A COMPUTER PROGRAM TO PREDICT THE QUALITY OF LONGLEAF PINE SEED CROPS

Daniel J. Leduc and Shi-Jean S. Sung

Abstract—Longleaf pine (*Pinus palustris* Mill.) has good seed years at irregular intervals. Although previous researchers found significant relationships between weather variables and size of the cone crop for a given year, they have stopped short of developing a predictive model. In this study, seed crops were classified as bumper, good to fair, and poor to failed. A canonical discriminant analysis based on weather data was performed to develop a classification function. We then developed a computer program that implemented the results of this canonical discriminant analysis to predict the class of cone crop for a given year. This prediction can be made as early as 18 months prior to seed maturity. This model should greatly help in planning site preparation or seed harvesting activities.

INTRODUCTION

Longleaf pine (*Pinus palustris* Mill.) has long been known to produce irregular cone crops. As far back as 1922, a Forest Service, U.S. Department of Agriculture report stated that longleaf pine bears seed in good quantities only once every 5 to 7 years (Mattoon 1922). This irregularity complicates management efforts in natural regeneration and even potential seed harvests for artificial regeneration.

Given this variation, it would be useful to predict what the quality of the cone crop might be for a given year. Many authors have noticed a correlation between weather and longleaf pine cone crops (Chen and others 2016, Leduc and others 2016, Pederson 1999, Shoulders 1967). Leduc and others (2016) found many weak but sometimes significant correlations between the current year cone crop and weather variables for the preceding three years. Furthermore, they used the results of a canonical discriminant analysis to show that even the nonsignificant variables contributed to an observable difference in classification between bumper, fair-to-good, and poor-to-failed seed crops. We sought to make it easy to use the many weather variables to generate actual concrete predictions for longleaf pine cone crop quality.

METHODS

Data

Longleaf pine cones have been counted during the spring of each year since 1958 at the Escambia Experimental Forest in Alabama, and this count was expanded to nine other locations across the South in subsequent years. This dataset ranges from Louisiana to North Carolina (Leduc and others 2016) and is maintained by the Southern Research Station at Auburn, AL. We obtained monthly values for average temperature, high temperature, low temperature, cooling-degree days, heating-degree days, precipitation, and Palmer drought severity index (PDSI) for each location and year since 1958 (NOAA 2014). Using these variables together resulted in 390 observations from which a model was developed.

Model Development

We wanted to find an empirical predictive model that would allow us to utilize many input variables to classify cone crops into bumper, good-to-fair, and poor-to-failed classes (see table 1 for class definitions). Several methodologies were tested. Among the modeling methods tested were genetic algorithms (System Dynamics International 1997) and neural networks (NeuralWare 1991). However, these methods proved unsatisfactory. We then used more traditional statistics

Table 1—Definitions of longleaf pine cone crop quality used in this study

Crop quality	Cones per tree
Bumper crop	≥100
Fair-to-good crop	25 to 99
Poor-to-failed crop	<25

Author information: Daniel J. Leduc, Information Technology Specialist, and Shi-Jean S. Sung, Research Plant Physiologist, U.S. Department of Agriculture, Forest Service, Southern Research Station, Pineville, LA 71360.

Kirschman, Julia E., comp. 2018. Proceedings of the 19th biennial southern silvicultural research conference. e-Gen. Tech. Rep. SRS-234. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 444 p.

including canonical discriminant analysis (SAS 2004). The initial model was built to predict cone crop quality with data up to March of the crop (that is, seed maturity) year; subsequent models were built to use data only up to December of the previous year (that is, the pollination year), September of the pollination year, and June of the pollination year. The intention was that the best model would use the latest data available, but models with less data could be used when a longer planning horizon for natural regeneration and/or cone harvest was desired at the cost of some loss of accuracy.

RESULTS

The Model

Two canonical vectors were required to classify the seed crop into the three crop quality classes defined in table 1. We checked model quality by predicting the crops for all of the data used to create the model. While this is not ideal, it allowed us to use all of our limited data to develop the model. The percent of successful predictions is shown in table 2. Figure 1 shows the canonical scores for each of the observations in the dataset when the full 27 months of data is used in the model. The three classes are distinctly different for the most part, but there is some overlap that results in uncertainty for the predictions.

In order to use the results of the canonical discriminant analysis, the vector of canonical scores is multiplied by the vector of standardized weather data to get a canonical score of the data. Each of the weather variables is standardized by subtracting its mean value and dividing by its standard deviation. The result of this multiplication is a model canonical score. We have two vectors of canonical scores, so we calculated two model canonical scores and these can be visualized as coordinates on an x-y graph as shown in figure 2. The next step is to see how far the calculated score is from the mean values for each of the cone crop classes. This distance is calculated as a Euclidian distance (ED) as shown in equation 1:

$$ED = \sqrt{X^{2+}Y^{2}} \qquad (1)$$

where X and Y are the respective distances in the horizontal and vertical directions from the calculated score to the mean score of a given class.

For example, figure 2 shows the crop would be considered bumper since the calculated ED is closest to the bumper mean, but one would accept this conclusion tentatively since the score is also close to the mean for the poor-to-failed class.

The Computer Program

Description—The calculations to predict a cone crop class are not difficult, but they are numerous. In the 27-month model (June of crop year), 185 variables are standardized by subtracting their means and dividing by their standard deviations, multiplied by the vectors for the canonical scores one and two, and the results summed. The calculations are not complex but are sufficiently tedious to make manual calculation unlikely. For this reason, a program was written in Visual BASIC® (2012 Microsoft Corporation) to perform the calculations. This program is called LongCones, and it is available on the Southern Research Station Web site (https://srs.fs.usda.gov/longleaf/tools/) as a tool of the Restoring and Managing Longleaf Pine Ecosystems unit.

Requirements—This program was written on Windows 7® and should also run on the more recent versions of Windows®. The program was written for a screen resolution of 1920x1200, but the windows can be scaled. An internet connection is needed to update the weather data.

Table 2—Successful crop quality predictions using various models

Model	Overall	Bumper	Good to fair	Poor to failed
		correct prediction percentage		
March of seed year (27-month)	90	95	82	92
December of pollination year (24-month)	87	95	78	90
September of pollination year (21-month)	85	90	79	87
June of pollination year (18-month)	82	95	76	85

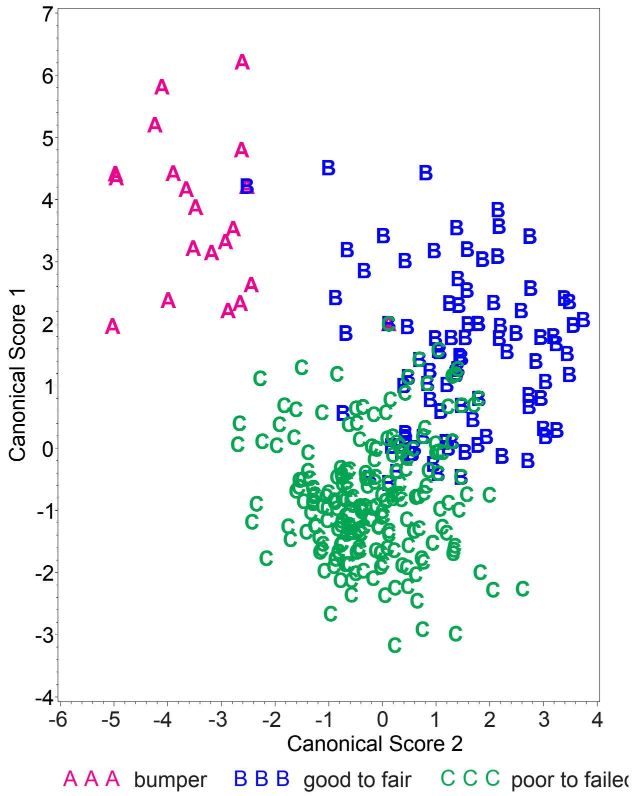


Figure 1—Results of canonical discriminant analysis showing that the calculated scores for each of the three classes of cone crops are mostly distinct from each other.

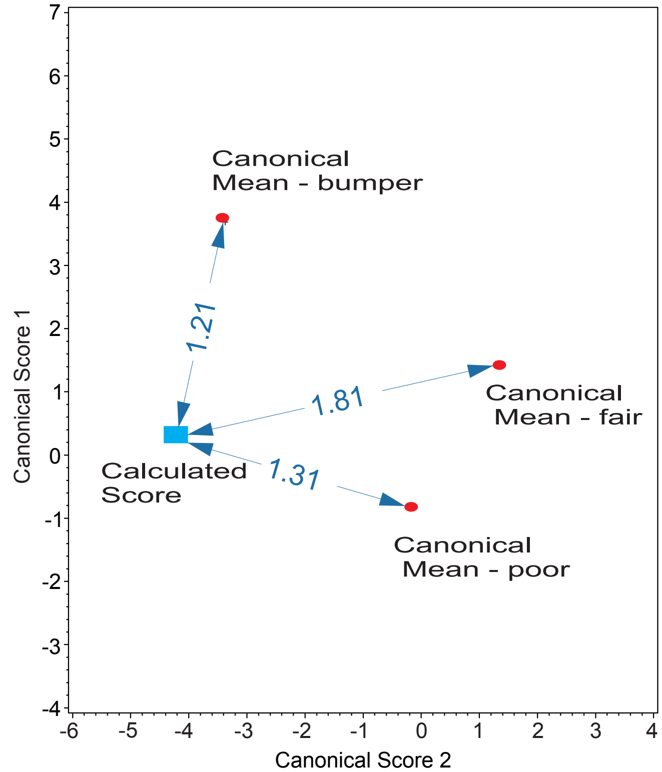


Figure 2—Graphical representation of how a new cone crop class is determined by the model. A new canonical score is calculated by multiplying canonical vectors one and two by the standardized input variables. The two numbers obtained are used as coordinates, and the distance from these coordinates to each of the class means is determined. The closest class is determined to be the new predicted class.

User's guide—Once the program has been installed using the longcones-setup.exe program downloaded from the Web site, the user can begin to use it by double-clicking on the icon. This will bring up a brief introductory splash screen followed by a screen that looks like figure 3. The largest part of this screen is occupied by a map of the areas where this model might be applicable. However, only the climate divisions highlighted in green actually contain stands used in developing this model. (The map shown in figure 3 is simply a reference.) The user must select the State, climate division, and year of interest using the buttons and text boxes below the map. When done, the user simply clicks on the "Get Data" button to continue.

This program uses data files from the NOAA website (https://www1.ncdc.noaa.gov/pub/data/cirs/climdiv/) copied to the user's computer at C:\ProgramData\LongCones. NOAA updates these files about once per month, and the user can update the local files by clicking on the button "Update Weather Data."

Once the user has clicked on "Get Data," the screen will go blank for a few seconds while the appropriate weather data is loaded. This resulting screen (fig. 4) shows the user all of the available weather data for the climate division and year for which they are trying to predict a cone crop. The data is shown for reference and can be changed by the user. Users can use more specific local data or do a sensitivity analysis to determine the effects of individual variables. Only available data will be shown, and this can affect the predictions that can be made. One oddity is that the months of July and August do not show heating degree days (HDD); this is because, in all of the historical data used, these values were always zero. Since this constant value of zero is problematic with the canonical discriminant function, HDD values for July and August were not used in the model. When the user is satisfied with the data shown, a single click on the "Calculate" button will do the necessary calculations to show model results. The user can also click "Go Back" to select different climate data or "Quit" to end the program.

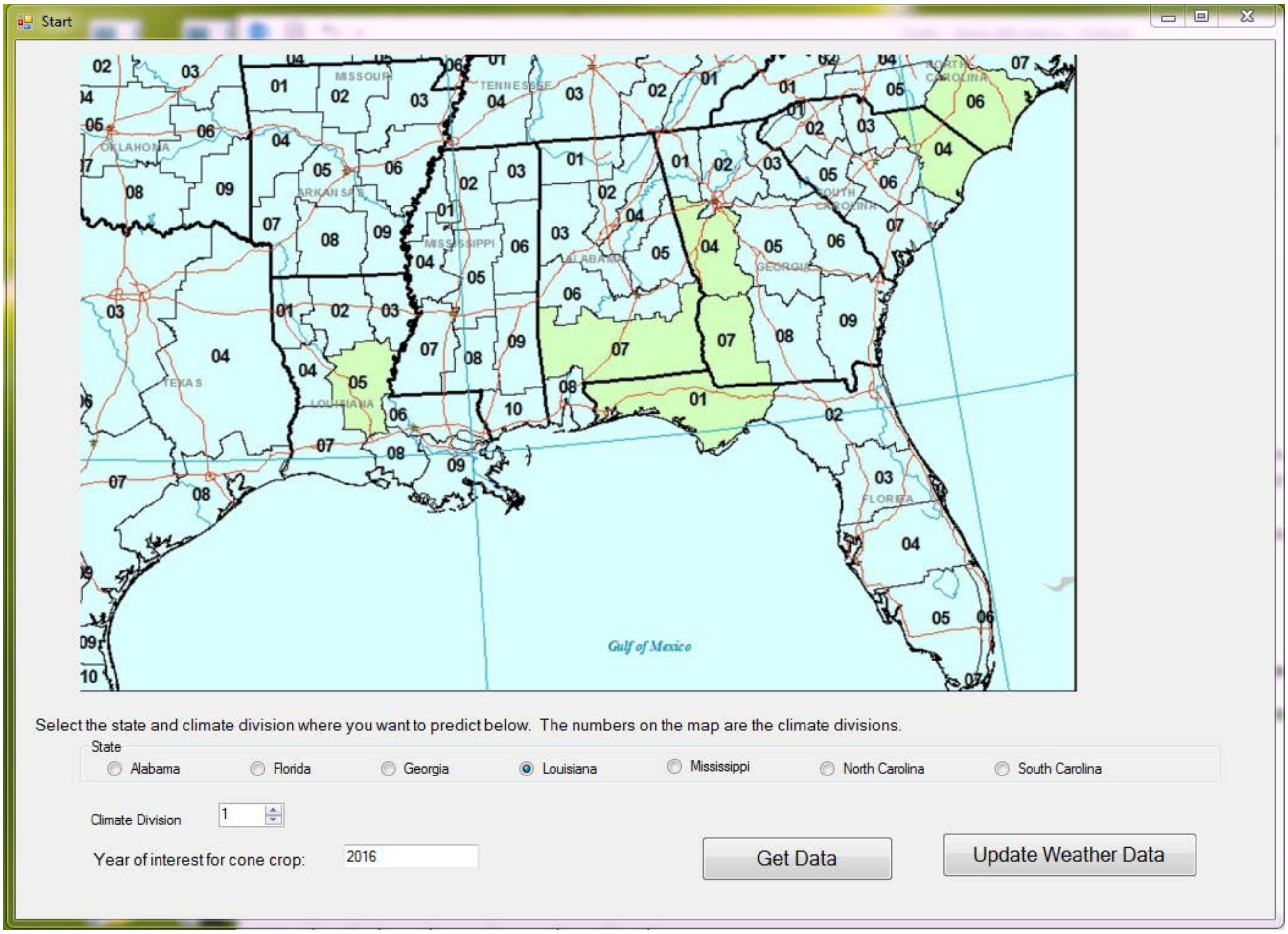
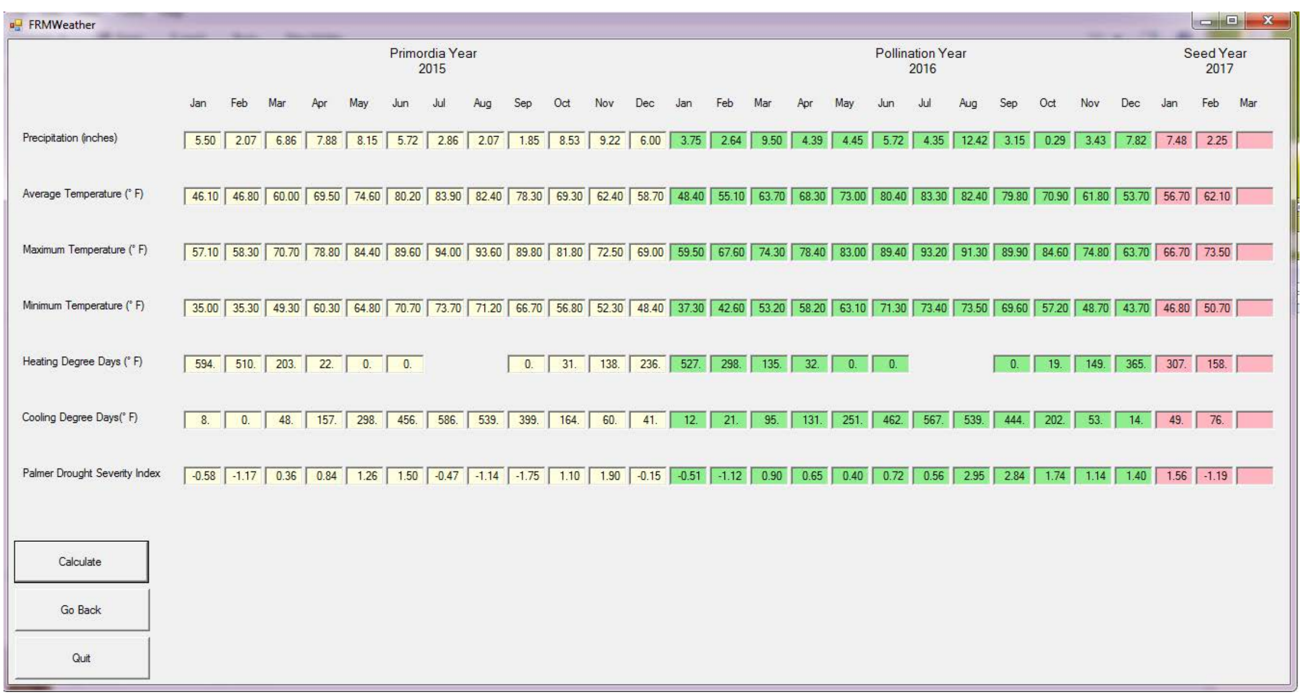


Figure 3—Opening screen from the program LongCones. This is the screen where the user will pick the location of interest and the prediction year.



	Primordia Year 2015												Pollination Year 2016												Seed Year 2017		
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Precipitation (inches)	5.50	2.07	6.86	7.88	8.15	5.72	2.86	2.07	1.85	8.53	9.22	6.00	3.75	2.64	9.50	4.39	4.45	5.72	4.35	12.42	3.15	0.29	3.43	7.82	7.48	2.25	
Average Temperature (° F)	46.10	46.80	60.00	69.50	74.60	80.20	83.90	82.40	78.30	69.30	62.40	58.70	48.40	55.10	63.70	68.30	73.00	80.40	83.30	82.40	79.80	70.90	61.80	53.70	56.70	62.10	
Maximum Temperature (° F)	57.10	58.30	70.70	78.80	84.40	89.60	94.00	93.60	89.80	81.80	72.50	69.00	59.50	67.60	74.30	78.40	83.00	89.40	93.20	91.30	89.90	84.60	74.80	63.70	66.70	73.50	
Minimum Temperature (° F)	35.00	35.30	49.30	60.30	64.80	70.70	73.70	71.20	66.70	56.80	52.30	48.40	37.30	42.60	53.20	58.20	63.10	71.30	73.40	73.50	69.60	57.20	48.70	43.70	46.80	50.70	
Heating Degree Days (° F)	594.	510.	203.	22.	0.	0.			0.	31.	138.	236.	527.	298.	135.	32.	0.	0.			0.	19.	149.	365.	307.	158.	
Cooling Degree Days(° F)	8.	0.	48.	157.	298.	456.	586.	539.	399.	164.	60.	41.	12.	21.	95.	131.	251.	462.	567.	539.	444.	202.	53.	14.	49.	76.	
Palmer Drought Severity Index	-0.58	-1.17	0.36	0.84	1.26	1.50	-0.47	-1.14	-1.75	1.10	1.90	-0.15	-0.51	-1.12	0.90	0.65	0.40	0.72	0.56	2.95	2.84	1.74	1.14	1.40	1.56	-1.19	

Figure 4—Weather data screen from the program LongCones. This is a presentation of the climate data for the location and timeframe from which the user will make model predictions. The data can be edited or the "Calculate" button can be clicked to make predictions.

Figure 5 shows the final results of the program. There are four models, namely: (1) a model that only uses data up through June of the pollination year, (2) a model that uses data through September of the pollination year, (3) a model that uses data up through December of the pollination year, and (4) a model that uses data up through March of the seed year. In figure 5, only three results are shown since the data for March of 2017 was not yet available. Results are only presented for those models that have sufficient data for the calculation.

In addition to estimated crop quality class, the weather data screen (fig. 5) also presents distances from the mean of all classes. These can be used to judge how confident the user can be in the results. In the example, using data only through June of the pollination year resulted in the prediction of a bumper crop, but the distances to the means indicated that while the bumper class mean was closest, the poor-to-failed class mean was also very close. With the additional data added in the next two models, through September and through December of the pollination year, the distance from the bumper mean increases while the distance to the poor-to-failed mean decreases. Another check on the model is to see how many of the four models agree. In validation, having four models in agreement increased the reliability by 31 percent over having three models in agreement.

DISCUSSION

We envision that this model will be used primarily to predict quality of cone crops, which will enhance longleaf pine regeneration planning. However, as with all models, caution must be taken since there is an element of error. At its best in March of the seed year, the model is correct 90 percent of the time, but in June of the previous year it is only correct 82 percent of the time.

Aside from the practical predictions that can be made, another use for this model is testing effects of variables on cone crops. In predicting the 2017 cone crop for climate division 5 in Louisiana, the June of the pollination year model predicted a bumper crop, but subsequent models predicted a poor-to-failed crop. The 1984 crop for the same location was actually a bumper crop, and all of the models predicted that it would be a bumper crop. Using the 1984 data as a guide and a little trial and error, it was discovered that simply changing the average temperature for November of the pollination year from 61.8 to 57.7 °F made the December model predict a bumper crop. The September model still predicted a poor-to-failed crop unless the rainfall of August of the pollination year was changed from 12.42 to 5.56 inches. Changing both variables resulted in all of the available models predicting a bumper crop. Unfortunately, the user of this model might have difficulty in reproducing the above result. NOAA updates climate data values monthly, and these updates do more than just add to the data collection. They also adjust values for the most recent two calendar years (NOAA 2014). The data files current at this writing do not produce the phenomenon described above.

Figure 5—Weather data screen from the program LongCones after the “Calculate” button has been clicked. Results for the models with sufficient input data are shown.

The implications of the sensitivity of the model to small changes is something that needs to be investigated further. By changing the August rainfall from 12.42 to 5.56 inches, a change from an extreme value to a moderate value, the change in distance to a bumper crop changed from 3.4 to 1.3. However, by changing November temperature from 61.8 to 57.7 °F, a moderate change, the ED to the bumper crop mean changed from 6.0 to 47.7 and the ED to the poor-to-failed crop changed from 1.1 to 52.6. Thus, bumper class is the closest result, but at that distance it might not be totally reliable. The threshold for the distance rendering the model meaningless is another area for future investigation.

Finally, the model might be criticized for over parameterization since the full model for March of the seed year (27-month) has 370 parameters and is based on 390 observations. Previous work (Leduc and others 2016) has shown that all of these variables are necessary to get class separation, as shown in figure 1. Thus, the model may not be as applicable in a general sense as is desired. However, the models that terminate in earlier months have fewer parameters and the same number of observations. The June of the pollination year model (18-month) has only 248 parameters, which gives some freedom for generalization. In applying this model, one should consider the variation in models depending on endpoint to help judge reliability.

CONCLUSION

A model was developed that predicts longleaf pine crops and could be a useful tool in planning regeneration strategies. The computer program makes the model easy to implement. However, the model is only the starting point for investigating prediction reliability and potentially finding strategies to increase future cone crops.

ACKNOWLEDGMENTS

We thank the many people who have contributed longleaf pine cone counts since 1958 and especially Dale G. Brockway who made this data available to us.

LITERATURE CITED

Chen, X.; Guo, Q.; Brockway, D.G. 2016. Analyzing the complexity of cone production in longleaf pine by multiscale entropy. Journal of Sustainable Forestry. 25(2):172-182.

Leduc, D.J.; Sung, S.-J.S.; Brockway, D.G.; Sayer, M.A.S. 2016. Weather effects on the success of longleaf pine cone crops. 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: 535-541.

Mattoon, W.R. 1922. Longleaf pine. Bull. 1061. Washington, DC: U.S. Department of Agriculture. 49 p.

Neuralware Inc. 1991. Neural computing. Pittsburgh, PA: NeuralWare Inc. 355 p.

NOAA. 2014. Divisional-readme.txt. https://www1.ncdc.noaa.gov/pub/data/cirs/climdiv/ [Date accessed: May 8, 2017].

NOAA. 2017. Index of /pub/data/cirs/climdiv. https://www1.ncdc.noaa.gov/pub/data/cirs/climdiv/ [Date accessed: May 8, 2017].

Pederson, N.; Kush, J.S.; Meldahl, R.S. 1999. Correlating climate and longleaf pine cone crops: is there a connection? In: Kush, J.S., comp. Rep. No. 4. Longleaf pine: a forward look, proceedings of the 2nd Longleaf Alliance conference. Charleston, SC: Longleaf Alliance: 140-143.

SAS Institute Inc. 2004. SAS/STAT 9.1 user's guide. Cary, NC: SAS Institute Inc. 5121 p.

Shoulders, E. 1967. Fertilizer application, inherent fruitfulness and rainfall affect flowering of longleaf pine. Forest Science. 13(4): 376-383.

System Dynamics International. 1997. E user's manual. St. Louis, MO: System Dynamics International, Inc. 386 p.