefore, we conducted a transformation for slope using max-normal technique in SAS EM which automatically selects the most appropriate transformation and creates new transformed variable. The transformed slope (SLOPEN) had a roughly symmetric distribution with skewness of -0.96.

We randomly separated the original dataset into training and validation datasets, and assigned 70% of the data into the training and the 30% to validation data set as described by Tan et al. (2006) and Sarma (2013). Later, the training data set was used to develop the predictive models throughout the model building process, and the validation data set was used to evaluate the performance of models built using the training data set. Fig. III-2 summarizes the main points discussed in section 2.2 & 2.3 diagrammatically.

3. Data analysis

In order to identify the major factors affecting the shortleaf pine regeneration level, both LR and DT models are developed using the data described in section 2.1. As mentioned earlier, all models are built using the training data, then their relative performance is evaluated using the validation data. In all models, the binary variable SLPN is set as the target. Variables, BA, AGE, SI, SITE, YEAR, THINNING, ALTITUDE, PRECIPITATION, SLOPEN, EASTNESS and NORTHNESS, described in Tables III-1 and III-2, are considered as potential inputs in each model. In this section, we will describe all models that we have built for predicting the shortleaf pine regeneration.

As for LR models, three models were developed and called as LR1, LR2 and LR3 models. The logit link function, introduced in Eq. 1, was used in all three models. In LR1 model, all inputs (Table III-2) were entered. Similarly, in LR2 model, we used all inputs as in LR1 but the stepwise selection method ($\alpha = 0.05$) was applied for selecting significant inputs to be kept in the final model. In the third model, LR3, two factor interaction and polynomial terms were tested using the stepwise method. Both validation ASE and MR were used to determine the best performing model among the three LR models.

Alternatively, two DT models were developed for the same objective. The first model (DT1) used Gini as the splitting criteria and the second one (DT2) used entropy. At each partitioning opportunity, the maximum number of splits was controlled at 2. One may want to increase the maximum number of splits to obtain additional modeling resolution. The maximum tree depth was controlled at the default level, 5. Therefore, we did not get any DT with more than 2 branches and a depth more than 5. The steps of data preparation and analysis in SAS EM are displayed in Fig. III-3. Sarma (2013) has described the DT model development methods in detail. In the following section, we introduce the results of the five models and select the best model to be applied for future predictions.

4. Results

This section summarizes the main results of all models described in section 3. First, we introduce the relative performance of the LR models and report the results of the best LR model. The results of DT models are reported in a similar fashion. Next, we compare the accuracy of best DT and LR models. We further emphasize the effect of thinning on the shortleaf pine regeneration.

4.1. Results of LR models

The performance of LR1, LR2 and LR3 models are summarized in Table III-3. All three models are statistically significant ($P<0.001$) but the stepwise LR2 model has the smallest validation MR and ASE (Table III-3). The LR2 model shows that the additive effect of AGE, SI, and PRECIPITATION in conjunction to thinning levels has the significant influence on shortleaf pine regeneration (Table III-4). Estimated coefficients, odds ratios and significance levels of each of the four factors are presented in Table III-4. The odd ratios of obtaining better shortleaf pine regeneration in the plots with low residual basal area (thinning level A) was 8.0 times higher than in the plots with high residual basal area (thinning level D)

4.2. Results of DT models

The default decision tree (DT1) model for the classification of shortleaf pine regeneration is displayed in Fig. III-4. To build the DT1model, all 11 input variables from training data set are used. The accuracy of the model is then assessed using the validation data set. From Fig. III-4, it is readily seen that this tree includes a total of 9 nodes from which 5 are leaf nodes; nodes 3, 5, 7 & 8 are true leafs, where no further splitting was performed as the nodes have a high purity level, while node 9 is stopping node. Summary statistics from both training and validation data sets are given for each node. The main statistic is the Gini value which reflects the purity level of each node. This DT uses only three input variables as given in Table III-5. Using these variables, four splits were made resulting in the validation MR of 7.6% which was computed by applying the DT1 model on the validation data. Classification details of DT1 model are given in Table III-6. Using entropy as the splitting criterion (DT2) gives the exact same results as DT1 (Table III-5).

Table III-7 summarizes the decision rules extracted from DT1 model. According to these rules, the shortleaf pine regeneration rate for a given plot can be classified as low or high after checking the status of three inputs (BA, SI and PRECIPITATION). Suppose, for example, a forest manager wants to predict the shortleaf pine regeneration level of a plot that has basal area of 15m^2 ha^{-1}, site index of 12 m and average annual precipitation of 1200 mm. Then using Fig. (III-4), starting from node 1, we see that the basal area test is satisfied (i.e., BA<18.771 m^2 ha^{-1}) and that leads us to node 2. Next, checking SI at node 2, DT shows the condition holds (i.e., SI<19.501m) and thus we move to node 4. Since the precipitation level is 1200 mm, the precipitation test at node 4 leads us to node 6. Finally, using the SI as the splitting criterion directs us to node 8 because the given site index is below 15.804m. As a result, the model DT1 predicts the given plot will have a high level of shortleaf pine regeneration with the MR of 7.6%.

The result of interactive DT can vary based on the interest of the forest managers. Here let's imagine an example. A forest manger somewhere in eastern Oklahoma has a shortleaf pine stand with the site index of less than 25 m and overstory basal area of less than 20 m^2 ha^{-1}. He/she assumes that the average annual precipitation is below 1100 mm $year^{-1}$ in the region. Here, the manager would like to estimate how different will be the regeneration in this site compared to the other sites where the basal area, site index and the precipitation are higher. We developed an interactive DT, utilizing the provided variable information, to estimate the shortleaf pine regeneration. Fig. III-5 shows the results of this hypothetical scenario.

4.3. Comparison of DT and LR models

In this section we compared the performance of all five models which have been discussed in the previous two sections. Using the validation MR as our criterion, we conclude, from Tables III-3 and III-5, that the default Gini decision tree (DT1) is the best model since it has the lowest MR among the five models. Another popular tool for model comparison is called the ROC curve which has been described in section 2.2.3. Fig. III-6 compares the ROC curves for two models (LR2 and DT1) for training and validation data. For the training data, DT1 performed slightly better. Model DT1 outperforms the model LR2 under the validation data. Thus, the results suggest that, in general, the decision tree models are viable alternatives to the logistic regression models to understand shortleaf pine regeneration patterns and in predicting its levels.

4.4. Effect of thinning on shortleaf pine regeneration

In section 4.1, thinning appeared as an important input for predicting shortleaf pine regeneration in the model LR2. As we mentioned earlier, plots in thinning level A (A=

BA<10.332 $m^2 ha^{-1}$) are about nine times more likely to have high regeneration than plots in thinning level D (D=BA≥ 24.108 $m^2 ha^{-1}$). For further exploration of the association between the SLPN and THINNING, we used Gamma (Γ) test as the measure of the strength of association between ordinal variables (Agresti, 2007). The percentage distribution of shortleaf pine regeneration levels (low regeneration vs high regeneration) along the thinning levels (A, B, C, and D) are displayed in Table III-8. For the distribution in Table III-8, the value of this measure is $\Gamma = -0.5474$ ($P < 0.001$) which implies that there exists a moderate but significant negative relationship between thinning levels and the shortleaf pine regeneration level. It suggests thinning level A has the highest and the level D has the lowest shortleaf pine regeneration.

4.5. Effect of thinning and time factor on shortleaf pine regeneration

Here, we applied the interactive DT model to assess how different the regeneration patterns are at various thinning levels over four measurement periods. Firstly, the DT was split based on the thinning levels. Overall, shortleaf pine regenerated at satisfactory level (High regeneration) in 20.49%, 7.83%, 6.03% and 4.08% of plots at thinning levels A, B, C and D respectively. Further we split the nodes based on the measurement years. Plots with high regeneration are a lot higher in thinning level A compared to the thinning level B, C and D (Fig. III-7). There is no single plot in thinning class D in third (2006) and fourth (2013) measurements that has the high regeneration (Fig. III-7).

5. Discussion

In the present study, we mainly assessed what factors have the most important effects on shortleaf pine regeneration in the long-term, and which model structure is the most accurate and potentially easiest to use. Undoubtedly, sufficient shortleaf pine stems need to be regenerated to assure the future of shortleaf pine dominant forests in southeastern United States. Current data shows that shortleaf pine regeneration on the sites in this study is critically low. Only 7.80 percent of the plots have shortleaf pine regeneration more than 1730 stems ha^{-1}.

Logistic regression models are commonly used to predict the probability of a categorical response variable in ecological studies (e.g. Lynch et al., 2003; Perry and Thill, 2008; Bisquert et al., 2012). On the other hand, DTs and ANNs are the most frequently used AI algorithms in ecological and environmental studies (Kim and Park, 2009; Kim et al., 2011). These algorithms are relatively accurate and stable (Vayssieres et al., 2000; Yang et al., 2003; Zhang et al., 2005; Moret et al., 2006; Sesnie et al., 2008). Successful past DT applications to ecological and environmental problems provide the motivation to strongly consider the DT models to study the effect of thinning and other inputs on shortleaf pine regeneration. Developing multiple forms of LR and DT models and comparing their performances provides insight on assessing the shortleaf pine regeneration in Arkansas and Oklahoma, USA.

The LR2 model (Table III-4) indicates a slightly negative effect of precipitation (odds ratio=0.989) but a highly negative effect of site index (odds ratio=0.682) in reducing the odds of high regeneration of shortleaf pine. As the site index increases by one unit, the odds of high regeneration decrease by 31.8%. Lawson (1986) and Lynch et al (2003) described similar negative effect of SI on shortleaf regeneration. This may be the case because high site index indicates better site which favors growth of a hardwood understory relative to shortleaf pine. Though it seems counterintuitive that higher precipitation amount would reduce the odds of shortleaf regeneration, this too may occur because higher precipitation levels favor growth of the hardwood understory at the expense of shortleaf regeneration. Plot age shows a slight but positive effect on shortleaf pine regeneration (odds ratio=1.024) which implies that the older the trees in the plot, the higher the odds of having high regeneration. Lynch et al. (2003) described that the plot age is another factor that can be used predict the success of shortleaf pine regeneration. Similarly, plots located in thinning level A are 8.703 times more likely to have high regeneration than plots located in thinning level D. Plots located in thinning levels B or C do not differ significantly, with respect to regeneration rate, from plots located in thinning level D. Lynch et al. (2003) mentioned that increasing amounts of overstory basal area affects the shortleaf pine regeneration negatively. The present study shows similar results. Here, we not only evaluate the effect of overstory basal area on regeneration but also at four thinning levels.

If we apply only the decision rules from default DT, the percentage of plots with high regeneration increases to 15.89%, 28.45%, 43.06% and 92.31% respectively at node 2, node 4, node 6, and node 8 (Fig. III-4). An MR of 7.6% is impressively low for the data of this kind. There are some variables, like PRECIPITATION that can't be controlled. But, model still provides the idea how the precipitation affects the shortleaf pine regeneration. Apparently intensive silvicultural treatments including aggressive hardwood control would be needed to be applied by forest managers to substantially increase regeneration levels. Therefore, DT models are simple and helpful to assess the present status of shortleaf pine regeneration, and also to determine the factors that are affecting the regeneration. Applying the rules suggested by the selected default DT model implies that thinning to sufficient levels and aggressive hardwood control could increase shortleaf pine regeneration in Arkansas and Oklahoma. We can always skip a variable that is not feasible to apply and try another one to achieve the similar results. An additional virtue of the DT is that it can be presented as a simple flow chart that is rather easy for forest managers to understand even if they have a limited statistical background.

The DT model can be modified in several ways to predict the target variable. The flexibility of the model makes the results managerially appealing. Here, an interactive DT (Fig. III-5) with variables SI, BA and PRECIPITATION is an example where we select what input variables to use and at what point to split. There is no single concrete DT model with fixed parameters. Forest managers can build the best DT model possible to answer their own sets of questions. By contrast, LR models do not offer such flexibilities. Here, as a forest manager, we focus on what variables play the most important role on shortleaf pine regeneration. The DT evaluates the importance of variables splits the tree on this basis. Fig. III-5 shows that if the site index is less than 25 m, overstory basal area is less than 20 m^2 ha^{-1} and if that particular site gets annual precipitation of less than 1100 mm, there is 60 percent chance that the plot has shortleaf pine density more than 1730 stems ha^{-1}. In many cases, the default model simply provides the best splitting options with lowest MR because machine selects the best possible purity of Gini. However, a forest manager might have other questions that are not specifically answered by the default DT model. In that case, the DT building process can be adjusted until the manager obtains the most desired result.

Thinning is a common silvicultural practice to manage the forests, but the primary purpose of thinning is not to promote the understory regeneration. However, this study indicates thinning has a great importance on overall understory regeneration. The declining percentages of "high regeneration" with plots thinned to low levels of residual basal area demonstrate the importance of overstory thinning on shortleaf pine regeneration. It is important to report how the time factor after thinning affects the shortleaf pine regeneration. Many plots with high thinning levels (low residual basal area) have consistently maintained "high regeneration" for 25 years. But, plots that were thinned at lower levels (high residual basal area) have critically low percentages of plots with "high regeneration". Interestingly, not a single plot has "high regeneration" after around 20 years of thinning in thinning level D (Fig. III-7). This clearly reflects the importance of thinning, and of course time interval after thinning, in maintaining high shortleaf pine regeneration. A significant impact of overstory shortleaf pine thinning over understory shortleaf pine is undeniable; moreover, thinning levels and time interval for thinning also have a significant impact on shortleaf pine regeneration. LR models did not show the significant effect of time (YEAR) after thinning on shortleaf pine regeneration. However, the time interval effect can be used in DT models to assess the high and low shortleaf pine regeneration patterns at various thinning levels. This demonstrates the advantage of DT models over the traditional LR models to predict shortleaf pine regeneration.

Shortleaf pine is shade intolerant species, and heavy thinning opens up more soil surface area for regeneration. When site index is good, other hardwood species utilize the regeneration opportunity more vigorously that shortleaf pine. We often find shortleaf pine regeneration better in poor sites where hardwood competition is less intense. A similar pattern may be occurring with precipitation trends. The effect of site index is high. The logistic regression model (LR2) did not indicate a significant difference on thinning level D compared to level B and C. But, the regeneration trend is different among these classes. In contrast to the LR2 model, the DT models clearly show how different the regeneration pattern is based on the thinning levels. This supports the contention that predictive modeling techniques such as DT have the potential to be useful to better understand the forestry data in general.

6. Conclusions

The DT models have many attractive attributes and performed better than LR models for prediction of shortleaf pine regeneration in Ozark and Ouachita national forests of Arkansas and Oklahoma. The models demonstrated that overstory shortleaf pine thinning positively affects the understory shortleaf pine regeneration. Site index, annual precipitation and overstory basal area are other important variables that affect the regeneration negatively. Regeneration prediction using DT models can be an attractive alternative method for forest managers who prefer faster, purer, and easier data driven solutions to manage their shortleaf pine forests. Incorporation of inputs such as seed production rates, edaphic properties of study sites, hardwood regeneration data, and ice damage records in future applications can make the DT model even more accurate for predicting the shortleaf pine regeneration levels. In addition the DT approach can also be applied in similar ecological studies where data contain nominal or ordinal target variables.

References

Agresti, A., 2007. Building and applying logistic regression models; Categorical Data Analysis, Second Edition: 211-266. doi:10.1002/0471249688.ch6

Aquino, L.D., Tullis, J.A., Stephen, F.M., 2008. Modeling red oak borer, Enaphalodes rufulus (Haldeman), damage using in situ and ancillary landscape data. Forest Ecol Manag 255(3-4): 931-939. doi:10.1016/j.foreco.2007.10.011

Bisquert, M., Caselles, E., Sánchez, J.M., Caselles, V., 2012. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. Int J Wildland Fire. 21(8): 1025-1029. doi:10.1071/WF11105

Budhathoki, C.B., 2006. Mixed-effects modeling of shortleaf pine (*Pinus echinata* Mill.) growth data, Ph. D. dissertation, Oklahoma State University, Stillwater, OK.

Budhathoki, C.B., Lynch, T.B., Guldin, J.M., 2008. A Mixed-Effects Model for the dbh–Height Relationship of Shortleaf Pine (*Pinus echinata* Mill.). Southern J. Appl. Forestry 32(1): 5-11.

Budhathoki, C.B., Lynch, T.B., Guldin, J.M., 2008. Nonlinear mixed modeling of basal area growth for shortleaf pine. Forest Ecol Manag 255(8): 3440-3446. doi:10.1016/j.foreco.2008.02.035

Budhathoki, C.B., Lynch, T.B., Guldin, J.M., 2010. Development of a shortleaf pine individual-tree growth equation using non-linear mixed modeling techniques. In: Stanturf, J.A., eds. Proceedings of the 14th biennial southern silvicultural research conference. Gen. Tech. Rep. SRS-121. Asheville, NC: USDA Forest Service, Southern Research Station: 519-520.

Bui, E.N., Henderson, B.L., Viergever, K., 2006. Knowledge discovery from models of soil properties developed through data mining. Ecol Model 191(3): 431-446. doi:10.1016/j.ecolmodel.2005.05.021

Dhar, S., 2011. Identity verification of a person by classifying digitized handwritten signature. Department of computer science and engineering (MS thesis). Jadavpur University, India.

Kabrick, J.M., Dey, D.C., Shifley, S.R., Villwock, J.L., 2010. Early survival and growth of planted shortleaf pine seedlings as a function of initial size and overstory stocking. P. 277-286 *in* 17th central hardwood forest conference. U.S. Department of Agriculture, Forest Service, Northern Research Station, Lexington, KY.

KC, A., Lynch, T.B., Guldin, G.M., 2016. Shortleaf pine (*Pinus echinata* Mill.) and hardwood regeneration after thinning natural shortleaf pine forests in southern United States. In: Schweitzer, C.J., Clatterbuck, W.K., Oswalt, C.M., eds. Proceedings of the 18th biennial southern silvicultural research conference. e–Gen. Tech. Rep. SRS–212. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 614

KC, A., Lynch, T.B., Stevenson, D., Wilson, D., Guldin, J.M., Heinermann, B., Holeman, R., Wilson, D., Anderson, K., 2015. Development of understory tree vegetation after thinning naturally occurring shortleaf pine forests. In: Gordon, H.A., Connor, K.F., Haywood, J.D., eds. Proceedings of the 17th biennial southern silvicultural research conference. e-Gen. Tech. Rep. SRS-203. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 99.

Kim, K., Yoo, K., Ki, D., Son, I.S., Oh, K.J., Park, J., 2011. Decision-Tree-based data mining and rule induction for predicting and mapping soil bacterial diversity. Environ Monit Assess 178(1-4): 595-610.

Kim, K.S., Park, J.H., 2009. A survey of applications of artificial intelligence algorithms in eco-environmental modelling. Environmental Engineering Research, 14(2): 102-110. doi:10.4491/eer.2009.14.2.102

Lawson, E.R., 1986. National regeneration of shortleaf pine. In: Murphy, P.A. (Ed.), Proceedings of a Symposium on the Shortleaf Pine Ecosystem. Univ. Arkansas Coop. Ext. Serv., Monicello, AR, 53-63.

Lees, B.G., Ritman, K., 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environ Manage 15(6): 823-831.

Lilly, C.J., Will, R.E., Tauer, C.G., Guldin, J.M., Spetich, M.A., 2012. Factors affecting the sprouting of shortleaf pine rootstock following prescribed fire. Forest Ecol Manag 265: 13-19. doi:10.1016/j.foreco.2011.10.020

Lynch, T.B., Nkouka, J., Huebschmann, M.M., Guldin, J.M., 2003. Maximum likelihood estimation for predicting the probability of obtaining variable shortleaf pine regeneration densities. Forest science 49(4): 577-584.

McWilliams, W.H., Sheffield, R.M., Hansen, M., Birch, T.W., 1986. The shortleaf resource. P. 9-24 *in* Symposium on the shortleaf pine ecosystem. Monticello, AR; Cooperative extension service, Little Rock, AR.

Moret, S.L., Langford, W.T., Margineantu, D.D., 2006. Learning to predict channel stability using biogeomorphic features. Ecol Model 191(1): 47-57. doi:10.1016/j.ecolmodel.2005.08.011

Moser, W.K., Hansen, M., McWilliams, W.H., Sheffield, R.M., 2006. Shortleaf pine composition and structure in the United States. P. 19-27 *in* Shortleaf pine restoration and ecology in the Ozarks. U.S. Department of Agriculture, Forest Service, Northern Research Station, Springfield, MO.

Perry, R.W., Thill, R.E., Leslie, D.M., 2008. Scale-dependent effects of landscape structure and composition on diurnal roost selection by forest bats.J. Wildl. Manag 72(4): 913-925. doi:10.2193/2006-435

Piramuthu, S., 2004. Evaluating feature selection methods for learning in data mining applications. Eur J Oper Res 156(2): 483-494. doi:10.1016/S0377-2217(02)00911-6

Roberts, D.W., 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66(3): 123-131.

Sarma, K.S., 2013. Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications. SAS Institute.

Sesnie, S.E., Gessler, P.E., Finegan, B., Thessler, S., 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. Remote Sens. Environ. 112(5): 2145-2159. doi:10.1016/j.rse.2007.08.025

Stevenson, D.J., Lynch, T.B., Saud, P., Heineman, R., Holeman, R., Wilson, D., Anderson, K., Cerny, C., Guldin, J.M., 2016. Winter in the Ouchitas- A severe winter storm signature in Pinus echinata in the Ouachita Mountains of Oklahoma and Arkansas, USA. Dendrochronologia 37: 107-115. doi:10.1016/j.dendro.2016.01.001

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240(4857): 1285-1293. doi:10.1126/science.3287615

Tan, P.N., Steinbach, M., Kumar, V., 2006. Introduction to data mining. Link: http://www-users.cs.umn.edu/~kumar/dmbook/index.php (cited on September 23, 2015).

Tirelli, T., Pessani, D., 2011. Importance of feature selection in decision-tree and artificial-neural-network ecological applications. Alburnus alburnus alborella: A practical example. Ecol Inform 6(5): 309-315. doi:10.1016/j.ecoinf.2010.11.001

Valipour, M., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. Meteorol. Appl. 22: 592–598. doi:10.1002/met.1491.

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez Dam reservoir. J. Hydrol. 476: 433–441. doi: 10.1002/met.1491

Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H., 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. J Veg Sci 11(5): 679-694. doi: 10.2307/3236575

Yang, C.C., Prasher, S.O., Enright, P., Madramootoo, C., Burgess, M., Goel, P.K., Callum, I., 2003. Application of decision tree technology for image classification using remote sensing data. Agr Syst 76(3): 1101-1117. doi:10.1016/S0308-521X(02)00051-3

Zhang, B., Valentine, I., Kemp, P.D., 2005. A decision tree approach modelling functional group abundance in a pasture ecosystem. Agric. Ecosyst. Environ. 110(3): 279-288. doi:10.1016/j.agee.2005.04.017

Zhang, D., Huebschmann, M.M., Lynch, T.B., Guldin, J.M., 2012. Growth projection and valuation of restoration of the shortleaf pine–bluestem grass ecosystem. Forest Policy Econ. 20: 10-15. doi:10.1016/j.forpol.2012.01.014

Table III-1 General description of inputs showing assigned variable names, minimum and maximum values.

S.N.	Variable description (Unit)	Variable name	Values [Min, max]
1	Shortleaf pine regeneration (density ha^{-1})	SLP	[0, 13344]
2	Overstory basal area (m^2 ha^{-1})	BA	[3.97, 48.68]
3	Average age of the sample plot (years)	AGE	[33, 119]
4	Site Index (m)	SI	[12.33, 26.64]
5	Site	SITE	Ozark/Ouachita
6	Years after thinning	YEAR	1996/2001/2006/2010
7	Thinning levels	THINNING	A/B/C/D
8	Altitude of the plot (m)	ALTITUDE	[177, 481]
9	Average annual precipitation (mm)	PRECIPITATION	[987, 1491]
10	Slope of the plot	SLOPE	[89.90, 90.00]
11	Aspect of the plot	ASPECT	[0, 354]

a. Residual basal area (Thinning levels) are: A= (BA<10.332 m^2 ha^{-1}); B= (10.332 m^2 ha^{-1}≤ BA<17.22 m^2 ha^{-1}); C= (17.22 m^2 ha^{-1} ≤ BA<24.108 m^2 ha^{-1}); and D= (BA≥ 24.108 m^2 ha^{-1}).

Note: Overstory basal area (m^2 ha^{-1}) has not been used together with Thinning levels.

Table III-2 Input variables of interval and ratio scale with imputed values and other descriptive summary statistics.

S.N.	Input	Imputed value	Mean(SD)	Skewness	Kurtosis
1	ALTITUDE	291.52	291.51(71.28)	0.63	-0.26
2	PRECIPITATION	1265.1	1265.18(106.35)	-0.20	0.02
3	SLOPE	89.990	89.98(0.01)	-2.94	-24.56
4	EASTNESS	0.060	0.06(0.73)	-0.12	-1.56
5	NORTHNESS	0.070	0.07(0.67)	-0.10	-1.42
6	SLOPEN[b]	NA[a]	0.66 (0.24)	-0.97	0.36
7	SI	NA	18.98(3.20)	0.14	-0.67
8	AGE	NA	76.28(20.64)	-0.09	-1.05
9	BA	NA	21.21(9.21)	0.19	-0.79

a. No imputation was performed.

b. $SLOPEN = [\max(\text{SLOPE} - 89.90, 0) / 0.094]^4$.

Table III-3 Fit statistics, model significance and significant variables for three logistic regression (LR) models.

Model	Validation MR	Validation ASE	*P*-value[a]	Significant variables
LR1	0.1052	0.0915	<0.001	PRECIPITATION, AGE, SI
LR2	0.0861	0.0871	<0.001	PRECIPITATION, AGE, SI, THINNING
LR3	0.1052	0.0876	<0.001	PRECIPITATION, THINNING, SI^2, $EASTNESS^2$, $SLOPE^2$, SLOPE*AGE

a. *P*-value from the Likelihood ratio test for the model significance.

MR= Misclassification Rate

ASE= Average Square Error

Table III-4 Summary statistics of the best performing logistic regression model (LR2)

Variable	Coefficient (SE)	Odds ratio	*P*-value
INTERCEPT	15.807 (3.554)	-	<0.001
PRECIPITATION	-0.011 (0.002)	0.989	<0.001
AGE	0.024 (0.009)	1.024	0.0130
SI	-0.383 (0.085)	0.682	<0.001
THINNING CLASS A vs D	2.164 (0.410)	8.703	<0.001
THINNING CLASS B vs D	0.421 (0.420)	1.523	0.3168
THINNING CLASS C vs D	-1.587 (0.802)	0.205	0.0477

Table III-5 Fit statistics of the three Decision Tree (DT) models.

DT model	Validation MR	Validation ASE	Input variables
DT1	0.076	0.069	SI, PRECIPITATION, BA
DT2	0.076	0.069	SI, PRECIPITATION, BA
Interactive (Gini)	0.090	0.077	PRECIPITATION, SI, BA, AGE

Table III-6 Prediction results from the application of the default Decision Tree (DT1) model on validation data.

Actual Target	Predicted Target	Result	Count	Percentage
High regeneration	High regeneration	TP	5	2.392
Low regeneration	High regeneration	FP	2	0.956
High regeneration	Low regeneration	FN	14	6.698
Low regeneration	Low regeneration	TN	188	89.95

Table III-7 Decision rules derived from the default Decision Tree (DT) model (DT1).

Node ID	Condition	Decision
3	If BA ≥ 18.77	Low regeneration
5	If BA <18.77 and SI ≥19.50	Low regeneration
7	If BA <18.77, SI <19.50 and PRECIPITATION ≥1271.09	Low regeneration
8	If BA <18.77, PRECIPATION <1271.09 and SI <15.80	High regeneration

Table III-8 Percentage distribution of response variable (SLPN) in four thinning levels

Thinning Class	SLPN (Low)	SLPN (High)	Total (%)
A	20.83	4.45	25.29
B	23.28	2.16	25.43
C	25.29	1.01	26.29
D	22.41	0.57	22.99
Total	91.81	8.19	100

Note: Low and High represent “Low Regeneration” and “High Regeneration” respectively.

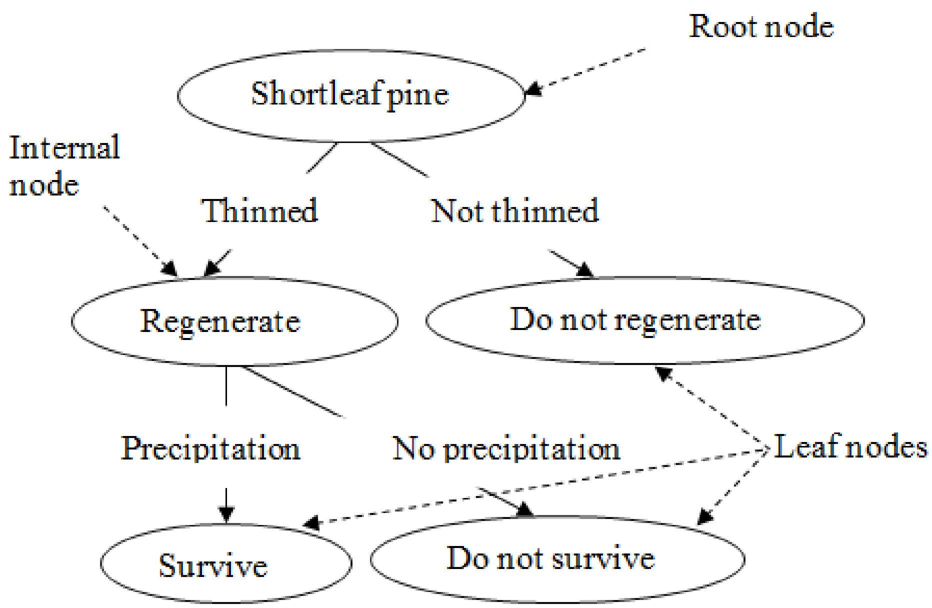

Fig. III-1 Schematic illustration of a hypothetical Decision Tree. Two hypothetical inputs (Thinning and Precipitation) are used to split the Decision Tree. Any node that is not further splitting is leaf node.

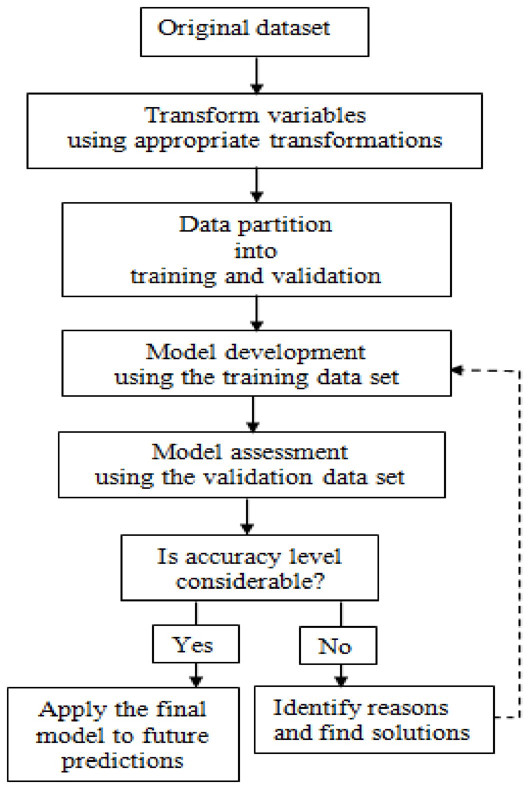

Fig. III-2 Schematic presentation of the predictive model building process.

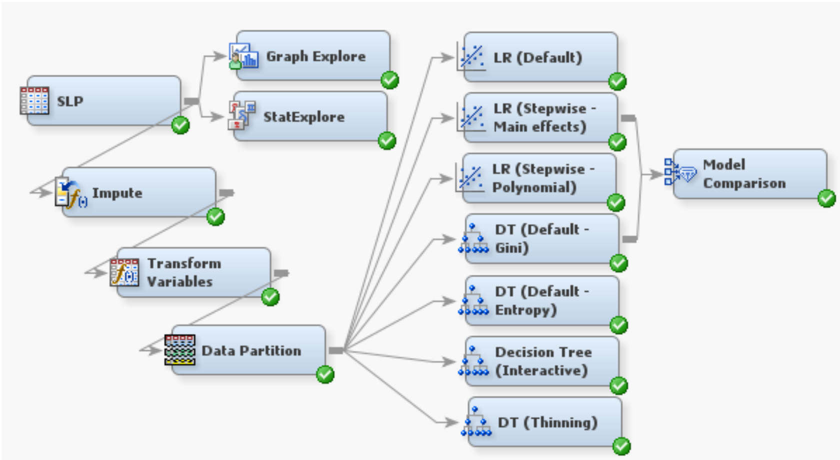

Fig. III-3 Flow diagram for model development and comparison on SAS Enterprise Miner 12.3. SLP is the name of a dataset that we used to build all the models. Arrows make the connections between nodes, and the data analysis process moves a step forward. Graph Explore and StatExplore were used to conduct the descriptive analysis of the data before building the Decision Tree and Logistic Regression models.

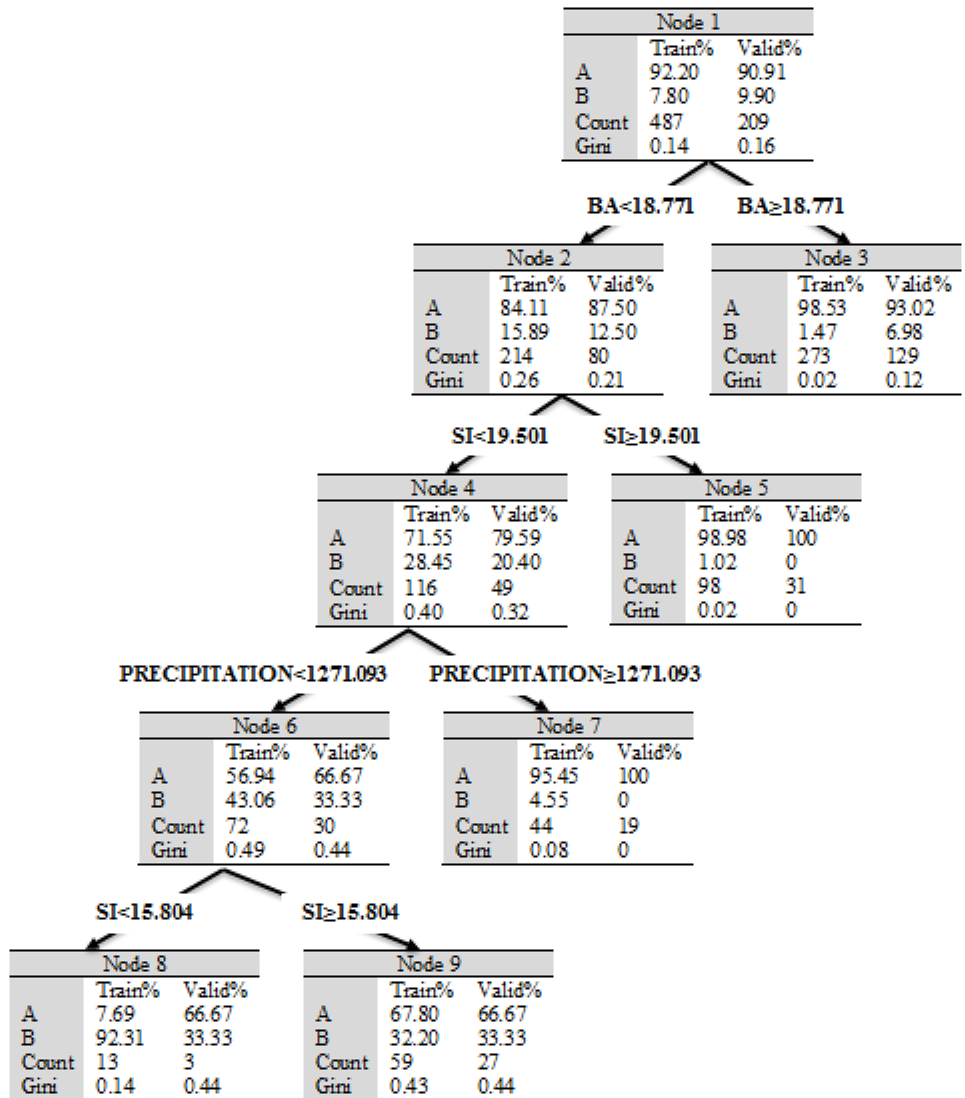

Fig. III-4 Default Decision Tree (DT1) model to predict the shortleaf pine regeneration. Class A and B are number of plots with low (<1730 stems ha^{-1}) and high (≥1730 stems ha^{-1}) regeneration respectively. Results from training data (Train %) on DT1 has been validated using validation data (Valid%).

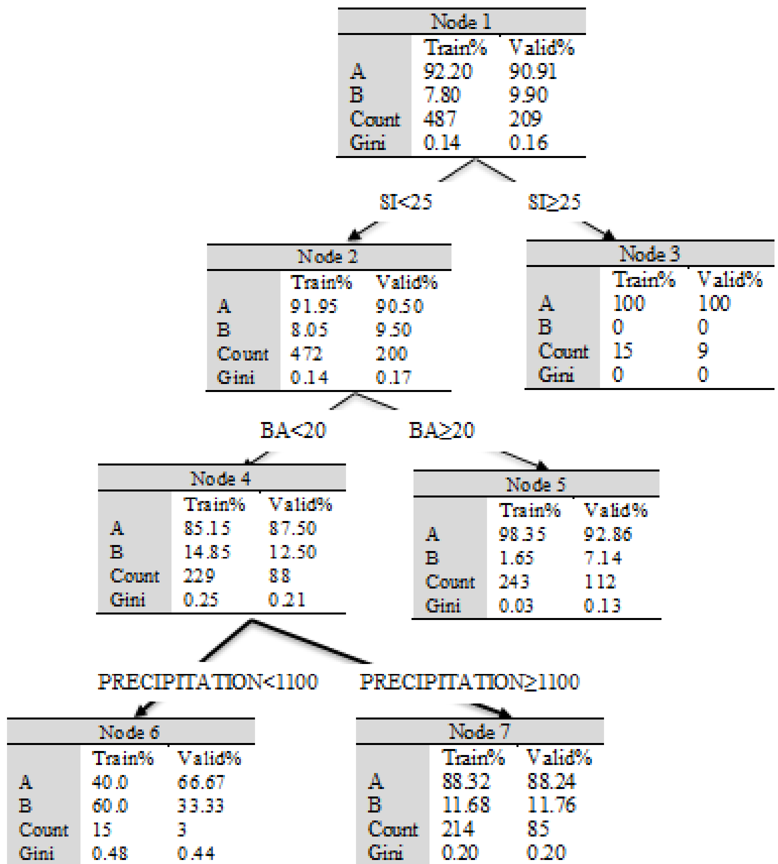

Fig. III-5 An example of an interactive DT model where the model split three times using three input variables (Site index, Overstory basal area and Precipitation). Because it is an interactive DT model, the values of input variables were selected by the authors.

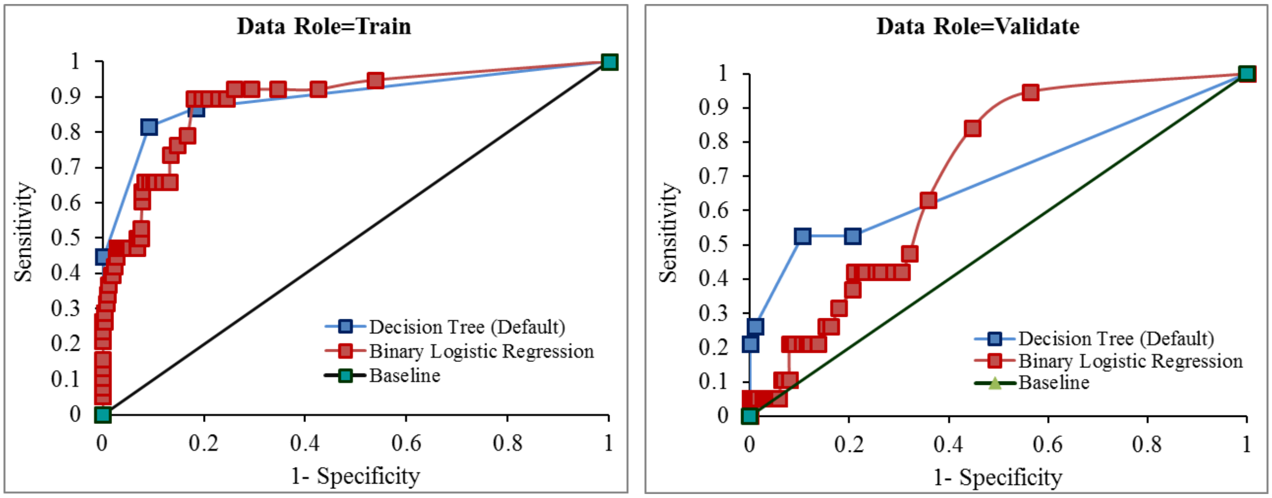

Fig. III-6 ROC (Receiver Operating Characteristics) curves comparing the performances of default Decision Tree (DT1) model with the best performing Logistic Regression (LR2) model. Sensitivity and the specificity are described in methods section.

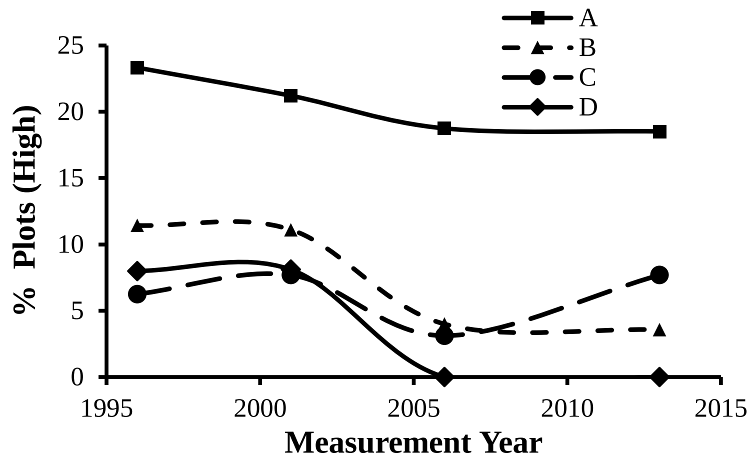

Fig. III-7 Percent of plots with high shortleaf pine regeneration (B>1730 stems ha^{-1}) over four measurement periods at four thinning levels (A<10 m^2 ha^{-1}, 10 m^2 ha^{-1} ≤B< 17 m^2 ha^{-1}, 17 m^2 ha^{-1} ≤C< 24 m^2 ha^{-1}, and D≥ 24 m^2 ha^{-1}). These four thinning levels were created using the shortleaf pine residual basal area. Results were extracted using the interactive Decision Tree (DT2) model.

CHAPTER IV

Predicting shortleaf pine regeneration after thinning in Arkansas and Oklahoma USA: A comparison of logistic regression, artificial neural network, and support vector machine methods

Abstract

Shortleaf pine (*Pinus echinata* Mill.) forests have been one of the most common forest types in the southeastern United States. But in recent years, the standing volume of the shortleaf pine is declining in the region. This study aimed to develop, evaluate, and compare the performance of logistic regression, artificial neural network, and support vector machine models to predict shortleaf pine regeneration in Arkansas and Oklahoma, USA. The predictors were multiple overstory shortleaf pine characteristics, climatic and topographic information, and the target variable was the understory shortleaf pine density. The best performing logistic regression model showed precipitation, plot age, site index, and overstory thinning were the significant inputs affecting understory shortleaf pine density with validation misclassification rate of 8 percent. The best performing artificial neural network model predicted the shortleaf pine density with validation misclassification rate of 7.6 percent, and cumulative lift of 5, 2.5 and 1.66 at depth of 20, 40 and 60 respectively. Similarly, the best performing support vector machine model predicted the shortleaf pine density with validation misclassification rate of 9 percent, and cumulative lift of 3.79, 2.10 and 1.39 at depth of 20, 40 and 60 respectively. An artificial neural network model performed best to predict the shortleaf pine density in Arkansas and Oklahoma.

The authors presume the results from this study can be extrapolated to the other naturally occurring shortleaf pine-oak mixed forests in southeastern United States.

Keywords:

Shortleaf pine, regeneration, logistic regression, artificial neural network, support vector machine, cumulative depth

1. Introduction

Historically, shortleaf pine (*Pinus echinata* Mill.) forests were one of the most common forest types in the southeastern United States (McWilliams et al., 1986; Kabrick et al., 2010), and shortleaf pine is considered one of the most important tree species in Arkansas and eastern Oklahoma, USA (Zhang et al., 2012). Shortleaf pine has been desirable in the region in terms of timber production for southern pine lumber which is typically used in building and home construction. It is also a source of southern pine pulpwood for the pulp and paper industry. Shortleaf pine is particularly desirable for red cockaded woodpecker habitat (Zhang et al., 2012), and it is also important from the tourism and recreation perspectives (Lawson and Kitchens, 1983). Despite its importance, shortleaf pine populations have been declining in recent years (Moser et al., 2006; KC et al., 2015; KC et al., 2016). KC et al. (2015) suggested that the current rate of regeneration of shortleaf pine seedlings is not adequate to maintain the shortleaf pine dominated forests in the long-term in Ozark and Ouachita Mountains of Arkansas and Oklahoma. In many naturally occurring shortleaf pine-hardwood mixed forests, hardwood regeneration dominates shortleaf pine saplings. Long term low shortleaf pine regeneration coupled with the continual hardwood domination might substantially affect the sustainability of shortleaf pine forests in this region. Therefore, multi-aged understory shortleaf pine seedlings, saplings and trees are desired to offer better ecosystem restoration. This also helps forests to transition to uneven-aged forests from an even-aged condition. Therefore, continuous and consistent shortleaf pine regeneration is often desired in the shortleaf pine forests of Arkansas and Oklahoma, USA.

In many ecological studies, data are complex and nonlinear (Lek et al., 1996; Gevrey et al., 2003; Ozesmi et al., 2006). Multiple studies conducted in the past suggested that predictive modeling/ machine learning techniques are strong and effective tools for assessment of such complex nonlinear patterns from ecological data (Almeida, 2002; Ozesmi et al., 2006). Artificial intelligence (AI) algorithms, such as logistic regression (LR), classification and regression trees (CART), artificial neural network (ANN), support vector machines (SVM), random forests, and ensemble models have been widely used in recent years because of their flexibility, speed, and accuracy (Aquino et al., 2008). However, only few of these techniques have been used in forest management (e.g. Jensen et al., 1999; Bisquert et al., 2012).

Initially, ANN models were perceived as black box models (Gevrey et al., 2003), and many ecologists were hesitant to apply these techniques. However, these models have been widely applied in recent years to answer the variety of ecology related questions. Frequently ANN models outperform the linear models (Ozesmi et al., 2006), because they detect non-linear patterns better than the linear and LR models. For example, ANN models have been applied in studies such as water quality (Awad, 2014), fisheries (Huse and Giske, 1998; Gebler et al., 2014), modeling microbial community structures (Santos et al., 2014), among others. ANN models can extract the nonlinear patterns that exist in large and complex data sets (Noble et al., 2000; Mele and Crowley, 2008; Santos et al., 2014), and do not need *a priori* hypotheses to guide model development. Similarly, LR models have been popular for prediction of regeneration for several forest tree species. For example, Larsen et al. (1997) used LR models to predict the probability of occurrence for oak regeneration in the Missouri Ozarks. Lynch et al. (2003) used LR models to predict the shortleaf pine (*Pinus echinata* Mill.) regeneration in Arkansas and Oklahoma, USA.

SVM models are a supervised learning method based on statistical learning theory (Vapnik, 1998). These models have rarely if ever been used in forest management but are popular in many other ecological studies. For example, Acevedo et al. (2009) applied an SVM model to classify the calls of nine frogs and three bird species, and reported that it performed best among all tested models by correctly classifying the calls 94.95 percent of the time. Hu and Davis (2005) applied SVM models to identify the plankton taxa and reported that the method reduced the classification error rate from 39 to 28 percent. SVM models often provide better fit statistics compared to traditional regression models (Gevrey et al., 2003; Aquino et al., 2008).

In order to better understand the present and future status of shortleaf pine forests in Arkansas and Oklahoma and to develop efficient management programs, the development of an efficient statistical/ predictive model is needed to assess the major factors influencing shortleaf pine regeneration. In this study, we developed multiple forms of LR, ANN and SVM models that predicted shortleaf pine regeneration in Ozark and Ouachita national forests in Oklahoma and Arkansas, USA. Additionally, we compared the performance of LR, ANN and SVM models based on their fit statistics to select the best performing model to predict the shortleaf pine regeneration. To the best of our knowledge, this study is first in kind to use and compare predictive modeling techniques to assess shortleaf pine regeneration in the southeastern USA.

2. Materials and Methods

2.1. Study area and data collection

The USDA Forest Service Southern Research Station and the then Department of Forestry, Oklahoma State University jointly established 180 permanent study plots in the Ozark and Ouachita National Forests during the period from 1985 to 1987. Study plots were located in the Ozark National Forest (latitude 35.3º N, longitude 93.1º W) and the Ouachita National Forest (latitude 34.0º N, longitude 94.7º W) in southeastern Oklahoma (Lynch et al., 2003). Out of 180 plots, 133 plots were from the Ouachita National Forest and 47 plots were from the Ozark

National Forest. Overstory measurement plots were circular with 809.371 m^2 in area and a 16.063 m radius. The 10.05 m isolation buffers were created outside the plots and were treated similarly to the plots.

Overstory shortleaf pine characteristics such as diameter at breast height (dbh), tree age, site index were measured when establishing the plots in 1985. Here, overstory represents all the shortleaf pine trees in plots that were remained after thinning in 1985. Shortleaf and hardwood understory trees are the cohort regenerated after thinning and hardwood control in 1985 and which were taller than 1.3m in 1995, hereafter these will be termed "understory". Shortleaf pine overstory characteristics from all 180 plots have been measured at approximately 5 year intervals since they were established in 1985. While establishing the plots, understory hardwoods exceeding 2.54 cm in diameter at ground level were removed using herbicides. Hardwoods were also removed from the isolation buffer area to eliminate hardwood competition. The understory woody-vegetation started regenerating a few years after plot establishment. During the time of third overstory measurement in 1995, four 20.23 m^2 subplots were created inside all of the 809.371 m^2 plots to measure the understory woody vegetation regeneration. All of the woody vegetation and shortleaf pine regeneration available inside the subplot larger than 1.37 m in height were measured. Only two subplots were measured within each overstory plot in 1995 but all four subplots were measured in each overstory plot in subsequent measurements.

2.2. Data management and exploratory analysis

We first assigned the understory shortleaf pine stem density ha^{-1} (SLP) as the target variable. Overstory shortleaf pine basal area, average tree age in plot, site index, sites, year of measurement, thinning classes, altitude and the average annual precipitation were assigned as the predictors. Site, thinning class and year of measurement were the only class predictor variables.

Four thinning classes were assigned based on the overstory basal area from 1985 (Table IV-1). Such basal area classes were created by thinning the overstory shortleaf pine from below in 1985. Most plots were also thinned from below in 2000 to bring the overstory basal area level down to similar level of 1985. The target variable was understory shortleaf pine density. Hereafter, we call it shortleaf pine regeneration density. For the target variable, regeneration density < 1730 stems ha^{-1} were assigned to class **A** and ≥ 1730 stems ha^{-1} were assigned to class **B**, and the new binary variable was denoted as SLPN. These classes were created based on previous studies in which shortleaf pine regeneration density exceeding 1730 stems ha^{-1} was indicative of adequate or high regeneration, whereas, shortleaf pine regeneration density below 1730 stems ha^{-1} were indicative of poor regeneration (Lynch et al., 2003). Climatic and topographic variables were extracted using the GPS locations of the plots from Arkansas and Oklahoma. All analyses were performed using SAS Enterprise Miner software version 12.3 (SAS Institute Inc. USA; hereafter SAS EM).

Additionally, we checked the symmetry assumption using measures of skewness and kurtosis. The symmetry for the interval inputs was not violated (Table IV-2). We conducted the Pearson's product-moment correlation analysis to detect the correlation between the inputs, as well as their correlation with the target variable. Any predictor variable that had a correlation greater than 0.70 (Dormann et al., 2013) and the variance inflation factor greater than 10 (O'brien, 2007) were excluded from the model building process.

We randomly separated the original data into training and validation datasets, and assigned 70% (487 observations) of the data into the training and 30% (209 observations) to validation data set as described by Tan et al. (2006) and Sarma (2013). Later, the training data set was used to develop LR, ANN and SVM models and the validation data set was used to evaluate the performance of models built using the training data set.

2.3. Logistic Regression models

We used logit as the link function (Eq. 1) to predict shortleaf pine regeneration as high or low. The link function can be algebraically reformulated as an event probability function (Eq. 2). We fitted models using the maximum likelihood method (McCullagh and Nelder, 1989) and used the stepwise selection method to select the best performing model. Various forms of LR models were developed using polynomial and interaction effects. We selected a simple yet good performing LR model to predict the shortleaf pine regeneration. Fit statistics such as Akaike's information criterion (AIC), Schwarz's Bayesian criterion (SBC), average square error (ASE), mean square error (MSE) and the misclassification rate (MR) were used to select the best performing model among various LR models. The validation dataset was used to control the overfitting of models. After selecting the best performing LR model, we tabulated parameter estimates and odds ratios obtained from the selected LR model and interpreted them accordingly.

$$\log\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 X_1 + ... + \beta_P X_P \qquad (1)$$

$$p = \frac{1}{1+\exp\left[-(\beta_0 + \beta_1 X_1 + ... + \beta_P X_P)\right]} \qquad (2)$$

2.4. Artificial Neural Network Models

The level of regeneration of shortleaf pine was predicted using multi-layered feed forwarding neural network model (Fig. IV-1). ANN uses complex nonlinear transformations and provides the probability of target variables using mathematical functions (Sarma, 2013). This powerful nonlinear regression technique (Bishop, 1995; Ripley, 1996) was inspired by theories about how the brain works (Kuhn and Johnson, 2013; Baesens, 2014). Baran et al. (1996) and Lek et al. (1996) have described ANN models in detail. The SAS EM software offers a node for ANN where variable transformation, filtering, composite variable creation and the model estimation are done simultaneously in such a way that a specified error function is minimized (Sarma, 2013). Technically, ANN is a sequence of input and output layers. There could be several hidden layers between the input and final output layers. Here, we used three hidden units inside the hidden layers. Every output layer was treated as the input layer at the next level to create another output layer until we obtained the final output layer.

The target and output layers perform two operations: combination and activation. Units use the target layer formula to combine the inputs, and they are then called target layer combination functions, and formulas used for transforming the combined values are called target layer activation functions. The combination and the activation functions in both the hidden layers and in the target layer are key elements of the architecture of an ANN (Sarma, 2013). The final result of the neural network largely depends on the selection of the hidden layer combination functions, and SAS Enterprise Miner software has a wide range of choices for those functions. We used multilayer perception, generalized linear model, user, ordinary radial-equal width, ordinary radial-unequal width, and normalized radial-equal width architectures to construct the network. Sarma (2013) described these functions in detail. We assessed the fit statistics such as AIC, SBC, ASE, MSE, and MR and selected the best performing model among various ANN models.

The cumulative lift chart is used to determine the predictive capability of the ANN models. Sometime, cumulative lift charts are also referred to as the gain chart. In the chart, x-axis represents the percentile or depth and the y-axis represents the lift. Indeed, the x-axis contains the cumulative number of cases with decreasing probability. Cumulative depths are the percentile of the data after applying the model. Cumulative lift is an approach that selects the lowest possible samples and achieves the most impactful results. Therefore, the decision makers can use the lift chart to take the better management decisions.

2.5. Support Vector Machine

SVM is a supervised machine-learning method that can be used to perform regression and classification analysis (Base SAS 9.4 Procedures Guide, 2015). In many problems, finite dimensional space is not linearly separable and the original space needs to be mapped into a higher dimensional space (Kampichler et al., 2010). This makes the separation easier (Base SAS 9.4 Procedures Guide, 2015). SVMs use sigmoidal nonlinear kernel (Gunn, 1998; Williams, 2011; Were et al., 2015), polynomial, and radial basis kernel functions to project the data onto a new hyperspace where complex non-linear patterns can be represented in a simpler fashion. It aims to construct an optimal hyperplane in the new hyperspace that separates classes and creates the widest margin between their data (Were et al., 2015). SVM model is a binary classifier (Kampichler et al., 2010; Nathan et al., 2011). It has been used in past to assess the behavior of domestic animals such as cats (Watanabe et al., 2005) and cows (Martiskainen, et al., 2009). The data structure used in this analysis fits the SVM model assumptions quite well. Therefore, it is worthy to assess how SVM model performs on correctly classifying the two classes (A and B) of understory shortleaf pine densities compared to LR and ANN models. As described earlier in section 2.5, the cumulative lift chart can also be used to determine the predictive capability of the SVM models.

2.6. Model comparison

Two layers of model comparison were created. Initially, we selected the best performing models in each group of LR, ANN and SVM. Then, we compared the performance of best performing models among each other. Receiver operating characteristics (ROC) curves and validation MR were used to compare selected models. The ROC curve is a graphical technique that describes and compares the accuracy of models by plotting the 1-specificity in X-axis and sensitivity on Y-axis (Akobeng, 2007). The area under the ROC curve represents the overall performance of the model (Akobeng, 2007).

3. Results

3.1. Logistic Regression models

We developed four LR models and evaluated their performances based on fit statistics (Table IV-3). The LR2 model had the lowest validation MR and consisted of the following significant inputs: PRECIPITATION, AGE, SI and THINNING (Table IV-3). Based on the fit statistics, the LR2 model performed the best of the LR models (Table IV-3). The parameter estimates and the odds ratio of significant variables are presented in table IV-4. The effect of nominal variable YEAR is not statistically significant ($P>0.05$). However, it is important to assess how time factor after thinning the overstory affects the regeneration. Therefore, we also present the parameter estimates and odds ratios of YEAR (Table IV-4).

3.2. Artificial Neural Network Models

We developed six ANN models by applying the different architectures for each model (Table IV-5). The ANN models only used the inputs that were significant in the LR2 models. Thus, PRECIPITATION, AGE, SI, and THINNING were used to develop the ANN models. All fit statistics suggested that the ANN3 model performed the best among all of the ANN models (Table IV-5). This model also had the lowest validation MSE and the highest AUC. The ANN3 model had the cumulative lift of 5, 2.5 and 1.66 percent on the depth of 20, 40 and 60 respectively.

3.3. Support vector Machines

We developed four SVM models by applying the different functions for each model (Table IV-6). Based on the fit statistics obtained from the validation data, the SVM2 model that used the Kernel polynomial function performed best among all the models (Table IV-6). The SVM2 model had a cumulative lift of 3.79, 2.10 and 1.39 percent on the depth of 20, 40 and 60 respectively.

3.4. Comparison of Model Performances

Here, we compared the performances of the selected LR (LR2), ANN (ANN3), and SVM (SVM: K-Polynomial) models. The detailed fit statistics and receiver operating characteristics (ROC) curves show that the ANN3 model was the best at predicting shortleaf pine regeneration (Fig. IV- 3), having the lowest training and validation MRs (Table IV-5). In terms of validation MR, the ANN3 model, which was developed by using the training data to predict the shortleaf pine regeneration, outperformed all other models. Other fit statistics (Tables IV-3, IV-5 and IV-6) also suggested that ANN3 performed better than other models for predicting shortleaf pine regeneration. Also, ANN3 provided the lowest number of false negatives (Table IV-7) compared to other models for both training and validation data. Hence, it was selected as the best performing model.

4. Discussion

Various forms of LR models have been used in past for shortleaf pine regeneration prediction (Nkouka, 1999; Lynch et. al., 2003), risk assessment (Jalkanen and Mattila, 2000), vegetation distribution prediction (Hilbert and Ostendorf, 2001), habitat evaluation (Pearce and Ferrier, 2000) and so on. In the present study, among four significant variables (SI, AGE, PRECIPITATION, and THINNING) from LR2 model, THINNING levels show the strongest effect on shortleaf pine regeneration (Table IV-4). The odds of high regeneration density (≥1730 stems ha^{-1}) gets low consistently in thinning levels B, C and D. Odds of getting high regeneration density reduces by 72.3%, 77.9% and 91.5% in thinning levels B, C, and D respectively compared to thinning level A. This result reflects the importance of overstory basal area level and practice of thinning on shortleaf pine regeneration. Lynch et al (2003) and Nkouka (1999) stated that overstory basal area affects the shortleaf pine regeneration negatively. Here, the overstory basal area information has not been used directly on LR models; however, as mentioned earlier in methods section, the four thinning levels represent the residual overstory shortleaf pine basal area after thinning in 1985. The shortleaf pine is shade intolerant (Baker et al., 1996). When a shade intolerant species, like shortleaf pine, fails to pose the intense competition and rapid height growth to remain in top canopy, they lag behind and succumb to hardwood competition (Baker et al., 1996). This study shows highly thinned plots have higher chance of having high shortleaf pine regeneration density. Thinning is a positive driving factor to promote the shortleaf pine regeneration and restore shortleaf pine forests for a long term.

The negative effect of site index on shortleaf pine regeneration density (Table IV-4) is not a surprising result. The finding is concordant with multiple studies conducted in past (e.g. Lawson 1986; Nkouka et al., 1999; Lynch et al., 2003). As the site index increases by one unit, the odds of high regeneration density (≥1730 stems ha^{-1}) decreases by 30.9 percent. The poor shortleaf pine regeneration in sites with higher site index is because such sites tend to be even more favorable for hardwood regeneration. Precipitation affects the shortleaf pine regeneration negatively (Table IV-4). With one unit increase in precipitation, the odds of high regeneration density decreases by 1.2 percent. This is not a strong effect; however, precipitation affecting the regeneration negatively is an interesting finding. We assume that hardwoods take more advantage of increased precipitation than t shortleaf pine. One of the reasons that shortleaf pine mostly grows on the sites with poor site quality is that hardwoods regenerate relatively poorly on such sites. As far as we know, this is the first study on shortleaf pine that asses the effect of precipitation on shortleaf pine regeneration. The age of the overstory shortleaf pine trees in plot is the only variable that has positive effect on understory shortleaf pine regeneration density. With one unit increase in age, the odds of high regeneration density increases by 3.8 percent.

Explanatory variables showing importance or significant in LR models can be further assessed by applying the predictive modeling techniques such as ANN and SVM to achieve the better fits (Ozesmi et al., 2006). Using only the significant inputs from LR models in ANN and SVM models is a common practice (Zurada et al., 1994; Gevrey et al., 2003). We applied the similar approaches and only used the significant variables (PRECIPITATION, AGE, SI, THINNING) from selected LR (LR2) model to build the ANN and SVM models. The ANN and SVM models are often considered “black boxes” or “data mining tools” (Intrator and Intrator, 2001), and detailed mathematical explanations of the predictive models are complex. ANN models are non-parametric in nature. While developing the ANN models, data do not require the transformation to match the desired distribution (Ozesmi and Ozesmi, 1999) because it goes through multiple transformations on various layers at model building process.

The cumulative lift charts generated by ANN and SVM models provide information on how much more likely we receive positive response (≥1730 stems ha^{-1}) at certain deciles of data than if we select a random plot (Sayad 2016). Based on cumulative lift for model ANN3, selected 20, 40 and 60 percent of the plots can have 5, 2.5 and 1.66 times of lift on high regeneration (≥1730 stems ha^{-1}) compared to selecting a random plot (Fig. IV- 4). That means, by selecting 20 percent of plots based on predictive model (ANN3 at present case) will provide 5 times more plots with high regeneration, as if we use no model. Similarly, in the case of SVM2 model, selected 20, 40 and 60 percent of the plots can have 3.79, 2.10 and 1.61 times of lift on high regeneration (≥1730 stems ha^{-1}) (Fig. IV-4). The comparison of cumulative lift shows that ANN3 performed better than the SVM2 model in predicting shortleaf pine regeneration for certain depth of data. The lift chart result suggests that the ANN3 model is better than the SVM2 model to predict shortleaf pine regeneration in plots. In this case, it detected higher percentage of plots (e.g. 1.66 times lift on 60 % plots for ANN3 vs 1.61 times lift on 60 % plots for SVM2) for a certain level of lift on high regeneration density.

Using the ANN3 model, forest managers can focus on the sites that are not regenerating at the desired level. KC et al. (under review, 2016) described how shortleaf pine regeneration can be improved in certain sites by applying the decision tree model, but we need to be able to locate the exact sites that are not regenerating well and the cumulative lift data provides such information. The cumulative lift chart (Fig. IV- 4) provides the lift (in y-axis) at certain decile (in x-axis), and the data can be separated easily at any decile level. Indeed, this is already a popular technique in the medical and business research (Shen et al., 2007; Das, 2010). Mostly, the business and medical studies use cumulative lift chart the other way. They target the first few deciles to receive the maximum response. But, we target last few deciles where the regeneration is critically low and plan accordingly so that shortleaf pine regenerate better on those sites too. To the best of our knowledge, this is the first paper stating that cumulative lift chart can be a helpful tool from a forest management perspective. Separating the poorly regenerating sites with satisfactorily regenerating sites using a cumulative lift chart helps to distribute the time and resources to sites where the interventions are needed to achieve certain shortleaf pine regeneration goals. Model ANN3 can be helpful in achieving such goals.

5. Conclusions

The ANN model (User as the architecture) predicted shortleaf pine regeneration with lowest validation misclassification rate. A cumulative lift chart provided an assessment of regeneration performance at various depths of data. The low MR of the ANN3 model on validation dataset further assures that the margin of error is low while drawing conclusions using the results from model ANN3. The selected predictive model (ANN3) can be an additional tool to the forest managers on making long term policy decisions on shortleaf pine forests management. Furthermore, we encourage future researchers to collect extra information such as seed distribution trends, hardwood regeneration, controlled burning, and edaphic factors that can affect shortleaf pine regeneration and reconstruct the ANN3 model. By doing so, we anticipate that the predictive power of the ANN3 will be further improved. This study can be a stepping stone for using predictive models to explore the non-linear patterns of ecological data particularly in the field of forest management in the future in the southeastern United States.

References

Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J., Aide, T.M., 2009. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecol Inform 4(4): 206-214. doi:10.1016/j.ecoinf.2009.06.005

Akobeng, A.K. 2007. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta paediatrica 96(5): 644-647. doi:10.1111/j.1651-2227.2006.00178.x

Almeida, J.S., 2002. Predictive non-linear modeling of complex data by artificial neural networks. Curr Opin Biotechnol 13(1): 72-76. doi:10.1016/S0958-1669(02)00288-4

Aquino, L.D., Tullis, J.A., Stephen, F.M., 2008. Modeling red oak borer, *Enaphalodes rufulus* (Haldeman), damage using *in situ* and ancillary landscape data. Forest Ecol Manag 255(3-4): 931-939. doi:10.1016/j.foreco.2007.10.011

Awad, M., 2014. Sea water chlorophyll-a estimation using hyperspectral images and supervised Artificial Neural Network. Ecol Inform 30(24): 60-68. doi:10.1016/j.ecoinf.2014.07.004

Baesens, B., 2014. Analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

Baker, J.B., Cain, M.D., Guldin, J.M., Murphy, P.A., Shelton, M.G., 1996. Uneven-aged silviculture for the loblolly and shortleaf pine forest cover types. USDA Forest Service, Southern Research Station. General Technical Report SO-118. P. 1.

Baran, P., Lek, S., Delacoste, M., Belaud, A., 1996. Stochastic models that predict trout population density or biomass on a mesohabitat scale. Hydrobiologia 337(1-3): 1-9.

doi:10.1007/BF00028502

Base SAS 9.4 Procedures Guide: Statistical Procedures. 2015. SAS Institute Inc, Cary, NC.

Bishop, C.M. 1995. Neural networks for pattern recognition. Oxford university press.

Bisquert, M., Caselles, E., Sánchez, J.M., Caselles, V., 2012. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. Int J Wildland Fire 21(8): 1025-1029. doi: http://dx.doi.org/10.1071/WF11105

Das, R. 2010. A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Syst Appl 37(2): 1568-1572. doi:10.1016/j.eswa.2009.06.040

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1): 27-46. doi:10.1111/j.1600-0587.2012.07348.x

Gebler, D., Kayzer, D., Szoszkiewicz, K., Budka, A., 2014. Artificial neural network modelling of macrophyte indices based on physico-chemical characteristics of water. Hydrobiologia 737(1): 215-24. doi: 10.1007/s10750-013-1585-7

Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 160(3):249-264. doi:10.1016/S0304-3800(02)00257-0

Gunn, S.R., 1998. Support vector machines for classification and regression. Technical Report, ISIS, Department of Electronics and Computer Science, University of Southampton, Southampton, UK.

Hilbert, D.W., Ostendorf, B., 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. Ecol Model 146(1): 311-327. doi:10.1016/S0304-3800(01)00323-4

Hu, Qiao., Davis, C.S., 2006. Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. Mar Ecol Prog Ser 306: 51-61. doi:10.3354/meps306051

Huse, G., Giske, J., 1998. Ecology in Mare Pentium: an individual-based spatio-temporal model for fish with adapted behaviour. Fish Res 37(1):163-178. doi:10.1016/S0165-7836(98)00134-9

Intrator, O., Intrator, N., 2001. Interpreting neural-network results: a simulation study. Comput Stat Data Anal 37(3): 373-393. doi:10.1016/S0167-9473(01)00016-0

Jalkanen, A., Mattila, U., 2000. Logistic regression models for wind and snow damage in northern Finland based on the National Forest Inventory data. Forest Ecol Manag 135(1): 315-330. doi:10.1016/S0378-1127(00)00289-9

Jensen, J., Qiu, F. Ji, M., 1999. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. International Journal of Remote Sensing 20(14): 2805-2822. doi: http://dx.doi.org/10.1080/014311699211804

Kabrick, J.M., Dey, D.C., Shifley, S.R., Villwock, J.L., 2010. Early survival and growth of planted shortleaf pine seedlings as a function of initial size and overstory stocking. P. 277-286 *in*

17th central hardwood forest conference. U.S. Department of Agriculture, Forest Service, Northern Research Station, Lexington, KY.

Kampichler, C., Wieland, R., Calme, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: A comparison of five machine-learning methods. Ecol Inform 5(6): 441-450. doi:10.1016/j.ecoinf.2010.06.003

KC, A., Lynch, T.B., Guldin, G.M., 2016. Shortleaf pine (*Pinus echinata* Mill.) and hardwood regeneration after thinning natural shortleaf pine forests in southern United States. In: Schweitzer, C.J., Clatterbuck, W.K., Oswalt, C.M., eds. Proceedings of the 18th biennial southern silvicultural research conference. e–Gen. Tech. Rep. SRS–212. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 614

KC, A., Lynch, T.B., Stevenson, D., Wilson, D., Guldin, J.M., Heinermann, B., Holeman, R., Wilson, D., Anderson, K., 2015. Development of understory tree vegetation after thinning naturally occurring shortleaf pine forests. In: Gordon, H.A., Connor, K.F., Haywood, J.D., eds. Proceedings of the 17th biennial southern silvicultural research conference. e-Gen. Tech. Rep. SRS-203. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 99.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer. doi:10.1007/978-1-4614-6849-3

Larsen, D.R., Metzger, M.A., Johnson, P.S., 1997. Oak regeneration and overstory density in the Missouri Ozarks. Can J Forest Res 27(6): 869-875.

Lawson, E.R., 1986. Natural regeneration of shortleaf pine. P. 53–63 *in* Murphy, P.A. (Ed.), Proceedings of a Symposium on the Shortleaf Pine Ecosystem. Univ. Arkansas Coop. Ext. Serv., Monticello, AR.

Lawson, E.R., Kitchens, R.N., 1983. Silvicultural systems for major forest types of the United States. USDA Handbook No. 445, USDA Forest Service, US Government Printing Office, P. 157-160.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996. Role of some environmental variables in trout abundance models using neural networks. Aquat Living Resour 9(1): 23-29. doi: http://dx.doi.org/10.1051/alr:1996004

Lynch, T.B., Nkouka, J., Huebschmann, M.M., Guldin, J.M., 2003. Maximum likelihood estimation for predicting the probability of obtaining variable shortleaf pine regeneration densities. Forest science 49(4): 577-584.

Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M., Mononen, J., 2009. Cow behavior pattern recognition using a three-dimensional accelerometer and support vector machines. Appl. Anim. Behav. Sci. 119(1): 32-38. doi:10.1016/j.applanim.2009.03.005

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Vol. 37. CRC press.

McWilliams, W.H., Sheffield, R.M., Hansen, M., Birch, T.W., 1986. The shortleaf resource. P. 9-24 *in* Symposium on the shortleaf pine ecosystem. Monticello, AR; Cooperative extension service, Little Rock, AR.

Mele, P.M., Crowley, D.E., 2008. Application of self-organizing maps for assessing soil biological quality. Agri. Ecosyst. Environ. 126(3): 139-152. doi:10.1016/j.agee.2007.12.008

Moser, W.K., Hansen, M., McWilliams, W.H., Sheffield, R.M., 2006. Shortleaf pine composition and structure in the United States. P. 19-27 *in* Shortleaf pine restoration and ecology in the Ozarks. U.S. Department of Agriculture, Forest Service, Northern Research Station, Springfield, MO.

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., Getz, W.M., 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. J. Exp. Biol. 215(6): 986-996. doi:10.1242/jeb.058602

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.S. thesis, Oklahoma State University, Stillwater, OK.

Noble, P.A., J.S. Almeida, and C.R. Lovell. 2000. Application of neural computing methods for interpreting phospholipid fatty acid profiles of natural microbial communities. Appl Environ Microbiol 66(2): 694-699. doi:10.1128/AEM.66.2.694-699.2000

O'brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity 41(5): 673-690. doi: 10.1007/s11135-006-9018-6

Ozesmi, S.L., Ozesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecol Model 116(1): 15-31. doi:10.1016/S03043800(98)00149-5

Ozesmi, S.L., Tan, C.O., Ozesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol Model 195(1-2): 83-93. doi:10.1016/j.ecolmodel.2005.11.012

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133(3):225-245. doi:10.1016/S0304-3800(00)00322-7

Ripley, B.D., 1996. Pattern recognition and neural networks. Cambridge university press.

Santos, E.C., Armas, E.D., Crowley, D., Lambais, M.R., 2014. Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil. Soil Biol Biochem 69:101-109. doi:10.1016/j.soilbio.2013.10.049

Sarma, K.S., 2013. Predictive modeling with SAS Enterprise Miner: Practical solutions for business applications. SAS Institute.

Sayad, S., 2016. Model evaluation-classification. An introduction to data mining. Assessed on 04/23/2016. url: *http://www.saedsayad.com/data_mining_map.htm*Shen, A., Tong, R., Deng, Y., 2007. Application of classification models on credit card fraud detection. P. 1-4 *in* Service Systems and Service Management. International Conference on IEEE, Chengdu, China.

Tan, C.O., Ozesmi, U., Beklioglu, M., Per, E., Kurt, B., 2006. Predictive models in ecology: Comparison of performances and assessment of applicability. Ecol Inform 1(2): 195-211. doi:10.1016/j.ecoinf.2006.03.002

Vapnik, V., 1998. Statistical Learning Theory. Wiley.

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: a case study with the domestic cat. Appl. Anim. Behav. Sci 94(1): 117-131. doi:10.1016/j.applanim.2005.01.010

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol Indic 52: 394-403. doi:10.1016/j.ecolind.2014.12.028

Williams, G., 2011. Data mining with Rattle and R: the art of excavating data for knowledge discovery. Springer Science & Business Media.

Zhang, D., Huebschmann, M.M., Lynch, T.B., Guldin, J.M., 2012. Growth projection and valuation of restoration of the shortleaf pine–bluestem grass ecosystem. Forest Policy Econ 20: 10-15. doi:10.1016/j.forpol.2012.01.014

Zurada, J.M., Malinowski, A., Cloete, I. 1994. Sensitivity analysis for minimization of input data dimension for feedforward neural network. ISCAS'94. IEEE International Symposium on circuits and Systems, vol. 6, IIE Press, London. P. 447-450.

Table IV-1 Description of target and explanatory variables (inputs) and their assigned variable names, scale, maximum, and minimum values

Variable description (Unit)	Variable	Values [min, max]
Shortleaf pine regeneration (density ha^{-1})	SLP	[0,13344]
Overstory basal area (m^2 ha^{-1})	BA	[3.97,48.68]
Age of the sample plot (years)	AGE	[33,119]
Site Index (m)	SI	[12.33,26.64]
Site	SITE	Ozark/Ouachita
Year of measurement	YEAR	1996/2001/2006/2010
Thinning class[a] (m^2 ha^{-1})	THINNING	A/B/C/D
Altitude of plot location(m)	ALTITUDE	[177,481]
Annual precipitation (mm)	PRECIPITATION	[987,1491]

a. Thinning classes are: A= (Overstory shortleaf pine basal area<10 m^2 ha^{-1}),
B= (10 m^2 ha^{-1}≤ Overstory shortleaf pine basal area <17 m^2 ha^{-1}),
C= (17 m^2 ha^{-1}≤ Overstory shortleaf pine basal area <24 m^2 ha^{-1}), and
D= (Overstory shortleaf pine basal area≥ 24 m^2 ha^{-1}).

Table IV-2 Descriptive summary of interval inputs

Variable	Mean [SD]	Skewness	Kurtosis
ALTITUDE	291.51 [71.28]	0.63	-0.26
PRECIPITATION	1265.18 [106.35]	-0.20	0.02
SI	18.98 [3.20]	0.140	-0.67
AGE	76.28 [20.64]	-0.09	-1.05
BA	21.21 [9.21]	0.19	-0.79

Table IV-3 Logistic regression models and their fit statistics

Model	Validation MR	*P*-value	Significant Variables
LR1	0.11	<0.001	PRECIPITATION, AGE, SI
LR2	0.08	<0.001	PRECIPITATION, AGE, SI, THINNING
LR3	0.09	<0.001	PRECIPITATION, BA, AGE, SI, THINNING
LR4	0.11	<0.001	PRECIPITATION, THINNING, SI*SI

LR1 uses none as the model selection criteria.

LR2 doesn't use polynomial and interaction terms in the model and stepwise is the selection criteria.

LR3 uses interaction terms in the model and stepwise is the selection criteria.

LR4 uses polynomial and interaction terms in the model and stepwise is the selection criteria.

Table IV-4 Parameter estimates and odds ratio of inputs from LR2 model

Parameter	Estimate	Standard Error	*P*-value	Odds Ratio
INTERCEPT	15.158	2.78	<0.001	
PRECIPITATION	-0.012	0.001	<0.001	0.988
AGE	0.037	0.009	<0.001	1.038
SI	-0.370	0.009	<0.001	0.691
THINNING (B vs A)	0.029	0.295	0.920	0.277
THINNING (C vs A)	-0.196	0.359	0.584	0.221
THINNING (D vs A)	-1.147	0.430	<0.001	0.085
YEAR (2001vs 1996)	0.146	0.290	0.614	0.525
YEAR (2006 vs 1996)	-0.540	0.323	0.094	0.264
YEAR (2013 vs 1996)	-0.395	0.315	0.200	0.306

Table IV-5 Performance of ANN models on predicting shortleaf pine regeneration

Model	Architecture	NPE	Validation			
			MR	MSE	AUC	NWC
ANN1	MP	31	0.086	0.078	0.696	18
ANN2	GLM	31	0.090	0.089	0.698	20
ANN3	USER	29	0.086	0.076	0.753	18
ANN4	OR-EW	31	0.090	0.076	0.668	19
ANN5	OR-UW	30	0.095	0.082	0.683	20
ANN6	NR-EW	31	0.105	0.083	0.670	22

MP= Multilayer perception

GLM=Generalized linear model

OR-EW=Ordinary radial-equal width

OR-UW=Ordinary radial-equal width

NR-EW=Normalized radial- equal width

NPE= Number of parameter estimates in the model

AUC= Area under receiver operating curve (ROC)

NWC= Number of wrong classification

Table IV-6 Performance of SVM models on predicting shortleaf pine regeneration

Model	Function	Validation		
		MR	AUC	NWC
SVM1	Kernel Linear	0.09	0.695	19
SVM2	Kernel Polynomial	0.09	0.720	19
SVM3	Kernel RBF	0.09	0.683	19
SVM4	Kernel sigmoidal	0.13	0.52	29

Kernel RBF= Kernel Radial basis function

AUC= Area under ROC curve for validation data set

NWC= Number of wrong classification on the validation dataset

Table IV-7 Event classification for selected LR, ANN, and SVM models for validation data.

Model	Data Role	FN	TP	FP	TP	Total Observation
LR2	VALIDATE	18	190	0	1	209
SVM:K-Polynomial	VALIDATE	18	189	1	1	209
ANN3	VALIDATE	9	183	7	10	209

TP= True Positive [Classifying High regeneration (≥1730 stems ha^{-1}) as High regeneration]

FP= False Positive [Classifying High regeneration as Low regeneration (<1730 stems ha^{-1})]

TN= True Negative [Classifying Low regeneration as Low regeneration]

FN= False Negative [Classifying Low regeneration as High regeneration]

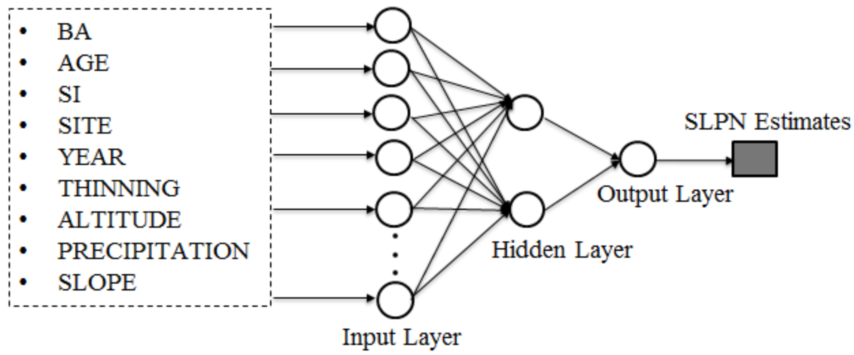

Fig. IV-1 Architecture of the MLP neural network for SLPN estimates.

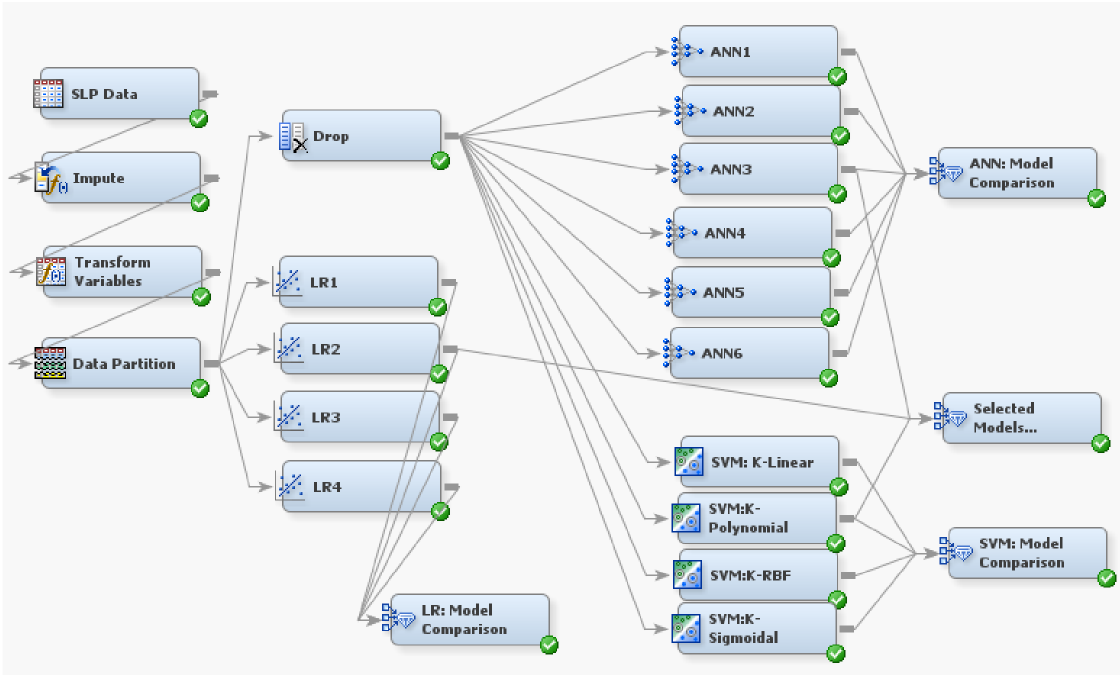

Fig. IV-2 Process flow map of the model building processes, in SAS EM software version 12.3 interface, and their comparison among the group of similar models and the selected models. "SLP Data" is a dataset that is cleaned and ready to be analyzed. Then, data has been imputed for the missing values and transformed as required. "Data Partition" node randomly separated data into two sets as training and validation data sets. "Drop" node dropped all the variables from the dataset that are not used by ANN and SVM models. "Model Comparison" node compared the performances of the models.

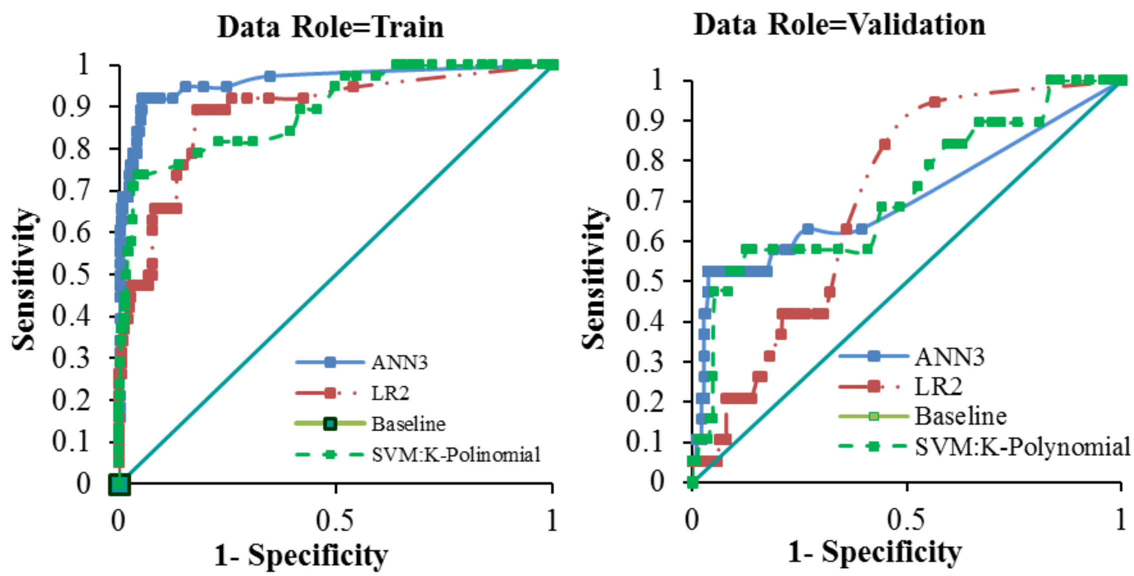

Fig. IV-3 Receiver operating characteristics (ROC) curves to compare the performances of selected LR (LR2), ANN (ANN3), and SVM (SVM2) models for training and validation data.

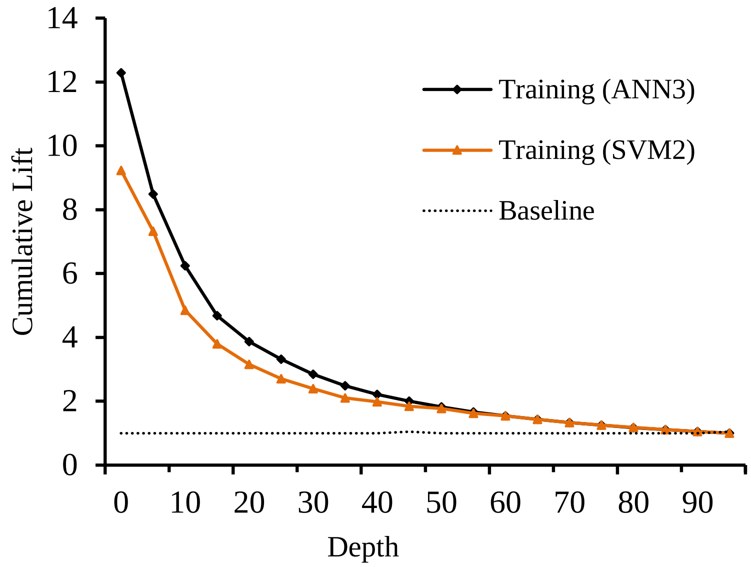

Fig. IV-4 Cumulative lift charts for ANN3 and SVM2 models with baseline.

CHAPTER V

CONCLUSIONS

The low levels of understory shortleaf pine density in this study indicate that on many ownerships where active management activities such as controlled burning are not being practiced, it will be difficult to replace an existing shortleaf overstory. This raises a serious concern for the future of shortleaf pine dominated-oak mixed forests in Arkansas and Oklahoma. Single tree selection thinning on uneven-aged shortleaf pine stands certainly improves the status of understory shortleaf pine density. Treatments to control competing hardwoods are essential at around 10-15 year intervals to maintain and develop shortleaf regeneration that is obtained from a previous single tree selection thinning. We propose 10-15 years interval because the understory shortleaf pine density level in present study was satisfactory until 1996. And, it sharply declined in 2001. This study concludes that not conducting any silvicultural treatment on the stands for around 15 years or longer nullifies the benefits that we received from the first thinning in terms of regeneration. This study suggests the continual intervention is mandatory to achieve healthy shortleaf pine regeneration naturally. In the present study, understory woody-plants dynamics changed significantly in later years. Understory shortleaf pine rarely survives to move to larger dbh classes if the silvicultural treatments are not applied frequently. The short term understory density may provide misleading results as we report good shortleaf pine regeneration in 1996 for number of plots. The status of understory shortleaf pine in 2013 is in a critically poor condition. Treatments like thinning from below, controlled burning, selective understory hardwood clearance could be the possibilities; where economically feasible. We conclude thinning from below every 10-15 years to keep the overstory basal area below17 m^2ha^{-1} would provide sufficient understory shortleaf pine in long-term if competing hardwood vegetation can be controlled.

The decision tree model is an attractive predictive modeling tool for prediction of shortleaf pine regeneration. These models demonstrated that overstory shortleaf pine thinning positively affects understory shortleaf pine regeneration. Site index, annual precipitation and overstory basal area are other important variables that affect the regeneration negatively. Regeneration prediction using DT models is an attractive alternative method for forest managers who prefer faster, purer, and easier data driven solutions to manage their shortleaf pine forests. The ANN model also predicted shortleaf pine regeneration with low validation misclassification rate. A cumulative lift chart provided an assessment of regeneration performance at various depths of data. The low MR of the ANN3 model on validation dataset further assures that the margin of error is low while drawing conclusions using the results from model ANN3. The selected predictive model (ANN3) can be an additional tool to the forest managers for making long term management decisions for shortleaf pine forests.

Incorporation of inputs such as seed production rates, edaphic factors, hardwood regeneration data, controlled burning, and ice damage records in future applications can make the DT and ANN models even more accurate for predicting the shortleaf pine regeneration levels. By adding extra information, we anticipate that the predictive power of the DT and ANN will be further improved. This study can be a stepping stone for using predictive models to explore the non-linear patterns of ecological data particularly in the field of forest management in the future in the southeastern United States. In addition the predictive modeling approach can also be applied in similar ecological studies.

VITA

Anup KC

Candidate for the Degree of

Doctor of Philosophy

Thesis: PREDICTING SHORTLEAF PINE (*Pinus echinata* Mill.) REGENERATION AFTER THINNING IN SOUTHEASTERN UNITED STATES

Major Field: Natural Resource Ecology and Management

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Natural Resource Ecology and Management at Oklahoma State University, Stillwater, Oklahoma in July, 2016.

Completed the requirements for the Graduate Certificate in Data Mining and Predictive Modeling at Oklahoma State University, Stillwater, Oklahoma in July, 2016.

Completed the requirements for the Master of Science in Aquaculture at University of Arkansas Pine Bluff, Pine Bluff, Arkansas in 2011.

Completed the requirements for the Bachelor of Science in Forestry at Institute of Forestry, Pokhara, Nepal in 2007.

Experience:

Research Associate (2012-2016), Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, Oklahoma.

Research Assistant (2009-2011), Department of Aquaculture and Fisheries, University of Arkansas Pine Bluff, Pine Bluff, Arkansas.