

---

## A Three-Step Approach To Model Tree Mortality in the State of Georgia

Qingmin Meng<sup>1</sup>, Chris J. Cieszewski<sup>2</sup>, Roger C. Lowe<sup>3</sup>, and Michal Zasada<sup>4</sup>

**Abstract.**—Tree mortality is one of the most complex phenomena of forest growth and yield. Many types of factors affect tree mortality, which is considered difficult to predict. This study presents a new systematic approach to simulate tree mortality based on the integration of statistical models and geographical information systems. This method begins with variable preselection using multiple linear regression models and logistic models and employs spatial autocorrelation detection and random sampling. Three random sampling methods are applied and compared to reduce the effects of spatial autocorrelation, and systematic random sampling significantly reduces the spatial autocorrelation among the observations and is used for the final variable selection and model fitting. Using Forest Inventory and Analysis (FIA) data for the State of Georgia, this systematic approach provides significant implications for future tree mortality studies and other spatial analysis in forestry or geography.

Forest tree mortality is an important factor in nutrient cycling as well as global climate warming because mortality and net primary production are two critical processes of forest carbon budgets (Brown and Schroeder 1999). In addition, a large portion of the threatened forest species lives in dead wood (Rouvinen *et al.* 2002). At the same time, forest tree mortality may reduce the productivity of forests and increase the risk of wildfires. Tree mortality, however, is considered difficult to predict.

The literature contains many reports from different studies on tree mortality. For example, Greene *et al.* (1992) and Pedersen and McCune (2002) conducted research on mortality rates, using

wind as the primary disturbance contributing to tree mortality, and found that the total biomass declined from 1979 to 1989 because the new biomass production was less than mortality. Pedersen and McCune (2002) modeled tree mortality rates as a function of diameter at breast height (d.b.h.), species, decades, and site index. In their study, they reconstructed tree mortality rates for the years 1968 to 1977 and 1978 and 1987 in oak-hickory forest. Rouvinen *et al.* (2002) researched tree mortality, its causes, and spatial pattern in four transects with a total area of 48.8 ha of Vienansalo wilderness in eastern Fennoscandia, Finland. They divided mortality into three categories: current, recent, and predicted mortality. They concluded that tree mortality was continuous at the landscape scale, although some spatial aggregations occurred. Osawa *et al.* (1986) conducted systematic research on forest tree mortality using Baxter State Park as the study area and compared tree mortality among various onsite topographical conditions. Basal area, d.b.h., or stand age is supposed to significantly contribute to tree mortality (Fridman and Stahl 2001, Yang *et al.* 2003, Monserud and Sterba 1999, Avila and Burkhart 1992, and Zhang *et al.* 1997).

In the 1960s and 1970s, linear and polynomial models were commonly used (e.g., Lee 1971). Osawa *et al.* (1986) concluded, however, that multiple regression analysis was unsuccessful in relating tree mortality to forest structural characteristics and topographical properties. Nonlinear models, especially the logistic functions, have been the most widely used functions for mortality modeling from Walker and Duncan (1967) and Neter and Maynes (1970). Guan and Gertner (1991) pointed out that the best function to model individual tree mortality may be the logistic function based on statistical tests.

We conducted a systematic study on tree mortality in the State of Georgia. In our research, we performed three steps: variable preselection based on original data, sampling and spatial autocorrelation comparison, and model fitting and selection.

---

<sup>1</sup> Graduate Student Assistant.

<sup>2</sup> Associate Professor, Corresponding Author, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA. Phone: 706-542-8169; fax: 706-542-8356; home page: [www.growthandyield.com/chris](http://www.growthandyield.com/chris).

<sup>3</sup> GIS Analyst, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA.

<sup>4</sup> Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA; Assistant Professor, Department of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Nowoursynowska 159, 02-776 Warsaw, Poland.

## Study Area and Data

Our study area was the entire State of Georgia, the largest State east of the Mississippi River, with an area of 152,576 square kilometers. We based our research on available, online, forest inventory data provided by the Forest Inventory and Analysis (FIA) program.

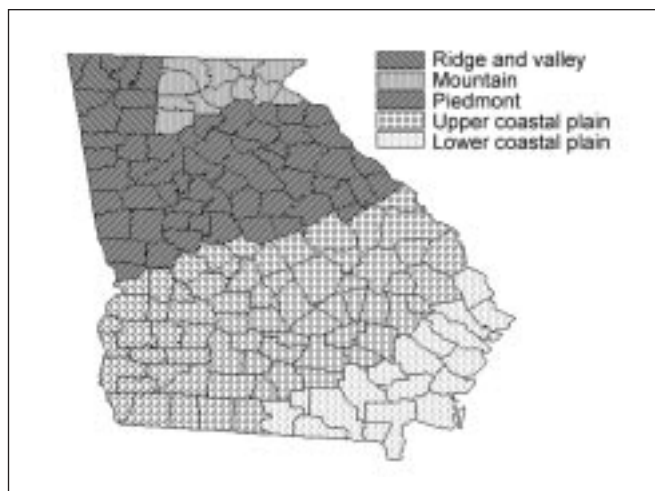
We analyzed plot data for the periodic 2001 inventories and considered every single FIA plot as the basic unit for our study. We processed the original data in this manner:

1. All forest type data, i.e. the data of all trees, was segregated into two broad forest categories, hardwoods and softwoods. All tree sizes inventoried by the FIA were included to account for any potential mortality of small trees.
2. The number of mortality trees per acre per year (TPAMORT) was used as the variable to calculate tree mortality on every plot and also as the response variable.
3. Latitude (LAT), longitude (LON), elevation (ELEV), condition proportion (CONDPROP), stand age (STDAGE), stand size code (STDSZCD), site productivity code (SITECLCD), slope (SLOP), aspect (ASPECT), physiographical code (PHYCLCD), growing-stock stocking code (GSSTKCD), stand treatment 1 code (TRTCD1), basal area of all live trees (BALIVE), current diameter (DIA), and trees per acre (TPACURR) are used as independent variables. STDSZCD, SITECLCD, PHYCLCD, GSSTKCD, and TRTCD1 are five categorical variables; the remaining 10 variables are numerical.
4. These data are aggregated into three groups: all trees, hardwood, and softwood.
5. In the sampling process, the State of Georgia is divided into five regions (fig 1): the northwest corner is a ridge and valley region; the northeast corner is a mountain region; north central Georgia is the piedmont region; south central Georgia is an upper coastal plain region; and the southeast corner is a lower coastal plain region. In each subregion, few differences exist in natural environment, landscape, or forest species.

## Methodology

We used four types of methods in our study. Multiple linear regression and logistic regression were applied for variable selection and model fitting. Sampling methods were used to

Figure 1.—Georgia's five study subregions.



make samples from the original data. For geographic information systems (GIS), Environmental Systems Research Institute, Inc. (ESRI) ArcInfo<sup>SM</sup> and Arcview<sup>®</sup> products are used to process location data and related attributed data and calculate the coefficients of spatial autocorrelation.

## Multiple Linear Regression and Logistic Regression

Equation 1 is a multiple linear regression function; equation 2 is a logistic regression function. The two functions, used for all trees, hardwood and softwood, have categorical variables. We applied a stepwise method to select variables in fitting logistic models:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{15} X_{15} \quad (1)$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{15} X_{15} \quad (2)$$

Where:  $Y$  is the numbers of tree mortality,  $p$  is the probability of tree mortality, and  $x_1, x_2, x_3, \dots$  and  $x_{15}$  are the variables of LAT, LON, ELEV, CONDPROP, STDAGE, STDSZCD, SITECLCD, SLOP, ASPECT, PHYCLCD, GSSTKCD, TRTCD1, BALIVE, DIA, and TPACURR.

## Sampling Methods

We used simple random sampling (SRS), the simplest form of random sampling. It is easy to perform and explain to others, a fair way to select a sample, and reasonable to generalize the

results from the sample back to the population. Second, we employed systematic random sampling (SYS), also fairly easy to perform. Third, we used a stratified random sampling (STS) method. For STS, the population of all trees, hardwood and softwood, were divided into five groups based to the five sub-regions (fig. 1) noted above.

### Geographic Information Systems

We used the following two types of coefficients of spatial autocorrelation this study. Equation 3 is Geary's coefficient C, and equation 4 is Moran's I coefficient.

$$C = \frac{\sum \sum W_{ij} C_{ij} / 2 \sum \sum W_{ij} \sigma^2}{\sum \sum W_{ij} C_{ij} / S^2 \sum \sum W_{ij}} \quad (3)$$

$$I = \frac{\sum \sum W_{ij} C_{ij} / S^2 \sum \sum W_{ij}}{\sum \sum W_{ij} \sigma^2} \quad (4)$$

where  $C_{ij}$  is the similarity of attributes,  $W_{ij}$  is the similarity of distance, and  $s^2$  and  $\delta^2$  is the variance of attributes.

## Results

### Variable Selection Based on Original Data

First, a multiple linear regression function is applied for all trees to select significant variables based on original data. In equation 1, the tree mortality numbers is the response, and the other 15 variables are independent variables. The residual plot indicates the absence of constant variance among the residuals. Then, a straightforward log transformation for the data of all trees, hardwood and softwood, is used. Next, the multiple linear regression models are fitted again for these three groups of data. The residual plots are good, and the linear models are acceptable. At the level of  $\alpha = 0.05$ , some differing variables exist, but some are also the same (table 1). For all trees, hardwood and

softwood, TPAGROW and DIA are significant variables, which means that tree density may be a critical factor for tree mortality, and tree size may be also a better variable for mortality prediction. For all trees and hardwood, ELEV is the other common variable that affects tree mortality. In addition, at the 0.05 level, the physiographic class variable is significant for hardwood mortality, and the treatment class variable is significant for softwood mortality.

The logistic models provided some of the same significant variables and some different ones, too, compared with multiple linear models. For all trees, only one variable, DIA, was still significant; the other four variables—BALIVE, CONPROP, SITECLC, and STDSZCD—are added, which means that basal area, condition proportion, site productivity class, and stand size class were important for mortality of all trees. For hardwood, the four variables DIA, TPAGROW, PHYSCLCD, and ELEV were still significant compared with the above multiple linear regression analysis, and two other variables, BALIVE and STDAGE, were added. Basal area and stand age were also important for hardwood mortality based on logistic regression analysis. For softwood, only one variable, TPAGROW, is still significant compared with the multiple linear regression, and three other variables that are important for softwood mortality were added: stand age, site productivity class, and stand size class.

### Sampling and Spatial Autocorrelation Calculation

Spatial autocorrelation is typically over looked in most tree mortality research. In our study, we calculated spatial autocorrelation for different kinds of sampling methods (table 2), and selected a better sampling method, SYS, for model fitting again.

### Variable Reselection and Model Fitting

For multiple regression functions, variables are reselected after log transformation of tree mortality data. Table 3 lists the significant variables.

Table 1.—Multiple linear regression analysis based on original data.

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	F-value	P-value		F-value	P-value		F-value	P-value
DIA	15.57	0.0001	DIA	30.74	0.0001	DIA	10.21	0.0015
TPAGROW	60.34	<.0001	TPAGROW	20.44	0.0001	TPAGROW	42.15	0.0001
ELEV	3.85	0.05	ELEV	9.97	0.0016	TRTCD	2.91	0.0338
			PHYSCLCD	1.87	0.0397			

Table 2.—Spatial autocorrelation comparison of different sampling methods.

	All trees		Hardwood		Softwood	
	Moran's I	Geary's C	Moran's I	Geary's C	Moran's I	Geary's C
No sampling	0.103	0.100	0.018	0.122	0.097	0.049
SRS	0.071	0.042	0.003	0.040	0.042	0.060
SYS	0.056	0.034	0.004	0.048	0.087	0.029
STRAT	0.138	0.059	0.006	0.036	-0.0009	0.039

Table 3.—Multiple regression analysis after SRS.

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	Wald-Chi	P		Wald-Chi	P		Wald-Chi	P
DIA	5.65	.0179	DIA	16.43	<.0001	TPAGROW	11.46	.0009
TPAGROW	22.12	<.0001	TPAGROW	3.91	.0488			
			ELEV	5.57	.0189			

Table 4.—Logistic regression analysis after systematical random sampling

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	Wald-Chi	P		Wald-Chi	P		Wald-Chi	P
BALIVE	9.02	0.0027	STDAGE	9.66	0.0019	STDSZCD	22.49	<0.0001
TPAGROW	6.42	0.0113	DIA	8.29	0.0040	STDAGE	7.87	0.0050
SITECLCD	12.35	0.0149	TPAGROW	5.32	0.0211	SITECLCD	11.36	0.0288
STDSZCD	12.74	0.0052						

Compared with outcomes of logistic regression without random sampling, the numbers of significant variables decreased, and some significant variables were no longer significant (table 4). For mortality of all trees, BALIVE, SITECLCD, and STDCLCD were still significant after SYS, and TPAGROW became a significant variable. For hardwood, STDAGE, DIA, and TPAGROW were still significant, but BALIVE, PHYSCLCD, and ELEV were no longer significant. For softwood, TPAGROW was no longer significant, but STDSZCD, STDAGE, and SITECLCD were still significant after SYS.

## Conclusions

Based on our analyses, the logistic mortality functions for all trees, hardwood and softwood, are equations 5, 6, and 7.  $\hat{Y}$  is the probability of tree mortality;  $ba$  is basal area;  $g$  is TPAGROW;

$s$  is the stand size class variable with from one to four classes;  $si$  is the site productivity class variable, which has from one to five classes;  $a$  is stand age; and  $d$  is DIA.

Tree mortality has several common characteristics. For all trees and softwood, STDSZCD classes 1 and 2 (large diameter and medium diameter classes) have the highest mortality probability; STDSZCD 4 (chaparral class) has a medium mortality probability; class 3 (small diameter class) has the lowest mortality probability. SITECLCD classes 2 (site productivity between 165 and about 224 cubic feet/acre/year) and 3 (site productivity between 124 and approximately 165 cubic feet/acre/year) have higher mortality probability than classes 4 (site productivity between 85 around 119 cubic feet/acre/year) and 5 (site productivity between 50 and about 84 cubic feet/acre/year). Class 1 (site productivity more than 225 cubic feet/acre/year) has a medium mortality probability. For all trees and hardwood, the probability of tree mortality slightly decreases as TPAGROW increases.

$$\hat{Y} = \frac{\ell \cdot \frac{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}}{1 + \ell \cdot \frac{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}}} \quad (5)$$

$$\hat{Y} = \frac{\ell \cdot \frac{-1.1575+0.0096a+0.0891d-0.0032g}{-1.1575+0.0096a+0.0891d-0.0032g}}{1 + \ell \cdot \frac{-1.1575+0.0096a+0.0891d-0.0032g}{-1.1575+0.0096a+0.0891d-0.0032g}} \quad (6)$$

$$\hat{Y} = \frac{\ell \cdot \frac{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}}{1 + \ell \cdot \frac{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}} \quad (7)$$

Much research indicates that logistic regressions appear to be the best method for individual tree mortality modeling and have been widely applied (Monserud, 1976; Monserud and Sterba 1999, Fridman and Stahl 2001, Woolons 1998, Yang *et al.* 2003). In our study, logistic models are selected, and tree mortality analysis is summarized based on these logistic models.

## Literature Cited

Avila, O.B.; Burkhardt, H.E. 1992. Modeling survival of loblolly pine trees in thinned and unthinned plantations. *Canada Journal of Forest Research*. 22: 1878–1882.

Brown, S.L.; Schroeder, P.E. 1999. Spatial patterns of above-ground production and mortality of woody biomass for Eastern U.S. forests. *Ecological Application*. 9(3): 968–980.

Fridman, J.; Stahl, G. 2001. A three-step approach for modeling tree mortality in Swedish forests. *Scandinavian Journal of Forest Research*. 16: 455–466.

Greene, S.E.; Harcombe, P.A.; Harmon, M.E.; Spycher, G. 1992. Patterns of growth, mortality and biomass change in a coastal picea-sitchensis-tsuga-heterophylla forest. *Journal of Vegetation Science*. 3(5): 697–706.

Guan, B.T.; Gertner, G. 1991. Using a parallel distributed-processing system to model individual tree mortality. *Forest Science*. 37(3): 871–885.

Hosmer, D.W.; Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley & Son. 309 p.

Lee, Y.L. 1971. Predicting mortality for even-aged stands of lodge-pole pine. *The Forestry Chronicle*. 47: 29–32.

Monserud, R.A. 1976. Simulation of tree mortality. *Forest Science*. 22: 428–444.

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

Neter, J.; Maynes, E.S. 1970. On the appropriateness of the correlation coefficient with a 0, 1 dependent variable. *Journal of the American Statistical Association*. 65: 501–509.

Osawa, A.; Spies, C.J.; Dimond, J.B. 1986. Patterns of tree mortality during an uncontrolled spruce budworm outbreak in Baxter State Park, 1983. *Maine Agricultural Experiment Station Technical Bulletin*. 121: 1–69.

---

Pedersen, B.S.; McCune, B. 2002. A non-invasive method for reconstructing the relative mortality rates of trees in mixed-age, mixed-species forests. *Forest Ecology and Management*. 155(1–3): 303–314.

Rouvinen, S.; Kuuluvainen, T.; Siitonen, J. 2002. Tree mortality in a *Pinus sylvestris* dominated boreal forest landscape in Vienansalo wilderness, eastern Fennoscandia. *Silva Fennica*. 36(1): 127–145.

Walker, R.H.; Duncan, D.B. 1967. Estimation of the probability of an event as a function of several independent variables. *Biometrika*. 54: 167–179.

Woollons, R.C. 1998. Even-aged stand mortality estimation through a two-step regression process. *Forest Ecology and Management*. 105: 189–195.

Yang, Y.; Titus, S.J.; Huang, S. 2003. Modeling individual tree mortality for white spruce in Alberta. *Ecological Modeling*. 163: 209–222.

Zhang, S.; Amateis, R.L.; Burket, H.E. 1997. Constraining individual tree diameter increment and survival models for loblolly pine plantations. *Forest Science*. 43(3): 414–423.