Empirical and process-based models for forest tree and stand growth simulation
21-27 September 97, Oeiras, Portugal

DIAMETER-GROWTH MODEL ACROSS SHORTLEAF PINE RANGE USING REGRESSION TREE ANALYSIS

Daniel Yaussy , Louis Iverson and Anantha Prasad
Department of Agriculture, Forest Service, USA

ABSTRACT

Diameter growth of a tree in most gap-phase models is limited by light, nutrients, moisture, and temperature. Growing-season temperature is represented by growing degree days (gdd), which is the sum of the average daily temperatures above a baseline temperature. Gap-phase models determine the north-south range of a species by the gdd limits at the north and south boundaries of the realized niche. An assumption of these models is that a species will reach maximum diameter growth at the midpoint of its gdd range and that growth will taper parabolically to gdd limits. One might assume that diameter growth would increase toward the southern edge of a species realized niche, and that factors other than temperature would determine the southern boundary.

The USDA Forest Service has remeasured the diameters of approximately 200 species of trees in the eastern United States, storing this information in a geo-referenced data base. Environmental data have been assembled from nationwide GIS coverages, including soils, digital elevation maps, climate data bases, and others. Using these data we developed and tested two methods in addition to the gap-phase model to model changes in annual diameter growth over the geographic range of species occurrence. Stepwise Regression (SR) and Regression Tree Analysis (RTA) were used to determine the environmental and geographic variables associated with different rates of diameter growth across the species range. SR provided a linear approach to model and predict diameter growth. RTA is an exploratory technique for uncovering structure in data and fitting models by recursive partitioning of the data. RTA is better at capturing interactions between variables than traditional linear models. These modeling techniques are demonstrated with shortleaf pine, Pinus echinata.

INTRODUCTION

Gap-phase (or gap) models are used widely to simulate forest response to management or environmental change (Urban *et al.* 1991; Shugart and Smith 1996). An important use of these models is simulating the effects of predicted global climate change on the long-term composition of forest tree species from increased greenhouse gases (Solomon 1986; Solomon and Bartlein 1992; Bowes and Sedjo 1993; Bugmann 1996; Talkkari and Hypén 1996). Individual-tree survivability and mortality are determined by the rate at which the tree's diameter increases. These models use competition for light, water, and nutrients to reduce the potential diameter growth of a tree. The first-generation gap models (based on the equations in JABOWA, Botkin *et al.* 1972) rely on an accumulation of the number of days and the degree to which the air temperature exceeds a threshold temperature within a calendar year (growing degree days or gdd) to determine the potential diameter growth for a tree. This potential growth is then decreased, by species, based on site-specific moisture, light, and nutrient availability. A north/south range for a species is defined by the gdd limits of the species' realized niche. To maintain a species within the assigned gdd range, potential diameter growth is assumed to be maximized at the midpoint of the gdd range, decrease parabolically toward the extremes, and equal zero at the extremes (figure 1). Botkin *et al.* (1972), the originators of the gap model, derived this relationship from the observations of relationships of net photosynthesis and immediate temperature and the apparent coincidence of species ranges and temperature isotherms (Botkin 1993). Multiple years of little or no diameter growth trigger mortality for the trees in these simulations. Thus, the tree species cannot survive beyond its assigned gdd range due to a lack of diameter growth. Many ecologists believe that the failure of a species to survive outside of its realized niche is due to factors other than temperature, such as competition from species better adapted to the ecological conditions at the extremes of a species' range (Bonan and Sirois 1992; Bugmann *et al.* 1996; Loehle 1996; Loehle and LeBlanc 1996; Schenk 1996). The assumption that diameter growth is so closely tied to gdd has been cited as a major flaw in the gap models that use this potential annual diameter growth equation (GAP). This equation is being replaced in newer versions (Bugmann *et al.* 1996).

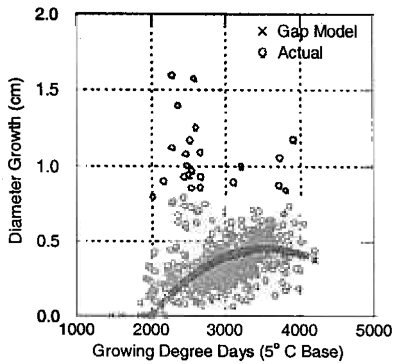

Figure 1. Actual county-level annual diameter-growth rates and those predicted by the GAP algorithm plotted over average growing degree days for each county.

The processes responsible for individual-tree diameter growth include the mechanisms that determine carbon allocation throughout the tree. Models of these processes have been developed for a several species (Korol *et al.* 1996, Retzlaff 1996), but many others still lack this information. Regional models will require parameterizations of many more species before process models of diameter growth are useful. Data are available to develop empirical regional models of annual diameter growth for most tree species in the United States.

Empirical growth modelers have found that basal-area growth is a better indicator of tree growth than change in diameter (Hilt 1983; Hilt and Teck 1989; Shifley 1987). Diameter growth of larger trees slows due to size while the increment of circumference of wood area may not decrease. We present a method to model the changes in both diameter and basal-area growth over the geographic range of species occurrence. The USDA Forest Service has remeasured the diameters of approximately 200 tree species in the eastern United States, storing this information in a geo-referenced data base (Hansen *et al.* 1992). Environmental data have been assembled from nationwide GIS coverages including soils, digital elevation maps, climate data bases, and others (Olson *et al.* 1980; Soil Conservation Service 1991; U.S. Environ. Prot. Agency 1993; U.S. Geol. Surv. 1987). Regression Tree Analysis (RTA) was used to determine the environmental and geographic variables associated with different rates of diameter and basal-area growth across the species range. We demonstrate this RTA modeling technique with shortleaf pine (*Pinus echinata*), and compare our predictions with those obtained with forward stepwise linear regression (SR) and the model used in the first-generation gap models (GAP).

STUDY LIMITATIONS

A study of this type carries a suite of assumptions and limitations that must be stated. These fall into the following categories: data inputs, analysis, and biology.

Data inputs.—When multiple GIS layers from disparate sources and scales are overlaid, there are error propagations throughout the data (Walsh *et al.* 1987). This impact is minimized in this study because a large sampling unit, the county, is the common spatial unit, though the scale of impact of some factors is much smaller than the county, i.e., some counties are highly diverse and some important ecological factors could be averaged out at the county scale. For example, small zones of high elevation, bottomlands, or unique soils, may be lost in the averaging process. Also, there is error associated with the Forest Service's Forest Inventory and Analysis (FIA) sampling, i.e., these data were not collected with the intent of developing diameter-growth models, intervals between remeasurements are inconsistent, and measurements of diameter at breast height (dbh) were not taken by the same people or possibly at the same height following long intervals between measurements.

Analysis.—Correlation among variables may cause some of the regression trees to be less interpretable than they might be if a fewer number of interpretable and independent variables were used. However, there are fewer problems related to multicolinearity with RTA than with linear methods, since the models are rule-based rather than parametric (Moore *et al.* 1991). Forward stepwise regression typically avoids such problems by using only the independent variables that most improve the fit of the model.

Biology.—Neither the RTA nor the stepwise methods described here account for physiological processes in the model outputs. We are not modeling the processes associated with diameter growth but modeling the empirical associations between diameter growth and large-scale ecological variables as surrogates to these processes. There is no guarantee that the model predictions will behave logically should events occur that alter the physiological processes.

One advantage of the diameter-growth model used by first-generation gap models was that it also was used to limit a species range. If the methods developed here are to replace the diameter-growth equation in these gap models, another method for determining a species' range must be implemented, such as those of Iverson *et al.* (in press) or Schwartz (1993, 1996; Schwartz *et al.* 1996). Iverson *et al.* Uses RTA to predict future ranges of tree species based on the same information used here. Schwartz (1996) estimates the rate at which a species can migrate across a fragmented landscape.

METHODS

Data

Data were extracted from several sources for the eastern United States. The county was chosen as the mapping unit because it is the reporting unit for much of these data and is of similarly size in the East. We evaluated 45 environmental variables for each of the 635 counties for which shortleaf pine was a dominant or codominant tree on more than one plot within the county (Table 1).

Table 1. County-level variables and descriptive statistics of the development data set.

Variable	Definition	Mean	Minim	Maxim
TAWC	Total available water capacity (cm)	18.1339	9.720	28.603
CEC	Cation exchange capacity	4.360	0.000	19.739
PH	Soil pH	4.205	2.067	6.603
PERM	Soil permeability rate (cm/hr)	4.867	1.161	22.245
CLAY	Percent clay (<0.002 mm size)	25.110	9.278	39.873
BD	Soil bulk density (gm/m^3)	0.838	0.115	1.510
INCH3	Weight % of rock 7.620-25.400 cm	2.021	0.000	11.938
NO10	% passing sieve No. 10 (coarse)	77.129	39.350	100.320
NO200	% passing sieve No. 200 (fine)	51.611	25.630	88.760
KFFACT	Soil erodibility factor, free of rock	0.293	0.122	0.465
OM	Organic matter content (% by	1.621	0.754	7.195
ROCKDE	Depth to bedrock (cm up to	135.179	78.461	152.400
SLOPE	Soil slope (%)	11.220	1.009	49.599
PSP	Potential soil productivity, m^3	5.702	0.228	8.161
SOILOK	Soils with no limitations (%)	1.868	0.000	14.590
SLSVLTS	Soils with severe limits (%)	20.356	0.910	63.860
SLVSLTS	Soils with very severe limits (%)	15.164	0.000	58.560
SLWET	Wet or stony level soils (%)	3.019	0.000	36.640
SLRANG	Soil suitable for range/forest (%)	11.298	0.000	40.260
SLWILD	Soil suitable for forest/wildlife (%)	18.384	0.000	93.390
SLNOAG	Soil not suitable for cultivation (%)	0.130	0.000	16.930
DISTLND	Disturbed land (%)	0.417	0.000	13.360
ALFISOL	Alfisol (%)	5.694	0.000	99.990
ENTISOL	Entisol (%)	1.748	0.000	98.46
INCEPTS	Inceptisol (%)	9.649	0.000	100.000

Table 1. County-level variables and descriptive statistics of the development data set (Cont.).

Variable	Definition	Mean	Minim	Maxim
MOLLISO	Mollisol (%)	1.721	0.000	78.490
SPODOS	Spodosol (%)	0.262	0.000	59.200
ULTISOL	Ultisol (%)	78.573	0.000	100.000
VERTISO	Vertisol (%)	0.660	0.000	79.980
GROWD	Number of frost-free days/year	218.978	158.000	281.000
ANNMOI	Annual precipitation/PET	41.880	-7.000	134.000
MINELV	Minimum elevation (m)	110.625	0.000	488.202
MAXELV	Maximum elevation (m)	354.050	17.994	1949.88
ELVCV	Elevation coefficient of variation	24.582	6.226	103.272
MAYSEP	Mean May-September temperature	23.569	15.900	27.500
JARPET	July-August precipitation/PET	0.966	0.300	2.200
JANT	Mean January temperature (°C)	4.854	-1.730	10.660
JULT	Mean July temperature (°C)	25.921	18.250	29.450
AVGT	Mean annual temperature (°C)	15.830	8.530	20.630
PET	Potential evapotranspiration	74.720	33.130	106.430
WINDS	Wind speed (m/second)	46.520	43.710	67.850
VP	vapor pressure (Pascals)	1382.33	974.000	1817.00
PPT	Annual precipitation (mm/yr)	1286.95	990.000	1822.00
DBHOLD	Mean initial DBH (cm)	28.052	10.932	52.724
GDD	Growing degree days (days/year)	3755.69	1819.07	5194.87

Diameter and basal-area growth values.— The Forest Service has a mandate to periodically determine the extent, condition, and volume of growth and removals of timber on the nation's forest land. This is done by six FIA units. Four of these produced a data base of standard format called the Eastwide Data Base (EWDB) for the 37 states from North Dakota to Texas and east. These data are stored in three record types (Hansen *et al.* 1992): county data, plot data, and tree data. Because the entire range of shortleaf pine is contained within the counties represented in the EWDB, that species was chosen for the modeling method demonstrated here. We used data from 5,148 forested plots that contained shortleaf pine to summarize desired county-level information needed for this study. Note that the EWDB does not provide information on precise plot location, so countywide averages were used for FIA data (and all other data) in the analyses.

We summarized the information for individual forested plots with shortleaf pine trees that were relatively free from competition for light. These trees were classified as dominant, codominant or open grown. Plots had been classified as poletimber or sawtimber stands indicating that the average dbh (outside bark) was greater than 12.7 cm (5 inches) or 27.94 cm (11 inches) at 1.37 m (4.5 feet), respectively. For each tree used in this study, dbh was measured twice over a period of 1 to 20 years. Basal area was calculated for each tree for both dbh measurements. The difference between the two dbh and basal-area measures was divided by the time interval between measurements to obtain annual rates of diameter and basal-area growth for each tree.

Average rates of growth were calculated for each county that contained at least two plots with shortleaf pine that met all of the specifications. Descriptive statistics for annual diameter and basal-area growth are presented in Table 2.

Table 2. County based annual basal-area and diameter growth for shortleaf pine for the development and validation data sets.

	Development		(n = 409)	Validation		(n = 113)
Variable	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Basal-area growth (mm^2)	19.952	2.500	102.750	19.070	2.330	47.540
Diameter growth (cm)	0.379	0.064	1.573	0.387	0.089	1.166

Climatic factors. — Monthly means (averaged from 1948 to 1987) of precipitation, temperature, potential evapotranspiration, windspeed, and vapor pressure for the current climate were extracted from a data base generated by the Environmental Protection Agency (1993). These data had been corrected and interpolated into 10 x 10 km grid cells for the conterminous United States. From these data we also calculated annual means of each of the variables mentioned, and selected two derived factors on the basis of their physiological importance to tree growth for this region: July-August ratio of precipitation to potential evapotranspiration (JARPET) (the time most prone to drought stress in the eastern United States), and May-September (i.e., growing season) mean temperature (MAYSEPT). The data were aggregated to county averages via weighted averaging by area. In addition, the GEOECOLOGY (Olson *et al.* 1980) county-level data base was used to provide county-level estimates of frost-free days

$$Gdd = \frac{365}{\pi}\left[\frac{tjul - tjan}{2} - \left(bt - \frac{tjul + tjan}{2}\right)\left(1 - \frac{bt - (tjul + tjan)/2}{tjul - tjan}\right)\right]$$

(GROWDAYS) and an annual ratio of precipitation to PET (ANNMOIST). Gdd values for each county were calculated using a sine-wave equation presented by Botkin (1993):

where:

π = 3.1416

tjul = average maximum temperature for July for each county (U.S. Environ. Prot. Agency 1993)

tjan = average minimum temperature for January for each county (U.S. Environ. Prot. Agency 1993)

bt = baseline temperature below which, it is assumed, no photosynthesis will occur.

Botkin (1993) assumes bt = 4.44°C (40°F), while Urban (1990) uses bt = 5.55°C (42°F). For this paper, we split the difference and assumed bt = 5°C (41°F).

Soil factors.-- The State Soil Geographic Data Base (STATSGO) was developed by the USDA Soil Conservation Service (now Natural Resource Conservation Service) to help meet that agency's mandate to collect, store, maintain, and distribute soil-survey information for U.S. lands. STATSGO includes data on the physical and chemical soil properties for about 18,000 soil series recognized in the nation (USDA Soil Conserv. Serv. 1991). STATSGO maps were compiled by generalizing more detailed soil-survey maps into soil associations in a scale more appropriate for regional analysis (1: 250,000).

Preprocessing was required before maps of particular attributes could be produced on a national scale. We selected 15 variables related to tree-species habitat: pH, available water capacity, organic matter, permeability, bulk density, cation-exchange capacity, depth to bedrock, K factor, slope, potential soil productivity, and several variables related to texture (e.g., percent clay, percent coarse fragments, percent volume of soil flowing through screens with meshes of various sizes). Calculations were performed by depth layer and spatial distribution, as described in Iverson *et al.* (1996), to create county weighted averages.

Additional soil information was obtained from the GEOECOLOGY data bases (Olson *et al.* 1980), including percentage of the county in each of seven soil orders, five levels of limitations for agriculture, level of disturbance, and suitability to range, forest, or wildlife.

Elevation.-- Maximum, minimum, and variation of elevation (i.e. coefficient of variation of elevation within the county) were derived for each county from 1:250,000 U.S. Geological Survey (USGS) Digital Elevation Model files obtained from the USGS internet site (U.S. Geol. Surv. 1987).

Regression tree analysis

A relatively new technique being used in the ecological sciences, RTA, uses iterative splitting of the data to develop empirical relationships between response and predictor variables rather than requiring the more restrictive distributional assumptions in classical regression functions. This alternative modeling approach creates models that are fitted by binary recursive partitioning whereby a data set is successively split into increasingly homogeneous subsets that elucidate relationships between predictor variables and the response variables (Clark and Pregibon 1992). Thus, RTA is much more flexible than classic statistical methods in uncovering structure in data that have variables that may be hierarchical, nonlinear, nonadditive, or categorical in nature. RTA is rapidly gaining in popularity as a means of devising prediction rules for rapid and repeated evaluation, as a screening method for variables, as a diagnostic technique to assess the adequacy of linear models, and for summarizing large multivariate data sets (Clark and Pregibon 1992).

In order to split the data, observations are ordered for each predictor variable. Calculating the value of each predictor variable produces the minimum variance within each resulting subset. The predictor variable that produces the most homogenous subsets is chosen to split the data; this process is repeated for each subset. RTA models the response as a discontinuous function of the predictors.

There are several key advantages in using RTA in our application, which covers a wide spatial domain, over classic statistical methods (Verbyla 1987; Michaelsen *et al.* 1994; Breiman et al 1984): (1) adeptness of RTA to capture nonadditive behavior, where relationships between the response variable and some predictor variables are conditional on the values of other predictors. Interaction effects can be elucidated effectively by first subsetting the data by the first predictor, then identifying two different predictors to subset the data at the next level of analysis, and so on, until the terminal nodes are reached. RTA does this subsetting without specifying the interaction terms in advance of the statistical analysis (as is required in multiple linear regression). For example, in our study, the factors associated with the northern limit for a species may differ greatly from the factors that regulate the southern limit of the species, (2) This advantage allows, in effect, a stratification of the country so that some variables may be most related to the diameter growth of a species for a particular region of the country, while a different set of variables may apply elsewhere. The same variables need not apply equally everywhere. See Breiman *et al.* (1984), Clark and Pregibon (1992), Michaelsen *et al.* (1994) for details on the regression and classification tree analysis. For this paper we used RTA with SYSTAT v. 7.0 (SPSS 1997).

Stepwise regression analysis

The SR procedure available in most statistical analysis packages (SAS 1990; SPSS 1997), can be used to create a parsimonious model from a large number of independent variables that may be correlated. The resulting model usually contains independent variables that are not highly correlated. P-values for entry into and removal from the model were set at 0.05. Tolerance for correlated independent variables was set at 1.0×10^{-11}. Since the process of entering and removing these variables from the model is iterative, the significance and fit statistics (e.g., R^2) usually associated with regression models are inflated and should be viewed skeptically (SPSS 1997). These statistics are valid for a priori models only, so the models should be tested for validity on an independent data set, as was done for this study.

The models parameterized with the development data set were applied to the data that was set aside for validation. Our interest was in comparing the newly developed algorithms with those used in gap models and the algorithms' abilities to predict diameter growth. Therefore, all comparisons were conducted with the estimates of diameter growth. Model performance was evaluated on accuracy and precision. As defined here, accuracy measures whether the predictions are consistently higher or lower than the actual values. Precision is a measure of the spread of the prediction errors. The statistics calculated for each measure are, for accuracy:

$$\text{Bias} = n_v^{-1} \sum_{i=1}^{n_v} (\hat{y}_i - y_i)$$

and precision:

$$Mean\ Squared\ Error = n_v^{-1} \sum_{i=1}^{n_v} (\hat{y}_i - y_i)^2$$

where:

n_v = number of counties in the validation data set
$\hat{y}_i$ = predicted diameter growth for county i
y_i = actual diameter growth for county i
i = county index ($1,...,n_v$).

Bias is a measure of how close the mean of the predicted values is to the mean of the actual observations. Mean Squared Error (MSE) measures the average squared distance between the predictions and the actual observations.

RESULTS

RTA Model

Plots of the residuals over the predicted values of the RTA models for diameter and basal-area growth indicated that the variance of the errors increased as the predictions increased. Therefore, natural log transformations were calculated for the variables and the models were reanalyzed (figure 2a-b).

The percent reduction in error (PRE) is a fit statistic used by SYSTAT with RTA analysis. PRE represents the proportion of variation in the data explained by the model similar to R^2 (SPSS 1997). The PRE statistic for the log diameter growth model (0.725) was higher than that for the log basal-area growth model (0.520). This is contrary to findings of other empirical modelers of diameter growth (Hilt 1983; Hilt and Teck 1989; Shifley 1987), who found that basal-area growth is a more predictable variable. However, the trees used in this study were constrained to be the largest, and fastest growing, so there were fewer differences in initial diameter than in the entire population of shortleaf pine trees. Thus, in the remainder of this paper, we focus on the results of the diameter-growth model.

If allowed, RTA would continue until each observation was identified as a node in the tree. We adjusted the tolerances of the procedure so that each node contained at least five observations. The tree fit for this study contained 31 terminal nodes. We "pruned" that tree to the most significant 21 nodes (figure 3). The first split in the tree was based on average temperature for the months of May through September, which is highly correlated to gdd ($r = 0.97$). Further divisions were determined by variables related to the suitability of the land to forestry and agriculture, soil chemistry, coarseness of the landforms, and initial dbh. The separations appear logical such that it is predicted that counties with warmer climates, less slope, and more permeable soils will have larger diameter growth rates.

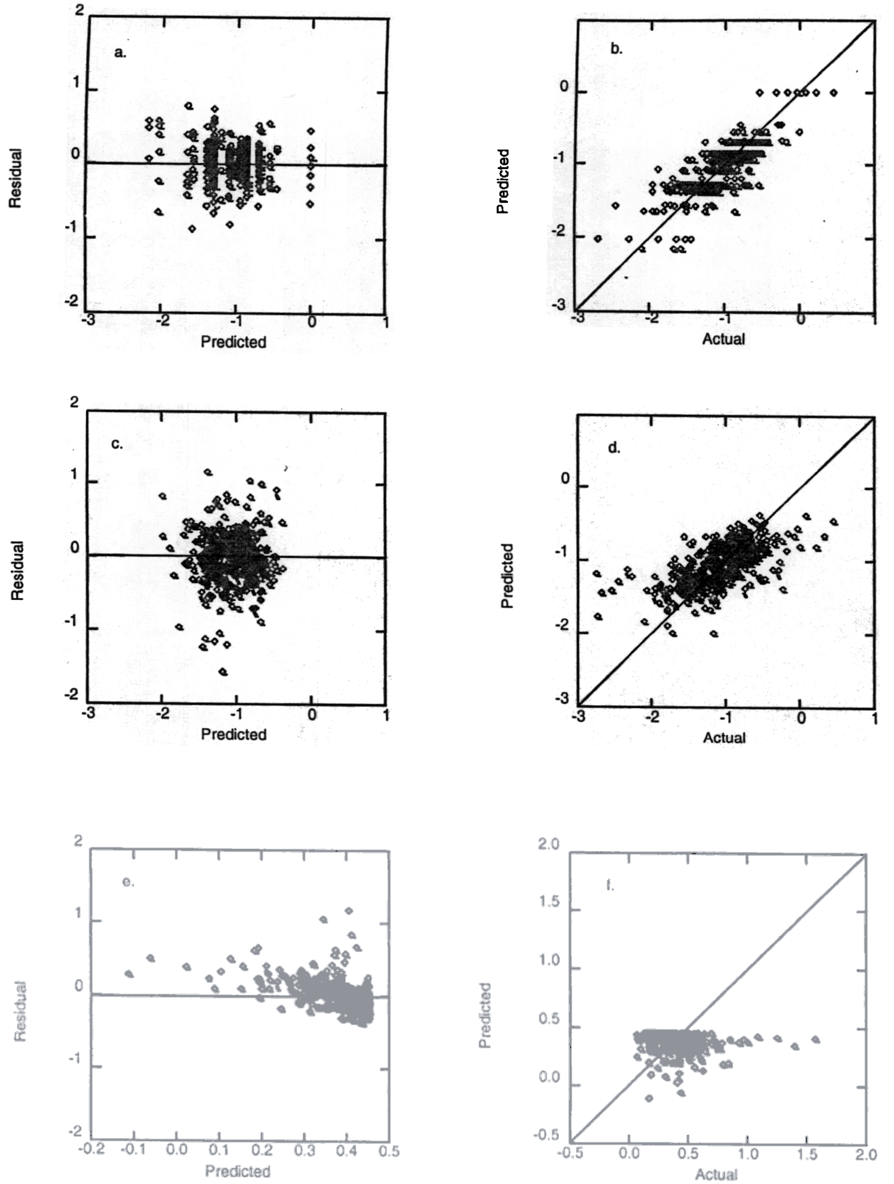

Figure 2. Residual versus predicted values and predicted versus actual values of ln(diameter growth), respectively, for RTA (a-b), SR (c-d), and diameter-growth values for GAP (e-f) for shortleaf pine (diameter growth in centimeters).

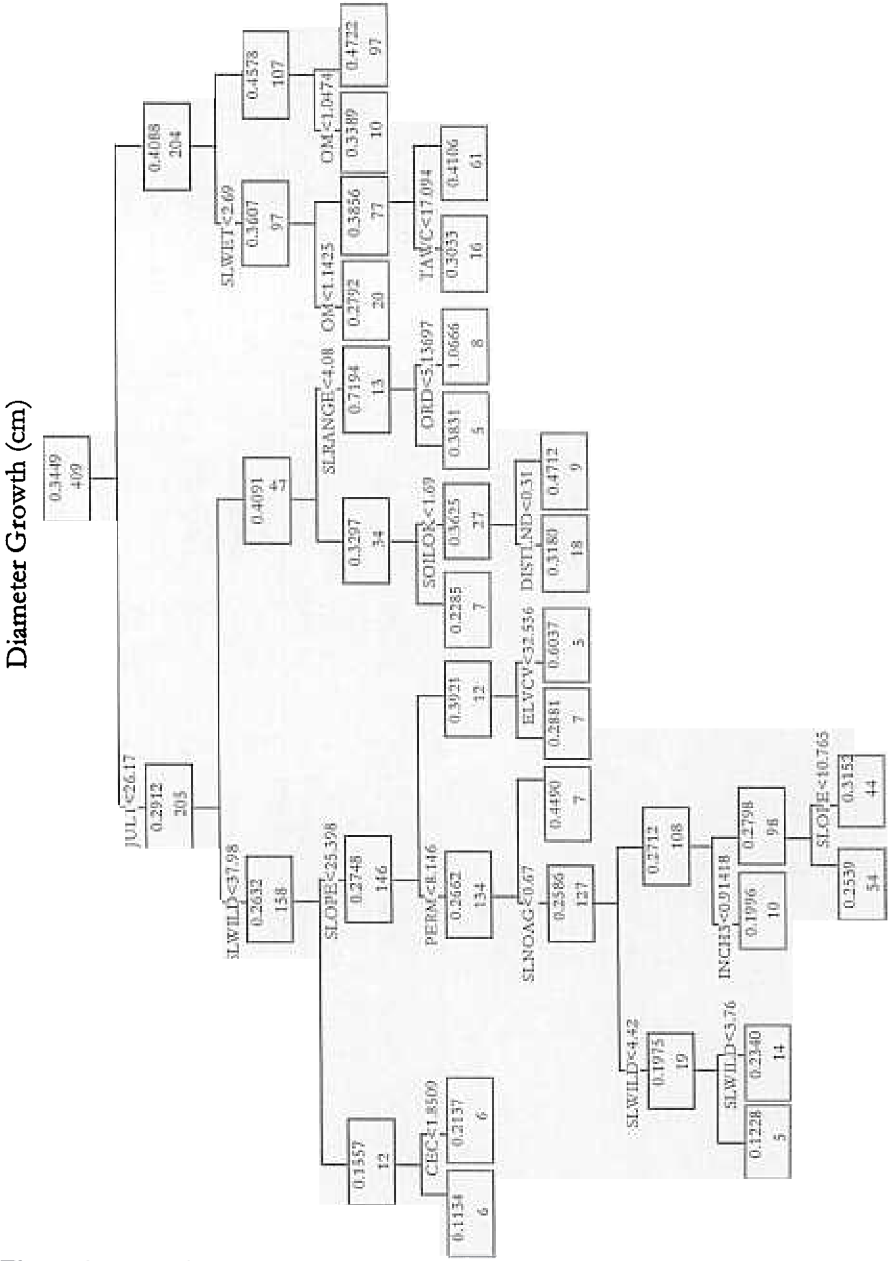

Figure 3. Regression tree of annual diameter growth (in centimeters) for shortleaf pine. The tree was constructed using ln(diameter growth) but was back-transformed for easier interpretation (see Table 1 for variable definitions). The first value in each box is the mean of diameter growth; the second value is the number of counties meeting the decision rule.

Stepwise model

When the SR procedure was used to fit diameter growth to the independent variables, a plot of the residuals again indicated that the variance was heterogenous. A natural log transformation was performed and the model was refit. All independent variables were available for selection regardless of possible multicolinearity. However, none of the variables chosen was highly correlated with each other. As in the RTA model, MAYSEPT (growing-season temperature) was the first entered, followed by variables indicating lands suitable for forests, nutrient availability, and the average initial dbh within the county (Table 3). Even though the R^2 value (0.383) is inflated due to the non-a priori character of the method, it is less than that of the PRE statistic of the RTA model. Plots of the residuals versus the predicted values and predicted versus actual values are shown in figure 2c-d. As in the RTA model, relationships between the independent and dependent variables are reasonable: counties that are warmer, flatter, and more fertile have larger diameter-growth rates. However, this model recognizes the negative relationship between ANNMOIST and diameter growth calculated from the FIA data. Diameter growth was not correlated to annual precipitation but was positively related to PET, probably because increased temperatures were needed to increase PET. And since PET is the divisor in the ANNMOIST calculation, a negative relationship would be expected. The RTA algorithm did not select a variable related to precipitation.

Table 3. Variables and coefficients developed by the SR model of the natural log of diameter growth (adjusted R^2=0.391); variables are in order of entry into the model.

Variable	Coefficient	p-value
Constant	-3.71437	0.00000
MAYSEPT	0.12791	0.00000
SLWILD	0.01006	0.00000
CEC	0.01125	0.02208
PH	0.11638	0.00035
SLWET	0.01180	0.00534
DBHOLD	-0.01220	0.00207
INCH3	0.04090	0.00077
OM	0.08504	0.00358
ANNMOIST	-0.00488	0.00039
PET	-0.00940	0.00076
SLOPE	-0.00854	0.04673

Gap Model

No algorithm was developed and no alterations were made to the existing GAP algorithm based on the development data set. Applying the GAP algorithm to this data set was for demonstration purposes only. Maximum diameter, maximum height, growth factor, and growing degree-day range for shortleaf pine as used by Solomon (1983) and the GAP algorithm used in ZELIG (Urban 1990) were combined with the gdd calculations presented previously to estimate diameter growth for the counties in our development data set; optimal soil moisture and fertility were assumed (figure 1). Modifying these estimates for low moisture availability and low fertility would not improve the estimates. The plot of residual versus predicted diameter-growth values and predicted versus actual diameter-growth values are presented in figure 2e-f.

Validation

Twenty-five percent of the observations (113 counties) were randomly set aside as a validation data set. Bias and MSE statistics were calculated for the models as applied to the validation data set (figure 4). The high level of bias for the SR and GAP algorithms indicates that these methods consistently overestimated diameter growth for the counties. However, the average error of the estimates for the SR algorithm was smaller based on the MSE. Both RTA and SR algorithms outperformed the GAP model in accuracy and precision.

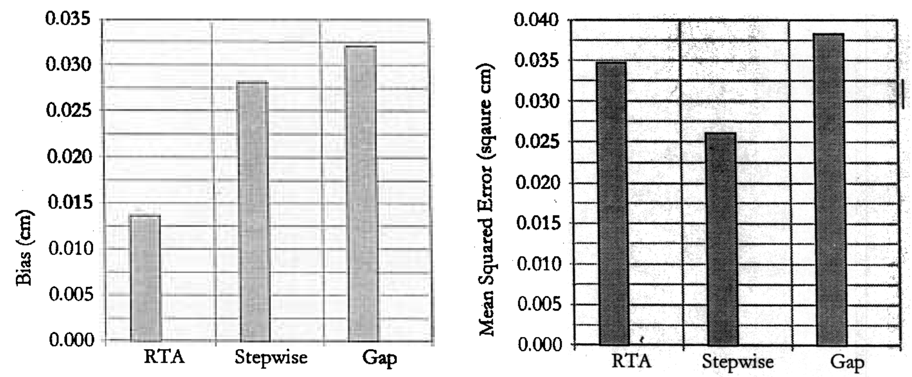

Figure 4. Comparison of the accuracy and precision of the RTA, SR, and GAP algorithms in predicting annual diameter growth of shortleaf pine when applied to an independent validation data set .

Diameter-growth measurements of the development and validation data sets were combined to produce the maps in figure 5. The predictions calculated using the RTA and SR approximate actual diameter growth better than the GAP predictions. The RTA algorithm tends to better identify the pockets of high growth rates at the northern range boundary than the SR algorithm.

CONCLUSIONS

RTA and SR provided algorithms that performed better than the GAP algorithm in predicting annual diameter growth based on statistical measures of accuracy and precision. The advantages of the GAP algorithm are that it is based on plant growth processes, is general to all species, performs with little site-specific data, and provides a mechanism to limit species ranges. This study showed that some of the assumptions used to develop the algorithm are false, at least for shortleaf pine. Diameter growth does not necessarily approach zero near the edges of a species gdd range. In fact, for shortleaf pine, some of the highest growth rates are near the range boundaries. Therefore, species ranges should not be determined solely by seasonal temperature ranges.

The site-specific data required by the RTA algorithm might seem obscure, but were readily calculated from publicly available GIS coverages. We intend to develop algorithms for more than 100 species of trees in the eastern United States, but at a finer scale. Algorithms of this type could replace the GAP algorithm in the gap models, allowing competition for light, water, and nutrients to determine species ranges as opposed to gdd limits. This information would provide for a persistence of species at the southern edge of their realized niche in simulations of climate change.

An efficient, realistic, general mechanistic model of diameter growth should be developed. These are being developed for certain species (Tree-BGC, Korol *et al.* 1996; TREGRO, Retzlaff *et al.* 1996). The scientific parameterizations of these algorithms for the tree species of the world will take time.

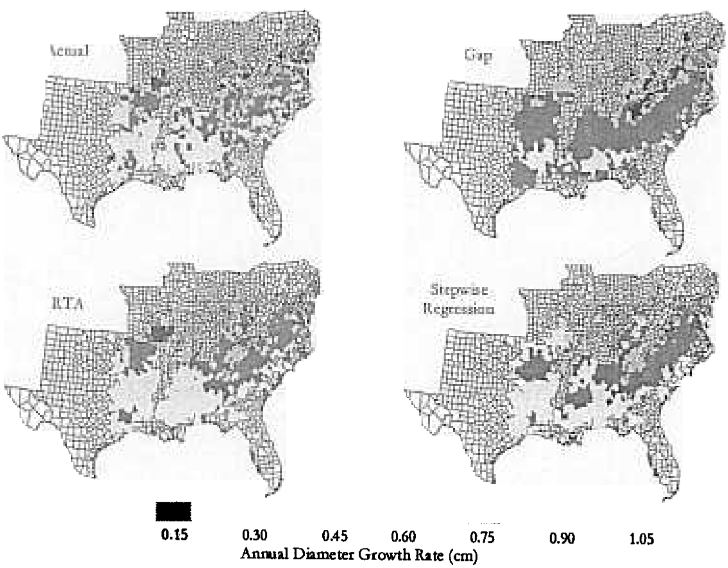

Figure 5. Comparison of actual diameter growth of shortleaf pine in the eastern United States and that predicted by three algorithms.

REFERENCES

Bonan, G. B. and L. Sirois. 1992. Air temperature, tree growth, and the northern and southern range limits to *Picea mariana*. Journal of Vegetation Science. 3:495-506.

Botkin, D.B. 1993. Forest dynamics: An ecological model. Oxford University Press, New York.

Botkin, D.B., J. F. Janak, and J. R. Wallis. 1972. Some ecological consequences of a computer model of forest growth. Journal of Ecology. 60:849-872.

Bowes, M. D. and R. A. Sedjo. 1993. Paper 3. Impacts and responses to climate change in forests of the Mink Region. Climatic Change. 24:63-82.

Breiman, L., J. Freidman, R. Olshen, and C. Stone. 1984. Classification and regression trees. Wadsworth, Belmont, CA.

Bugmann, H. K. M. 1996. A simplified forest model to study species composition along climate gradients. Ecology. 77(7):2055-2074.

Bugmann, H. K. M, Y. Xiaodong, M. T. Sykes, P. Martin, M. Lindner, P. V. Desanker, and S. G. Cumming. 1996. A comparison of forest gap models: Model structure and behaviour. Climatic Change. 6(4):289-309.

Clark, L. A. and D. Pregibon. 1992. Tree-based models. In: Chambers, J. M. and T. J. Hastie, eds. Statistical models in S. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. pp. 377-419.

Hansen, M. H., T. Frieswyk, J. F. Glover, and J. F. Kelly. 1992. The Eastwide Forest Inventory Data Base: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station.

Hilt, D. E. 1983. Individual-tree diameter growth model for managed, even-aged, upland oak stands. Res. Pap. NE- 533. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

Hilt, D. E. and R. M. Teck. 1989. NE-TWIGS: An individual-tree growth and yield projection system for the northeastern United States. The Compiler 7(2):10-16.

Iverson, L. R. and A. Prasad. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecology. [In press].

Iverson, L. R., A. Prasad, and C. T. Scott. 1996. Preparation of forest inventory and analysis (FIA) and state soil geographic data base (STATSGO) data for global change research in the eastern United States. In: Hom, J., R. Birdsey, and K. O'Brian, eds. Proceedings, 1995 meeting of the Northern Global Change Program. Gen. Tech. Rep. NE-214, Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station: pp. 209-214.

Korol, R. L., K. S. Milner, and S. W. Running. 1996. Testing a mechanistic model for predicting stand and tree growth. Forest Science. 42(2):139-153.

Loehle, C. 1996. Forest response to climate change: Do simulations predict unrealistic dieback? Journal of Forestry. 94(9):13-15.

Loehle, C. and D. LeBlanc. 1996. Model-based assessments of climate change effects on forests: A critical review. Ecological Modelling. 90:1-31.

Michaelsen, J., D. S. Schimel, M.A. Friedl, F. W. Davis, and R. C. Dubayah. 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. Journal of Vegetation Science. 5:673-686.

Moore, D. E., B. G. Lees, and S. M. Davey. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Journal of Environmental Management. 15:59-71.

Olson, R. J., C. J. Emerson, and M. K. Nungesser. 1980. GEOECOLOGY: a county-level environmental data base for the conterminous United States. Publ. No. 1537. Oak Ridge National Laboratory Environmental Sciences Division, Oak Ridge, TN.

Retzlaff, W. A., D. A. Weinstein, and J. A. Laurence, and B. Gollands. 1996. Simulated root dynamics of a 160-year-old sugar maple (*Acer saccharum* Marsh.) tree with and without ozone exposure using the TREGRO model. Tree Physiology. 16:1-7.

SAS Institute, Inc. 1990. SAS/STAT user's guide, release 6.03 edition. SAS Institute, Carey, NC.

Schenk, H. J.. 1996. Modelling the effects of temperature on growth and persistence of tree species: A critical review of tree population models. Ecological Modelling. 92(1):1-34.

Schwartz, M. W. 1993. Modeling the effect of habitat loss on potential rates of range change for trees in response to global warming. Biodiversity and Conservation. 2:51-6.

Schwartz, M. W. 1996. Assessing the ability of plants to respond to climatic change through distribution shifts. In: Horn, J., R. Birdsey, and K. O'Brian, eds. Proceedings, 1995 meeting of the Northern Global Change Program. Gen. Tech. Rep. NE-214. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

Schwartz, M. W., L. R. Iverson, and A. Prasad. 1996. Projected tree distribution shifts under global climate change in the fragmented Ohio landscape. Bulletin of the Ecological Society of America. 77:398.

Shifley, S. R. 1987. A generalized system of models forecasting Central States tree growth. Res. Pap. NC- 279. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station.

Shugart, H. H., Jr. and T. M. Smith. 1996. A review of forest patch models and their application to global change research. Climatic Change. 34(2):131-154.

Solomon, A.M. 1986. Transient response of forests to CO_2-induced climate change: simulation modeling experiments in eastern North America. Oecologia. 68:567-579.

Solomon, A. M. and P. J. Bartlein. 1992. Past and future climate change: response by mixed deciduous-coniferous forest ecosystems in northern Michigan. Canadian Journal of Forest Research. 22:1727-1738. SPSS, Inc. 1997. SYSTAT 7.0: New statistics. SPSS, Inc., Chicago, IL.

Talkkari, A. and H. Hypén. 1996. Development and assessment of a gap-type model to predict the effects of climate change on forests based on spatial forest data. Forest Ecology and Management. 83(3):217-228.

Urban, D.L. 1990. A versatile model to simulate forest pattern: A user's guide to ZELIG version 1.0. University of Virginia, Environmental Sciences Department. Charlottesville, VA.

Urban, Dean L., G. B. Bonan, T. M. Smith, and H. H. Shugart., Jr. 1991. Spatial applications of gap models. Forest Ecology and Management. 42(1,2):95-110.

U.S. Department of Agriculture, Soil Conservation Service. 1991. State soil geographic data base (STATSGO) data users guide. Misc. Publ. 1492. U.S. Department of Agriculture, Soil Conservation Service, Washington, DC.

U.S. Environmental Protection Agency. 1993. EPA-Corvallis model-derived climate database and 2xCO2 predictions for long-term mean monthly temperature, vapor pressure, wind velocity and potential evapotranspiration from the Regional Water Balance Model and precipitation from the PRISM model, for the conterminous United States. Digital raster data on a 10 x 10 km, 470 x 295 Albers Equal Area grid, in "Image Processing Workbench" format. U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR

U.S. Geological Survey. 1987. Digital elevation models: U.S. Geological Survey data user's guide 5. U.S. Geological Survey, Reston, VA.

Verbyla, D. L. 1987. Classification trees: a new discrimination tool. Canadian Journal of Forest Research. 17:1150-1152.

Walsh, S. J., D. R. Lightfoot, and D. R. Butler. 1987. Recognition and assessment of error in geographic information systems. Photogrammetric Engineering and Remote Sensing. 53:1423-1430.