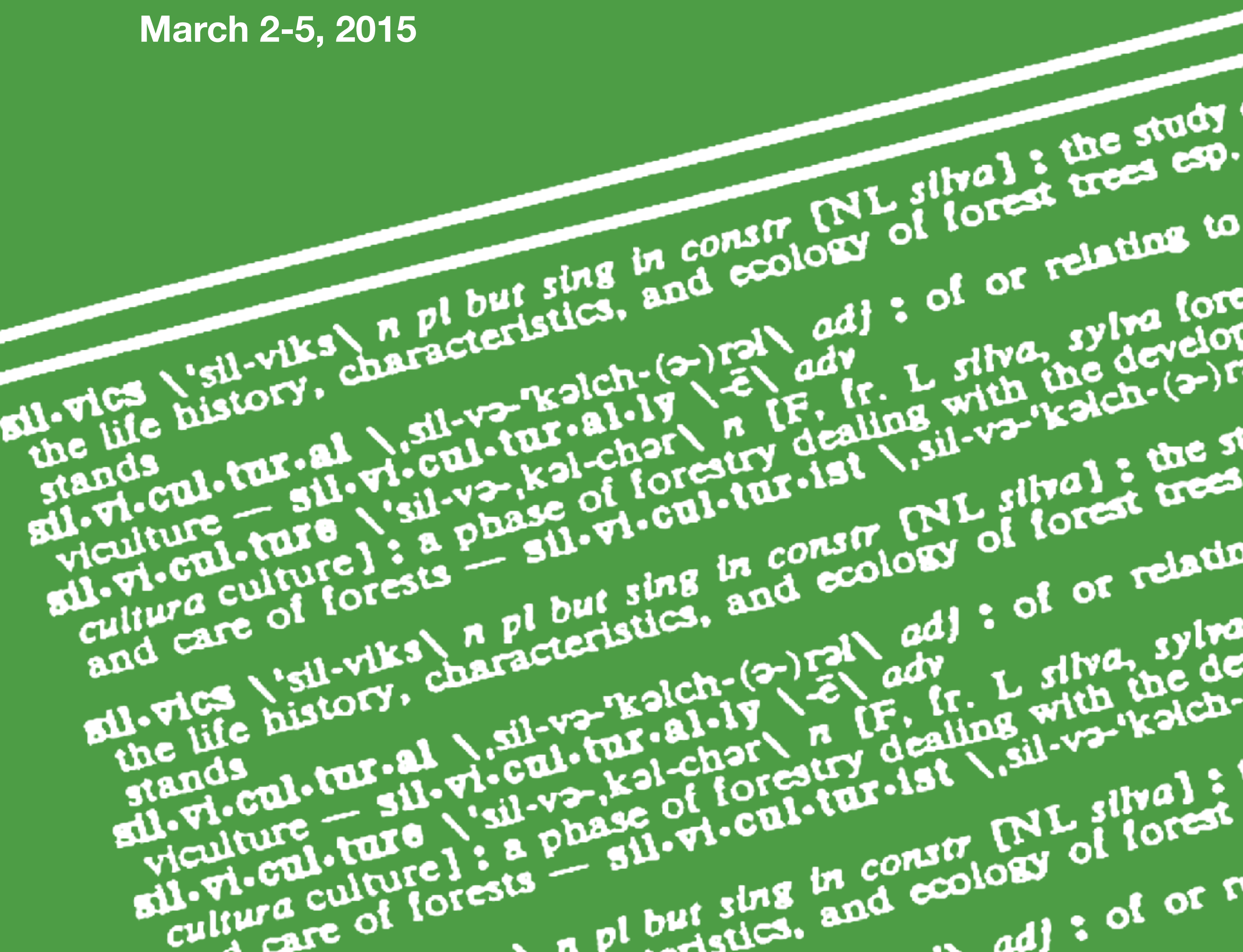
USDA
United States Department of Agriculture

Proceedings of the 18th Biennial Southern Silvicultural Research Conference

Knoxville, TN

March 2-5, 2015



Forest Service
Southern Research Station
e-General Technical Report SRS-212, January 2016

MODELING INDIVIDUAL TREE SURVIVAL

Quang V. Cao[1]

Abstract—Information provided by growth and yield models is the basis for forest managers to make decisions on how to manage their forests. Among different types of growth models, whole-stand models offer predictions at stand level, whereas individual-tree models give detailed information at tree level. The well-known logistic regression is commonly used to predict tree survival probability. In addition to the maximum likelihood approach, a new approach called CDF regression was introduced here to estimate parameters of the tree survival equation.

Each of the two above approaches was evaluated as follows: (1) unadjusted, (2) disaggregated from the whole-stand model, and (3) disaggregated from the combined estimator. Results from this study showed that the tree survival model, when adjusted from the combined estimator, produced the best-ranked two alternatives. The new method, CDF Regression, coupled with the combined estimator, was better than the Maximum Likelihood method in estimating parameters of the logistic regression equation.

INTRODUCTION

Among many different types of growth and yield models, individual-tree simulation models provide the most flexible outputs because growth of an individual tree is the basis for this type of models. Predicting tree survival is an important component of tree-level models. The probability that a tree survives a growing period has been modeled by use of logistic regressions (Hamilton 1974, Hamilton and Edwards 1976, Monserud 1976, Buchman 1979, 1983, Zhang and others 1997, Monserud and Sterba 1999) or other methods (Glover and Hool 1979, Amateis and others 1989, Guan and Gertner 1991a, 1991b).

Maximum likelihood estimation is the most common method for estimating the parameters of a logistic regression model. An alternative method, called CDF Regression, is introduced in this paper.

Stand-level prediction of survival can be predicted directly from a stand survival model, or indirectly by summing individual tree survival probabilities. The predictions could be improved by use of a combined estimator (Yue and others 2008, Zhang 2010), which is a weighted average of outputs from both types of models.

Disaggregation method is a method that links a tree-level model and a stand-level model (Ritchie and Hann 1997). In this method, outputs from the tree-level model are adjusted such that the resulting stand summary matches prediction from a stand-level model.

The objective of this study was to evaluate two methods of estimating parameters of the logistic regression model for predicting tree survival probabilities, Maximum Likelihood and CDF Regression. The evaluation was conducted under the following scenarios:

- Unadjusted tree survival model,
- Tree survival model adjusted by disaggregation from the stand survival outputs, and
- Tree survival model adjusted by disaggregation from the combined estimator.

METHODS

Data

Data used in this study were from 200 plots randomly selected from the Southwide Seed Source Study, which include 15 loblolly pine (*Pinus taeda* L.) seed sources planted at 13 locations across 10 southern states (Wells and Wakeley 1966). Each 0.0164 ha plot consisted of 49 trees, planted at a 1.8 m × 1.8 m spacing. Tree diameters and survival were recorded at ages 10, 15, 20, and 25 years, resulting in a total of 600 growth periods.

The data were randomly divided into two groups of 100 plots each (table 1). The leave-one-out evaluation scheme was applied in this study. Parameters of the tree survival model were estimated from one group, and then used to predict for the other group. The predictions

[1]Professor, School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803

Citation for proceedings: Schweitzer, Callie J.; Clatterbuck, Wayne K.; Oswalt, Christopher M., eds. 2016. 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 p.

Table 1—Stand and tree attributes for the 200 plots used in this study, by group

Group	Age	Stand attributes				Tree attribute	
		# obs.	*Hd*	*TPH*	*BA/ac*	*# obs.*	*Avg. DBH*
1	10	100	9.0	1987	21.52	3257	11.4
	15	100	13.2	1750	31.63	2868	14.7
	20	100	16.4	1303	33.63	2135	17.6
2	10	100	9.2	1977	22.08	3237	11.6
	15	100	13.4	1702	32.03	2788	15.0
	20	100	16.7	1243	33.22	2037	17.9

from both groups were used to compute evaluation statistics for the different methods.

Stand survival equation

The model developed by Cao (2006) was used in this study to predict stand-level survival:

$$\hat{N}_{2i} = \frac{N_{1i}}{1+exp(b_0+b_1RS_{1i}+b_2H_{1i}+b_3N_{1i}/A_{1i}+b_4/A_{1i})}, \quad (1)$$

where $\hat{N}_{2i}$ = predicted number of trees per hectare for plot i at the end of the 5-year growth period, A_{1i} = stand age in years for plot i at the beginning of the growth period, N_{1i} = number of trees per hectare at age A_{1i}, H_{1i} = dominant height in meters at age A_{1i}, $RS_{1i} = \frac{\sqrt{10000/N_{1i}}}{H_{1i}}$ = relative spacing at age A_{1i}.

Tree survival equation

The following logistic regression model was employed to predict tree survival probability ($\hat{p}_{1j}$) of tree *j* in plot *i* during the 5-year growth period:

$$\hat{p}_{ij} = \frac{1}{1+exp(b_0+b_1H_{1i}+b_2RS_{1i}+b_3d_{1ij}/Dq_{1i})}, \quad (2)$$

where Dq_{1i} = quadratic mean diameter of plot *i* at age A_{1i}, and d_{1ij} = dbh of tree *j* in plot *i* at age A_{1i}.

CDF Regression

Tree diameters in each plot were sorted from smallest to largest. Let $d_{1i(j)}$ be the j^{th} order statistic for tree dbh at age A_{1i} for plot *i*, with $d_{1i(j)}$ being the minimum dbh in plot *i* at age A_{1i}. The empirical CDF (cumulative distribution function) or tree survival, $F_{1i(j)}$, is defined for diameter $d_{1i(j)}$ as follows:

$$F_{i(j)} = F_{i(j\text{-}1)} + \delta_{i(j)}/n_{2i}\,. \quad (3)$$

where $F_{i(0)} = 0$, $\delta_{i(j)} = 0$ if the tree having diameter d_{1ij} is dead and 1 if it survives the 5-year growth period, and n_{2i} = total number of surviving trees in plot *i* at the end of the growth period. Subscript *j* varies from 1 to n_{1i}, where n_{1i} = total number of trees in plot *i* at age A_{1i}.

In the CDF Regression method, the parameters of the tree survival equation (2) were solved to minimize

$$z = \sum_i \sum_j \left(F_{i(j)} - \hat{F}_{i(j)}\right)^2, \quad (4)$$

where $\hat{F}_{i(j)} = \hat{F}_{i(j-1)} + \hat{p}_{i(j)}/n_{2i}$, $\hat{p}_{i(j)}$ = predicted tree survival probability for a tree having diameter $d_{1i(j)}$, and $\hat{F}_{i(0)} = 0$.

Figure 1 shows observed survival CDF for a sample plot and predicted CDF's from the Maximum Likelihood and CDF Regression methods.

Combined estimator

The combined estimator for plot *i* (N_i^C) is the weighted average of predictions from the stand-level model (N_i^S) and the tree-level model (N_i^T):

$$N^C = wN^T + (1-w)N^S, \quad (5)$$

where w were computed according to the least-squares method described by Tang (1992, 1994) and applied by Zhang and others (2010).

Disaggregation

The predicted tree survival probability ($\hat{p}_{ij}$) of tree *j* in plot *i* was adjusted as follows so that the resulting stand survival matched either the prediction from the stand survival model or the combined estimator:

$$\tilde{p}_{ij} = \hat{p}_{ij}^{\alpha_i}, \quad (6)$$

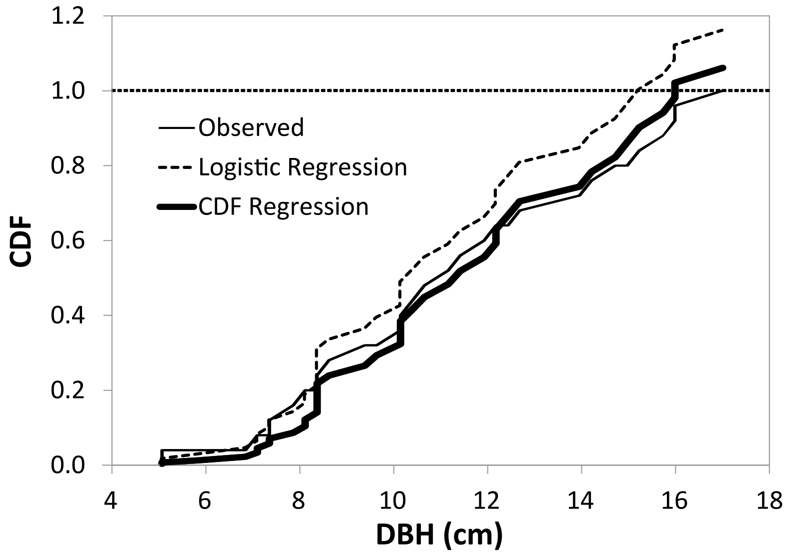


Figure 1—Observed survival CDF for a sample plot and predicted CDF's from the Maximum Likelihood and CDF Regression methods.

where $\tilde{p}_{ij}$ = adjusted tree survival probability, and
a_i = coefficient for plot i to adjust probabilities.

Evaluation Criteria

The performance of the different methods was evaluated at the stand level based on the following statistics.

Mean difference: $MD = \sum_i (N_{2i} - \hat{N}_{2i})/m$ (7)

Mean absolute difference: $MAD = \sum_i |N_{2i} - \hat{N}_{2i}|/m$ (8)

Fit index: $R^2 = \sum_i (N_{2i} - \hat{N}_{2i})^2 / \sum_i (N_{2i} - \bar{N}_2)^2$ (9)

where N_{2i} and $\hat{N}_{2i}$ = observed and predicted number of trees per hectare in plot i at the end of the growth period, $\bar{N}_2$ = the average number of trees per hectare at the end of the growth period, and m = total number of plots.

The stand-level evaluation included the whole-stand model, the individual-tree model (with two parameter estimation methods), and the combines estimator (also with two parameter estimation methods).

The tree-level evaluation statistics were:

Mean difference: $MD = \sum_i \sum_j (\delta_{ij} - \hat{p}_{ij}) / \sum_i n_{1i}$ (10)

where δ_{ij} = 0 if tree j in plot i was dead and 1 if it was alive, and n_i = number of trees in plot i.

Mean absolute difference:

$$MAD = \sum_i \sum_j |\delta_{ij} - \hat{p}_{ij}| / \sum_i n_{1i} \quad (11)$$

Log-likelihood:

$$-2lnL = -2\left[\sum_i \sum_j \hat{p}_{ij} \ln(\hat{p}_{ij}) + \sum_i \sum_j (1 - \hat{p}_{ij}) \ln(1 - \hat{p}_{ij})\right] \quad (12)$$

AUC: area under the ROC (Receiving Operating Characteristic) curve. The range for AUC is between 0.5 and 1. The higher the AUC value, the better the fit.

The tree-level evaluation involved six methods. Each of the two parameter estimation methods included three alternatives: the unconstrained model, the Disaggregation method in which outputs from the tree-level model was adjusted from the stand-level predictions, and the Combination method that adjusted the tree-level outputs to match the combined estimator.

The relative rank, developed by Poudel and Cao (2013), was used in this study to display the relative position of each method. The relative rank of method i is defined as

$$R_i = 1 + \frac{(k-1)(S_i - S_{min})}{S_{max} - S_{min}} \quad (13)$$

where R_i = the relative rank of method i $(i = 1, 2, ..., k)$, k = number of methods evaluated, S_i = the goodness-of-fit statistic produced by method i, S_{min} = the minimum value of S_i, and S_{max} = the maximum value of S_i.

RESULTS AND DISCUSSION

The individual-tree survival model, with parameters estimated by either the Maximum Likelihood or CDF Regression method, produced stand-level outputs that were inferior to the whole-stand survival model (table 2). This was expected because stand-level outputs from individual-tree models typically suffer from accumulation of errors (Qin and Cao 2006). On the other hand, the combined estimators did outperform the whole-stand model based on all three evaluation statistics (Table 2). The combined estimator from the CDF Regression method was slightly better than the one from the Maximum Likelihood method.

Table 3 shows the tree-level evaluation statistics for the six methods. The relative ranks for these methods are presented in Table 4.

Unconstrained models

For the unadjusted model, the Maximum Likelihood method clearly outperformed the CDF Regression method in predicting tree survival. It produced better statistics for all evaluation criteria.

Disaggregation

Based on the sum of the relative ranks, disaggregation was better for the CDF Regression than for the Maximum Likelihood method. This was true when disaggregation was either from the whole-stand model or from the combined estimator. The CDF Regression method was a compromise between optimizing for tree-level and stand-level survival prediction. As such, it made sense that this method performed well in conjunction with the combined estimator.

Radar plot

The radar plot based on the relative ranks of four criteria from all methods is shown in figure 2. The method resulting in the smallest area inside the box represents the best method. Figure 3 presents the relative ranks for the best three methods. The CDF Regression method, coupled with the combined estimator, overall ranked best in predicting tree survival, followed by the Maximum Likelihood method, unconstrained and disaggregated by the Combination method.

CONCLUSIONS

Results from this study showed that the Combination approach, in which outputs from the tree survival model were adjusted to match the combined estimator, was the best approach to predict both tree- and stand-level survival. The new method, CDF Regression, when disaggregated from the combined estimator, was better than the Maximum Likelihood method in estimating parameters of the logistic regression equation.

Table 2—Stand-level evaluation statistics for different types of models. Underlined, italic numbers denote the best statistic among the methods

Type	Parameter estimation	MD	MAD	R^2
Whole-stand model		-21.77	160.37	0.8295
Individual-tree model	Maximum Likelihood	***-0.05***	176.28	0.7963
	CDF Regression	55.77	180.36	0.7866
Combined estimator	Maximum Likelihood	-17.11	159.86	0.8322
	CDF Regression	-3.91	***159.40***	***0.8337***

Table 3—Tree-level evaluation statistics for different parameter estimation methods. Underlined, italic numbers denote the best statistic among the methods

Parameter estimation	Method	MD	MAD	-2lnL	AUC
Maximum Likelihood	Unconstrained	***0.0001***	0.2276	***0.7384***	0.7929
	Disaggregation	-0.0134	0.2141	0.7631	0.7821
	Combination	-0.0106	0.2166	0.7391	0.7932
CDF Regression	Unconstrained	0.0337	0.2315	0.7659	0.7907
	Disaggregation	-0.0134	***0.2057***	0.7788	0.7901
	Combination	-0.0026	0.2109	0.7480	***0.7997***

Table 4—Relative rankings for tree-level evaluation statistics. Underlined, italic numbers denote the highest rank among the methods

Parameter estimation	Method	Rank MD	Rank MAD	Rank -2lnL	Rank AUC	Total	Overall Rank
Maximum Likelihood	Unconstrained	***1.00***	5.24	***1.00***	2.93	10.16	2.33
	Disaggregation	3.00	2.61	4.07	6.00	15.68	4.39
	Combination	2.58	3.10	1.09	2.84	9.61	2.13
CDF Regression	Unconstrained	6.00	6.00	4.41	3.56	19.98	6.00
	Disaggregation	3.00	***1.00***	6.00	3.72	13.72	3.66
	Combination	1.40	2.01	2.20	***1.00***	6.60	***1.00***

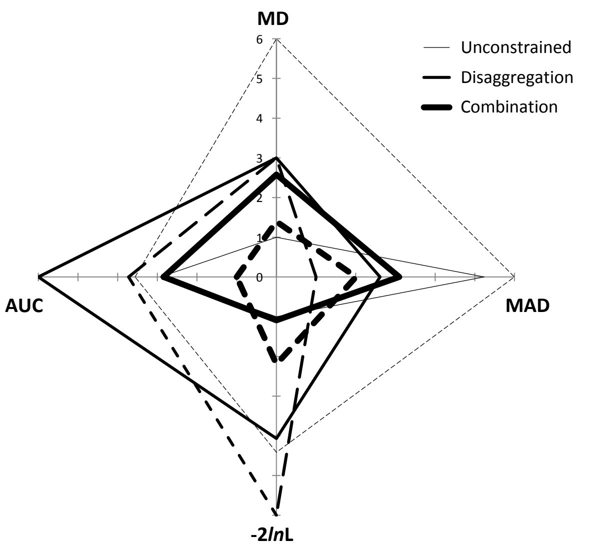


Figure 2—Overall comparison for the Maximum Likelihood (continuous lines) and CDF Regression (dashed lines) parameter estimation methods. Each method was evaluated unadjusted or disaggregated from either the stand model or the combined estimator. The method resulting in the smallest area inside the box represents the best method.

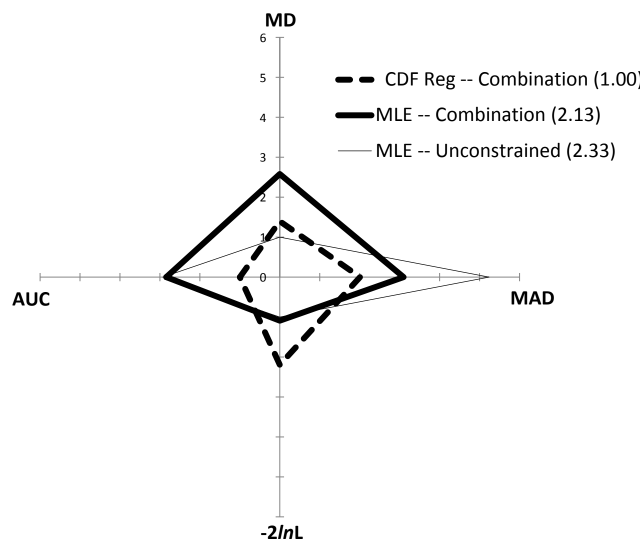


Figure 3—Relative ranks of the three best methods.

LITERATURE CITED

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

Buchman, R.G. 1979. Mortality functions. In: A generalized forest growth projection system applied to the Lake States region. Gen. Tech. Rep. NC-49. St. Paul, MN: U.S. Department of Agriculture Forest Service, North Central Forest Experiment: 47-55.

Buchman, R.G. 1983. Survival predictions for major Lake States tree species. Res. Pap. NC-233. St. Paul, MN: U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station. 10 p.

Cao, Q.V. 2006. Predictions of individual-tree and whole-stand attributes for loblolly pine plantations. Forest Ecology and Management. 236: 342-347

Glover, G.R.; Hool, J.N. 1979. A basal area ratio predictor of loblolly pine plantation mortality. Forest Science. 25: 275-282.

Guan, B.T.; Gertner, G. 1991a. Using a parallel distributed processing system to model individual tree mortality. Forest Science. 37: 871-885.

Guan, B.T.; Gertner, G. 1991b. Modeling red pine tree survival with an artificial neural network. Forest Science. 37: 1429-1440.

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

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

Monserud, R.A. 1976. Simulation of forest tree mortality. Forest Science. 22: 438-444.

Monserud R.A.; Sterba, H. 1999. Modeling individual tree mortality for Austrian forest species. Forest Ecology and Management. 113: 109-123.

Poudel, K.; Cao, Q.V. 2013. Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. Forest Science. 59: 243-252.

Qin, J.; Cao, Q. V. 2006. Using disaggregation to link individual-tree and whole-stand growth models. Canadian Journal of Forest Research. 36: 953-960.

Ritchie, M.W.; Hann, D.W. 1997. Implications of disaggregation in forest growth and yield modeling. Forest Science. 43: 223-233.

Tang, X. 1992. Forecasting theory with its applications (in Chinese with English abstract). Chengdu, China: University of Electronic Science and Technology of China: 288.

Tang, X. 1994. Some new results of the efficient frontier of portfolio investment. J. UEST China. 23:301-305.

Wells, O.O.; Wakeley, P. C. 1966. Geographic variation in survival, growth, and fusiform rust infection of planted loblolly pine. Forest Science Monograph. 11(Supplement to Number 3). 40 p.

Yue, C.F.; Kohnle, U; Hein, S. 2008. Combining tree- and stand-level models: a new approach to growth prediction. Forest Science. 54: 553-566.

Zhang, S.; Amateis, R.L.; H.E. Burkhart, H.E. 1997. Constraining individual tree diameter increment and survival models for loblolly pine plantations. Forest Science. 43: 414-423.

Zhang, X.; Lei, Y.; Cao, Q.V. 2010. Compatibility of stand basal area predictions based on forecast combination. Forest Science. 56: 552-557